

# Ensemble of Classifiers with Modification of Confidence Values

Robert Burduk, Paulina Baczyńska

► **To cite this version:**

Robert Burduk, Paulina Baczyńska. Ensemble of Classifiers with Modification of Confidence Values. 15th IFIP International Conference on Computer Information Systems and Industrial Management (CISIM), Sep 2016, Vilnius, Lithuania. pp.473-480, 10.1007/978-3-319-45378-1\_42 . hal-01637486

**HAL Id: hal-01637486**

**<https://hal.inria.fr/hal-01637486>**

Submitted on 17 Nov 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# Ensemble of Classifiers with Modification of Confidence Values

Robert Burduk, Paulina Baczyńska

Department of Systems and Computer Networks, Wrocław University of Technology,  
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland  
robert.burduk,paulina.baczynska@pwr.edu.pl

**Abstract.** In the classification task, the ensemble of classifiers have attracted more and more attention in pattern recognition communities. Generally, ensemble methods have the potential to significantly improve the prediction base classifier which are included in the team. In this paper, we propose the algorithm which modifies the confidence values. This values are obtained as an outputs of the base classifiers. The experiment results based on thirteen data sets show that the proposed method is a promising method for the development of multiple classifiers systems. We compared the proposed method with other known ensemble of classifiers and with all base classifiers.

**Keywords:** Multiple classifier system; Decision profile; Confidence value.

## 1 Introduction

For several years, in the field of supervised learning a number of base classifiers have been used in order to solve one classification task. The use of the multiple base classifier for a decision problem is known as an ensemble of classifiers (EoC) or as multiple classifiers systems (MCSs) [12], [15]. The building of MCSs consists of three main phases: generation, selection and integration (fusion) [4]. When injecting randomness into the learning algorithm or manipulating the training objects is done [3], [11], then we are talking about homogeneous classifiers. In the second approach the ensemble is composed of heterogeneous classifiers. It means, that some different learning algorithms are applied to the same data set.

The selection phase is related to the choice of a set of classifiers from the whole available pool of base classifiers [14]. Formally, if we choose one classifier then it is called the classifier selection. But if we choose a subset of base classifiers from the pool then it is called the ensemble selection or ensemble pruning. Generally, in the ensemble selection, there are two approaches: the static ensemble selection and the dynamic ensemble selection [2], [4].

A number of articles [7], [16],[21] present a large number of fusion methods. For example, in the third phase the simple majority voting scheme [19] is most popular. Generally, the final decision which is made in the third phase uses the prediction of the base classifiers and it is popular for its ability to fuse

together multiple classification outputs for the better accuracy of classification. The fusion methods can be divided into selection-based, fusion-based and hybrid ones according to their functioning [6]. The fusion strategies can be also divided into fixed and trainable ones [7]. Another division distinguishes class-conscious and class-indifferent integration methods [16].

In this work we will consider the modification confidence values method. The proposed method is based on information from decision profiles. The decision scheme used in training phase is created only from these confidence values that concern to the correct classification.

The remainder of this paper is organized as follows. Section 2 presents the concept of the base classifier and ensemble of classifiers. Section 3 contains the proposed modification confidence values method. The experimental evaluation, discussion and conclusions from the experiments are presented in Section 4. The paper is concluded by a final discussion.

## 2 Supervised classification

### 2.1 Base classifiers

The aim of the supervised classification is to assign an object to a specific class label. The object is represented by a set of  $d$  features, or attributes, viewed as  $d$ -dimensional feature vector  $x$ . The recognition algorithm maps the feature space  $x$  to the set of class labels  $\Omega$  according to the general formula:

$$\psi : X \rightarrow \Omega. \quad (1)$$

The recognition algorithm defines the classifier, which in the complex classification task with multiple classifiers is called a base classifier.

The output of a base classifier can be divided into three types [16].

- The abstract level – the classifier  $\psi$  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 and the label at the top being the first choice.
- The measurement level – the output of a classifier is represented by a confidence value (CV) 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. Generally, this level can provide richer information than the abstract and rank level.

In this work we consider the situation when each base classifier returns CVs. Additionally, before the final combination of base classifiers' outputs the CVs modification process is carried out.

## 2.2 Ensemble of classifiers

Let us assume that  $k \in \{1, 2, \dots, K\}$  different classifiers  $\Psi_1, \Psi_2, \dots, \Psi_K$  are available to solve the classification task. 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:

$$s_\omega(x) = \sum_{k=1}^K p_k(\omega|x), \quad \omega \in \Omega, \quad (2)$$

where  $p_k(\omega|x)$  is CV for class label  $\omega$  returned by classifier  $k$ .

The final decision of MCSs is made following the maximum rule:

$$\Psi_S(x) = \arg \max_{\omega} s_\omega(x). \quad (3)$$

In the presented method (3) CV 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 on 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 [5].

Decision template (DT) is another approaches to build the MCSs. DT was proposed in [17]. In this MCS model DTs are calculated based on training set. One DT per class label. In the operation phase the similarity between each DT and outputs of base classifiers for object  $x$  is computed. The class label with the closest DT is assigned to object  $x$ . In this paper algorithm with DT is labelled  $\Psi_{DT}$  and it is used as one of the reference classifiers.

## 3 Modification of confidence values algorithm

### 3.1 Training phase

[16] The proposed algorithm of the modification CVs values uses DPs. DP is a matrix containing CVs for each base classifier, i.e.:

$$DP(x) = \begin{bmatrix} p_1(0|x) & \dots & p_1(\Omega|x) \\ \vdots & \dots & \vdots \\ p_K(0|x) & \dots & p_K(\Omega|x) \end{bmatrix}. \quad (4)$$

In the first step of the algorithm we remove CVs which relate to the misclassification on the training set. This set contains  $N$  labelled examples  $\{(x_1, \bar{\omega}_1), \dots, (x_N, \bar{\omega}_N)\}$ , where  $\bar{\omega}_i$  is the true class label of the object described by feature vector  $x_i$ . CVs are removed according to the formula:

$$p'_k(\omega|x) = \begin{cases} p_k(\omega|x), & \text{if } I(\Psi(x), \bar{\omega}) = 1 \\ 0, & \text{if } I(\Psi(x), \bar{\omega}) = 0. \end{cases} \quad (5)$$

where  $I(\Psi(x), \bar{\omega})$  is an indicator function having the value 1 in the case of the correct classification of the object described by feature vector  $x$ , i.e. when  $\Psi(x) = \bar{\omega}$ .

In the next step, our algorithm, the decision scheme (DS) is calculated according to the formula:

$$DS(\beta) = \begin{bmatrix} ds(\beta)_{10} & \dots & ds(\beta)_{1\Omega} \\ \vdots & \dots & \vdots \\ ds(\beta)_{K0} & \dots & ds(\beta)_{K\Omega} \end{bmatrix}, \quad (6)$$

where

$$ds(\beta)_{k\omega} = \bar{ds}_{k\omega} + \beta \sqrt{\frac{\sum_{n=1}^N (p'_k(\omega_n|x_n) - \bar{ds}_{k\omega})^2}{N-1}} \quad (7)$$

and

$$\bar{ds}_{k\omega} = \frac{\sum_{n=1}^N p'_k(\omega_n|x_n)}{N}. \quad (8)$$

The parameter  $\beta$  in our algorithm determines how we compute DS elements. For example, if  $\beta = 0$ , then  $ds_{k\omega}$  is the average of appropriate DFs received after the condition (5).

### 3.2 Operation phase

During the operation phase the modification of CVs is carried out using  $DS(\beta)$  calculated in the training phase. For the new object  $x$  being recognized, the outputs of the base classifiers construct  $DP(x)$  as in (4). The modification CVs from  $DP(x)$  is performed with the use of  $DS(\beta)$  calculated in (6) according to the formula:

$$p'_k(\omega|x) = \begin{cases} \bar{m} * p_k(\omega|x) & \text{if } p_k(\omega|x) \geq ds(1)_{k\omega} \\ m * p_k(\omega|x) & \text{if } ds(0)_{k\omega} < p_k(\omega|x) < ds(1)_{k\omega}, \\ \underline{m} * p_k(\omega|x) & \text{if } p_k(\omega|x) \leq ds(0)_{k\omega} \end{cases} \quad (9)$$

where  $\bar{m}$ ,  $m$  and  $\underline{m}$  define how ordinal CVs are modified. The modification process taking into account the equation (9) causes the ensemble of the classifier method to use the modified CVs. The algorithm using the proposed method is denoted as  $\Psi^{MCV}$ . In the experiment modified CVs are combined according to the sum method (3).

## 4 Experimental studies

In the experiments we use 13 data sets. Nine of them come from the Keel Project [1] and four com from UCI Repository [9] (Blood, Breast Cancer Wisconsin, Indian Liver Patient and Mammographic Mass UCI). The details of the data sets are included in Tab. 1. In the experiment 16 base classifiers were used from four different classification models. The first group of four base classifiers works

according to  $k - NN$  rule, the second group uses the Support Vector Machines models. The next group uses the Neural Network model and the last group uses the decision trees algorithms. The base classifiers are labelled as  $\Psi_1, \dots, \Psi_{16}$ .

The main aim of the experiments was to compare the quality of classifications of the proposed modification CVs algorithms  $\Psi_{MCV}$  with the base classifiers  $\Psi_1, \dots, \Psi_{16}$  and their ensemble without the selection  $\Psi_{DT}$  and  $\Psi_S$ . The parameters of  $\Psi_{MCV}$  algorithm were established on  $\bar{m} = 1.25$ ,  $m = 1$  and  $\underline{m} = 0.75$ . In the experiments the feature selection process [13], [18] was not performed and we have used the standard 10-fold-cross-validation method.

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

Data set	example attribute ration (0/1)		
Cylinder Bands	365	19	1.7
Haberman's Survival	306	3	2.8
Hepatitis	80	19	0.2
Mammographic Mass	830	5	1.1
Parkinson	197	23	0.3
Pima Indians Diabetes	768	8	1.9
South African Hearth	462	9	1.9
Spectf Heart	267	44	0.3
Statlog (Heart)	270	13	1.3
Blood	748	5	3.2
Breast Cancer Wisconsin	699	10	1.9
Indian Liver Patient	583	10	0.4
Mammographic Mass UCI	961	6	1.2

The classification error with the mean ranks obtained by the Friedman test for classification algorithms used in the experiments are presented in Tab. 2. Considering only the mean ranks obtained by the Friedman test the best result achieved proposed in the work algorithm labelled  $\Psi_{MCV}$ . The obtained results were compared also by the post-hoc test [20]. The critical difference (CD) for this test at  $p = 0.05$  is equal to  $CD = 7.76$  – for 19 methods used for classification and 13 data sets. We can conclude that the post-hoc Nemenyi test detects significant differences between the proposed algorithm  $\Psi_{MCV}$  and the four base classifiers  $\Psi_4$ ,  $\Psi_6$ ,  $\Psi_8$ , and  $\Psi_{14}$ . The classifier method labelled  $\Psi_S$  is statistically better than the three base classifiers  $\Psi_6$ ,  $\Psi_8$  and  $\Psi_{14}$ . This confirms that the proposed algorithm is better than the ensemble of classifiers with the used sum method. It should be noted, however, that the difference in average ranks is not large enough to point to the significant differences between  $\Psi_{MCV}$  and  $\Psi_S$  algorithms.

**Table 2.** Classification error and mean rank positions for the base classifiers ( $\Psi_1, \dots, \Psi_{16}$ ), algorithms  $\Psi_S$ ,  $\Psi_{DT}$  and the proposed algorithm  $\Psi_{MCV}$  produced by the Friedman test

Clas.	Data set												Rank	
	Bands	Blood	Cance	Haber	Heart	Hepat	Liver	MamU	Mam	Park	Pima	Sahea		Spectf
$\Psi_1$	0.62	0.77	0.95	0.75	0.68	0.83	0.69	0.81	0.79	0.86	0.73	0.67	0.80	10.70
$\Psi_2$	0.56	0.79	0.96	0.76	0.69	0.83	0.70	0.81	0.77	0.81	0.73	0.69	0.78	10.58
$\Psi_3$	0.61	0.80	0.96	0.74	0.68	0.84	0.70	0.81	0.78	0.80	0.74	0.68	0.77	10.50
$\Psi_4$	0.63	0.76	0.95	0.75	0.70	0.84	0.65	0.81	0.77	0.86	0.72	0.64	0.80	11.81
$\Psi_5$	0.64	0.80	0.98	0.74	0.79	0.85	0.73	0.80	0.78	0.91	0.67	0.69	0.80	7.31
$\Psi_6$	0.64	0.74	0.87	0.69	0.56	0.84	0.28	0.70	0.74	0.45	0.66	0.66	0.25	16.54
$\Psi_7$	0.64	0.79	0.94	0.75	0.56	0.84	0.73	0.79	0.79	0.73	0.59	0.67	0.80	11.50
$\Psi_8$	0.64	0.77	0.93	0.71	0.56	0.84	0.74	0.80	0.76	0.74	0.66	0.66	0.80	13.04
$\Psi_9$	0.64	0.81	0.96	0.74	0.77	0.76	0.70	0.83	0.81	0.87	0.64	0.66	0.78	9.66
$\Psi_{10}$	0.65	0.81	0.95	0.72	0.77	0.76	0.70	0.84	0.81	0.88	0.65	0.66	0.78	9.47
$\Psi_{11}$	0.65	0.81	0.94	0.71	0.73	0.79	0.68	0.83	0.81	0.87	0.65	0.66	0.76	11.31
$\Psi_{12}$	0.66	0.80	0.96	0.71	0.77	0.85	0.69	0.82	0.81	0.93	0.66	0.66	0.79	8.27
$\Psi_{13}$	0.63	0.80	0.94	0.74	0.82	0.84	0.69	0.84	0.81	0.91	0.73	0.73	0.73	7.81
$\Psi_{14}$	0.61	0.78	0.94	0.67	0.74	0.85	0.62	0.82	0.80	0.85	0.72	0.66	0.75	12.85
$\Psi_{15}$	0.63	0.79	0.94	0.70	0.79	0.81	0.65	0.84	0.80	0.90	0.70	0.68	0.76	11.20
$\Psi_{16}$	0.66	0.80	0.94	0.73	0.80	0.85	0.66	0.83	0.80	0.88	0.72	0.72	0.69	8.50
$\Psi_{DP}$	0.65	0.69	0.97	0.73	0.78	0.89	0.64	0.81	0.79	0.84	0.71	0.65	0.78	10.89
$\Psi_S$	0.67	0.82	0.96	0.75	0.84	0.82	0.72	0.84	0.81	0.94	0.73	0.70	0.79	4.39
$\Psi_{MCV}$	0.68	0.82	0.97	0.74	0.83	0.84	0.70	0.84	0.81	0.93	0.74	0.71	0.80	3.74

## 5 Conclusion

In this paper we have proposed the methods that use information from the decision profiles and modified CVs received from the base classifiers. The aim of the experiments was to compare the proposed algorithm with all base classifiers and the ensemble classifiers based on the sum and decision profile methods. The experiments have been carried out on 13 benchmark data sets. We can conclude that the post-hoc Nemenyi test detects significant differences between the proposed algorithm  $\Psi_{MCV}$  and four base classifiers. While the algorithm  $\Psi_S$  classification results are statistically different from the three base classifiers. The obtained results show an improvement in the quality of the classification proposed algorithm  $\Psi_{MCV}$  with respect to all the base classifiers and the used reference ensemble of the classifiers' methods. Future work might involve the application of the proposed methods for various practical tasks [8], [10], [22] in which base classifiers are used. Additionally, the advantage of the proposed algorithm can be investigated as well as its ability to work in the parallel and distributed environment.

**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, Wrocław University of Technology.

## References

1. J. Alcalá, A. Fernández, J. Luengo, J. Derrac, S. García, L. Sánchez, and F. Herrera. Keel data-mining software tool: Data set repository, integration of algorithms

- and experimental analysis framework. *Journal of Multiple-Valued Logic and Soft Computing*, 17(255-287):11, 2010.
2. P. Baczyńska and R. Burduk. Ensemble selection based on discriminant functions in binary classification task. In *Intelligent Data Engineering and Automated Learning IDEAL 2015*, volume 9375 of *Lecture Notes in Computer Science*, pages 61–68. Springer International Publishing, 2015.
  3. L. Breiman. Randomizing outputs to increase prediction accuracy. *Machine Learning*, 40(3):229–242, 2000.
  4. A. S. Britto, R. Sabourin, and L. E. Oliveira. Dynamic selection of classifiers—a comprehensive review. *Pattern Recognition*, 47(11):3665–3680, 2014.
  5. R. Burduk. Classifier fusion with interval-valued weights. *Pattern Recognition Letters*, 34(14):1623–1629, 2013.
  6. A. M. Canuto, M. C. Abreu, L. de Melo Oliveira, J. C. Xavier, and A. d. M. Santos. Investigating the influence of the choice of the ensemble members in accuracy and diversity of selection-based and fusion-based methods for ensembles. *Pattern recognition letters*, 28(4):472–486, 2007.
  7. R. P. Duin. The combining classifier: to train or not to train? In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, volume 2, pages 765–770. IEEE, 2002.
  8. P. Forczmański and P. Łabędź. Recognition of occluded faces based on multi-subspace classification. In *Computer Information Systems and Industrial Management*, volume 8104 of *Lecture Notes in Computer Science*, pages 148–157. Springer, 2013.
  9. A. Frank and A. Asuncion. UCI machine learning repository, 2010.
  10. D. Frejlichowski. An algorithm for the automatic analysis of characters located on car license plates. In *Image Analysis and Recognition*, volume 7950 of *Lecture Notes in Computer Science*, pages 774–781. Springer, 2013.
  11. Y. Freund, R. E. Schapire, et al. Experiments with a new boosting algorithm. In *ICML*, volume 96, pages 148–156, 1996.
  12. G. Giacinto and F. Roli. An approach to the automatic design of multiple classifier systems. *Pattern Recognition Letters*, 22:25–33, 2001.
  13. H. H. Inbarani, A. T. Azar, and G. Jothi. Supervised hybrid feature selection based on pso and rough sets for medical diagnosis. *Computer methods and programs in biomedicine*, 113(1):175–185, 2014.
  14. K. Jackowski, B. Krawczyk, and M. Woźniak. Improved adaptive splitting and selection: The hybrid training method of a classifier based on a feature space partitioning. *International Journal of Neural Systems*, 24(03), 2014.
  15. M. Korytkowski, L. Rutkowski, and R. Scherer. From ensemble of fuzzy classifiers to single fuzzy rule base classifier. In *Artificial Intelligence and Soft Computing—ICAISC 2008*, volume 5097 of *Lecture Notes in Computer Science*, pages 265–272. Springer, 2008.
  16. L. I. Kuncheva. *Combining Pattern Classifiers: Methods and Algorithms*. John Wiley and Sons, Inc., 2004.
  17. L. I. Kuncheva, J. C. Bezdek, and R. P. Duin. Decision templates for multiple classifier fusion: an experimental comparison. *Pattern recognition*, 34(2):299–314, 2001.
  18. I. Rejer. Genetic algorithm with aggressive mutation for feature selection in bci feature space. *Pattern Analysis and Applications*, 18(3):485–492, 2015.
  19. D. Ruta and B. Gabrys. Classifier selection for majority voting. *Information Fusion*, 6(1):63–81, 2005.



20. B. Trawiński, M. Smętek, Z. Telec, and T. Lasota. Nonparametric statistical analysis for multiple comparison of machine learning regression algorithms. *International Journal of Applied Mathematics and Computer Science*, 22(4):867–881, 2012.
21. L. Xu, A. Krzyżak, and C. Y. Suen. Methods of combining multiple classifiers and their applications to handwriting recognition. *Systems, Man and Cybernetics, IEEE Transactions on*, 22(3):418–435, 1992.
22. R. Zdunek, M. Nowak, and E. Pliński. Statistical classification of soft solder alloys by laser-induced breakdown spectroscopy: review of methods. *Journal of the European Optical Society-Rapid publications*, 11, 2016.