Integration Base Classifiers Based on Their Decision Boundary

Robert Burduk(✉)

Department of Systems and Computer Networks,
Wroclaw University of Science and Technology,
Wybrzeze Wyspianskiego 27, 50-370 Wroclaw, Poland
robert.burduk@pwr.edu.pl

Abstract. Multiple classifier systems are used to improve the performance of base classifiers. One of the most important steps in the formation of multiple classifier systems is the integration process in which the base classifiers outputs are combined. The most commonly used classifiers outputs are class labels, the ranking list of possible classes or confidence levels. In this paper, we propose an integration process which takes place in the "geometry space". It means that we use the decision boundary in the integration process. The results of the experiment based on several data sets show that the proposed integration algorithm is a promising method for the development of multiple classifiers systems.

Keywords: Ensemble selection · Multiple classifier system · Decision boundary

1 Introduction

An ensemble of classifiers (EoC) or multiple classifiers systems (MCSs) [7,10] have been a very popular research topics during the last two decades. The main idea of EoC is to employ multiple classifier methods and combine their predictions in order to improve the prediction accuracy. Creating EoC is expected to enable better classification accuracy than in the case of the use of single classifiers (also known as base classifiers).

The task of constructing MCSs can be generally divided into three steps: generation, selection and integration [1]. In the first step a set of base classifiers is trained. There are two ways, in which base classifiers can be learned. The classifiers, which are called homogeneous are of the same type. However, randomness is introduced to the learning algorithms by initializing training objects with different weights, manipulating the training objects or using different features subspaces. The classifiers, which are called heterogeneous, belong to different machine learning algorithms, but they are trained on the same data set. In this paper, we will focus on homogeneous classifiers which are obtained by applying the same classification algorithm to different learning sets.

The second phase of building MCSs is related to the choice of a set of classifiers or one classifier from the whole available pool of base classifiers. If we

L. Rutkowski et al. (Eds.): ICAISC 2017, Part II, LNAI 10246, pp. 13–20, 2017.
DOI: 10.1007/978-3-319-59060-8_2

choose one classifier, this process will be called the classifier selection. But if we choose a subset of base classifiers from the pool, it will be called the ensemble selection. Generally, in the ensemble selection, there are two approaches: the static ensemble selection and the dynamic ensemble selection [1]. In the static classifier selection one set of classifiers is selected to create EoC during the training phase. This EoC is used in the classification of all the objects from the test set. The main problem in this case is to find a pertinent objective function for selecting the classifiers. Usually, the feature space in this selection method is divided into different disjunctive regions of competence and for each of them a different classifier selected from the pool is determined. In the dynamic classifier selection, also called instance-based, a specific subset of classifiers is selected for each unknown sample [2]. 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 or different decision regions [4]. The existing methods of the ensemble selection use the validation data set to create the so-called competence region or level of competence. These competencies can be computed by K nearest neighbours from the validation data set. In this paper, we will use the static classifier selection and regions of competence will be designated by the decision boundary of the base classifiers.

The integration process is widely discussed in the pattern recognition literature [13,18]. One of the existing way to categorize the integration process is by the outputs of the base classifiers selected in the previous step. Generally, the output of a base classifier can be divided into three types [11].

- 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 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 levels.

For example, when considering the abstract level, voting techniques [16] are most popular. As majority voting usually works well for classifiers with a similar accuracy, we will use this method as a baseline.

In this paper we propose the concept of the classifier integration process which takes place in the "geometry space". It means that we use the decision boundary in the integration process. The decision boundary is another type of information obtained from the base classifiers. In our approach, the decision boundary from the selected base classifiers is averaged in each region of competence separately.

Geometry reasoning is used in the formation of so-called "geometry-based ensemble". The method proposed in [12,14] used characteristic boundary points to define the decision boundary. Based on the characteristic boundary points, there is create a set of hyperplanes that are locally optimal from the point of view of the margin.

The remainder of this paper is organized as follows. Section 2 presents the basic concept of the classification problem and EoC. Section 3 describes the proposed method for the integration base classifiers in the "geometry space" which have been selected earlier (in particular we use Fisher linear discriminant method as a base classifier). The experimental evaluation is presented in Sect. 4. The discussion and conclusions from the experiments are presented in Sect. 5.

2 Basic Concept

Let us consider the binary classification task. It means that we have two class labels $\Omega = \{0, 1\}$. Each pattern is characterized by the feature vector x. The recognition algorithm Ψ maps the feature space x to the set of class labels Ω according to the general formula:

$$\Psi(x) \in \Omega. \tag{1}$$

Let us assume that $k \in \{1, 2, ..., K\}$ different classifiers $\Psi_1, \Psi_2, \ldots, \Psi_K$ 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 of all the classifiers' actions, their K responses are obtained. Usually all K base classifiers are applied to make the final decision of MCSs. Some methods select just one base classifier from the ensemble. The output of only this base classifier is used in the class label prediction for all objects. Another option is to select a subset of the base classifiers. Then, the combining method is needed to make the final decision of EoC.

The majority vote is a combining method that works at the abstract level. This voting method allows counting the base classifiers outputs as a vote for a class and assigns the input pattern to the class with the majority vote. The majority voting algorithm is as follows:

$$\Psi_{MV}(x) = \arg\max_{\omega} \sum_{k=1}^{K} I(\Psi_k(x), \omega), \tag{2}$$

where $I(\cdot)$ is the indicator function with the value 1 in the case of the correct classification of the object described by the feature vector x, i.e. when $\Psi_k(x) = \omega$. In the majority vote method each of the individual classifiers takes an equal part in building EoC. 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.

3 Proposed Method

The proposed method is based on the observation that the large majority of the integration process used the output of a base classifiers. In addition, the method called "geometry-based ensemble" used characteristic boundary points not the decision boundary [12,14]. Therefore, we propose the method of integrating (fusion) base classifiers based on their decision boundary. Since the proposed algorithm also uses the selection process, it is called *decision-boundary fusion with selection* and labelled Ψ_{DBFS}. The proposed method can be generally divided into five steps.

Step 1: Train each of base classifiers $\Psi_1, \Psi_2, \ldots, \Psi_K$ using different training sets by splitting according to the cross-validation rule.
Step 2: Divide the feature space in different separable decision regions. The regions can be found using points in which the decision boundaries of base classifiers are equal.
Step 3: Evaluate the base classifiers competence in each decision region based on the accuracy. The classification accuracy is computed taking into account the learning set of each base classifier separately.
Step 4: Select l best classifiers from all base classifiers for each decision regions, where $1 < l < K$.
Step 5: Define the decision boundary of the proposed EoC classifier Ψ_{DBFS} as an average decision boundary of the selected in the previous step base classifiers in the geometry space. The decision boundary of Ψ_{DBFS} is defined in each decision region separately. In this step we make the integration process of the selected base classifiers.

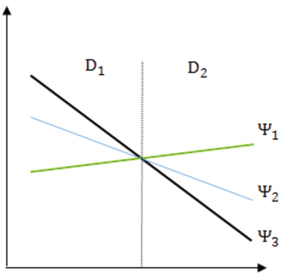
The decision boundary obtained in step 5 is applied to make the final decision of the proposed EoC. Graphical interpretation of the proposed method for two-dimensional data set and three base classifiers is shown in Fig. 1.

The method proposed above may be modified at various stages. For example, another division of the training set can be made using different subspaces of the feature space for different base classifiers or by using the bagging method. Another modification relates to step 2, when the competence regions can be found using a clustering method [9]. It should also take into account the fact that the method proposed in step 5 is suitable for linear classifiers.

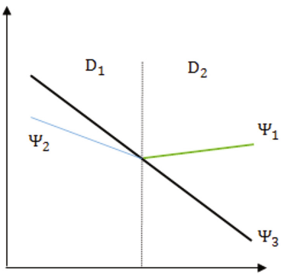
4 Experimental Studies

In the experiential research 6, benchmark data sets were used. Four of them come from the KEEL Project and two are synthetic data sets – Fig. 2. The details of the data sets are included in Table 1. All the data sets constitute two class problems. In the case of data sets with more than 2 features, the feature selection process [8,15] was performed to indicate two most informative features.

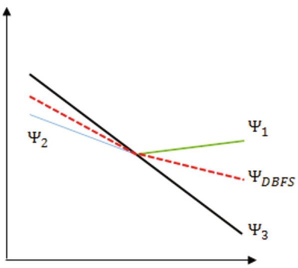
In the experiment 3 Fisher linear discriminant classifiers are used as base classifiers. This means that in the experiment we use an ensemble of the homogenous



(a) Base classifiers and decision regions

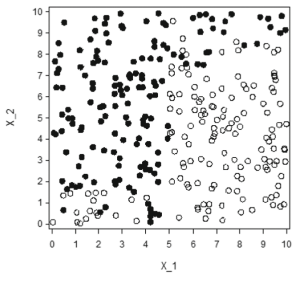


(b) Decision boundary of the base classifiers after selection

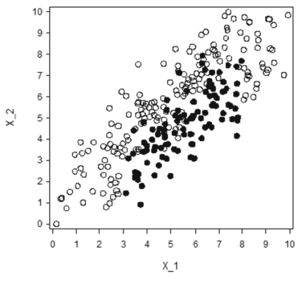


(c) Decision boundary of the proposed algorithm

Fig. 1. Example with two-dimensional data set and three base classifiers of the proposed method



(a) Synthetic(1)



(b) Synthetic (2)

Fig. 2. Syntectic data sets

base classifiers. Their diversity is created by learning either from subsets of the training patterns according to 3-cross-validation method. The learning process was repeated ten times. In each decision region two base classifiers are selected to perform “Step 5” from the algorithm proposed.

Table 2 shows the results of the classification error and the mean ranks obtained by the Friedman test for the proposed method Ψ_{DBFS} and the

Table 1. Description of data sets selected for the experiments

Data set	Example	Attribute	Ration (0/1)
Syntectic(1)$_{150}$	150	2	0.5
Syntectic(1)$_{300}$	300	2	0.5
Syntectic(1)$_{600}$	600	2	0.5
Syntectic(2)$_{150}$	150	2	0.5
Syntectic(2)$_{300}$	300	2	0.5
Syntectic(2)$_{600}$	600	2	0.5
Ionosphere	351	34	1.8
Pima Indians diabetes	768	8	1.9
Sonar	208	60	0.87
Ring$_{7400}$	7400	20	0.5
Ring$_{3700}$	3700	20	0.5
Ring$_{1850}$	1850	20	0.5

Table 2. Classification error and mean rank positions for the proposed method Ψ_{DBFS} and the majority voting method without selection Ψ_{MV} produced by the Friedman test

Data set	Ψ_{MV}	Ψ_{DBFS}
Syntectic(1)$_{150}$	20.36	19.40
Syntectic(1)$_{300}$	27.02	25.34
Syntectic(1)$_{600}$	23.82	22.89
Syntectic(2)$_{150}$	17.00	17.17
Syntectic(2)$_{300}$	17.29	16.36
Syntectic(2)$_{600}$	16.67	15.33
Ionosphere	18.70	19.20
Pima	25.02	25.10
Sonar	24.99	24.52
Ring$_{7400}$	35.03	33.83
Ring$_{3700}$	37.15	36.67
Ring$_{1850}$	39.53	39.36
Mean rank	1.25	1.75

majority voting method without selection Ψ_{MV}. The results were compared with the use of the post-hoc test [17]. This test is useful for pairwise comparisons of the methods considered. The critical difference (CD) for this test at $p = 0.05$, $p = 0.1$, equals $CD = 0.56$ and $CD = 0.47$ respectively. We can conclude that the post-hoc Nemenyi test detects significant differences between the proposed algorithm Ψ_{DBFS} and Ψ_{MV} method at $p = 0.10$. Additionally, at $p = 0.05$ the post-hoc test is not powerful enough to detect any significant differences between those algorithms, but the obtained difference between the mean ranks (0.5) is very close to $CD = 0.56$. This observation confirms that the algorithm Ψ_{DBFS} proposed in the paper can improve the quality of the classification as compared to the method without the selection. It should also be noted that the size of the data set does not allow formulating requests for increasing the data set size and the difference between considered algorithms.

5 Conclusion

In this paper we have proposed a concept of a classifier integration process taking place in the "geometry space". It means that we use the decision boundary in the integration process but we do not consider information produced by the base classifiers such as class labels, a ranking list of possible classes or confidence levels. In the proposed approach the selection process is carried out additionally, while the decision boundary from the selected base classifiers is averaged in each region of competence separately.

The experiments have been carried out on six benchmark data sets. The aim of the experiments was to compare the proposed algorithm Ψ_{DBFS} and the majority voting method without selection Ψ_{MV}. The results obtained show an improvement in the quality of the proposed method with respect to the majority voting method.

Future work might include another division of a training set using different subspaces of the feature space for different base classifiers, using the clustering method to partition the feature space in decision regions or application of the proposed methods for various practical tasks [3,5,6] in which base classifiers are used.

Acknowledgments. This work was supported by the statutory funds of the Department of Systems and Computer Networks, Wroclaw University of Science and Technology.

References

1. Britto, A.S., Sabourin, R., Oliveira, L.E.: Dynamic selection of classifiers-a comprehensive review. Pattern Recogn. **47**(11), 3665–3680 (2014)
2. Cavalin, P.R., Sabourin, R., Suen, C.Y.: Dynamic selection approaches for multiple classifier systems. Neural Comput. Appl. **22**(3–4), 673–688 (2013)
3. Cyganek, B., Woźniak, M.: Vehicle logo recognition with an ensemble of classifiers. In: Nguyen, N.T., Attachoo, B., Trawiński, B., Somboonviwat, K. (eds.) ACIIDS 2014. LNCS, vol. 8398, pp. 117–126. Springer, Cham (2014). doi:10.1007/978-3-319-05458-2_13
4. Didaci, L., Giacinto, G., Roli, F., Marcialis, G.L.: A study on the performances of dynamic classifier selection based on local accuracy estimation. Pattern Recogn. **38**, 2188–2191 (2005)

5. Forczmański, P., Łabędź, P.: Recognition of occluded faces based on multi-subspace classification. In: Saeed, K., Chaki, R., Cortesi, A., Wierzchoń, S. (eds.) CISIM 2013. LNCS, vol. 8104, pp. 148–157. Springer, Heidelberg (2013). doi:10.1007/978-3-642-40925-7_15
6. Frejlichowski, D.: An algorithm for the automatic analysis of characters located on car license plates. In: Kamel, M., Campilho, A. (eds.) ICIAR 2013. LNCS, vol. 7950, pp. 774–781. Springer, Heidelberg (2013). doi:10.1007/978-3-642-39094-4_89
7. Giacinto, G., Roli, F.: An approach to the automatic design of multiple classifier systems. Pattern Recogn. Lett. **22**, 25–33 (2001)
8. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. J. Mach. Learn. Res. **3**, 1157–1182 (2003)
9. 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**(03), 1430007 (2014)
10. Korytkowski, M., Rutkowski, L., Scherer, R.: From ensemble of fuzzy classifiers to single fuzzy rule base classifier. In: Rutkowski, L., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.) ICAISC 2008. LNCS (LNAI), vol. 5097, pp. 265–272. Springer, Heidelberg (2008). doi:10.1007/978-3-540-69731-2_26
11. Kuncheva, L.I.: Combining Pattern Classifiers: Methods and Algorithms. Wiley Inc., Hoboken (2004)
12. Li, Y., Meng, D., Gui, Z.: Random optimized geometric ensembles. Neurocomputing **94**, 159–163 (2012)
13. Ponti, Jr., M.P.: Combining classifiers: from the creation of ensembles to the decision fusion. In: 2011 24th SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T), pp. 1–10. IEEE (2011)
14. Pujol, O., Masip, D.: Geometry-based ensembles: toward a structural characterization of the classification boundary. IEEE Trans. Pattern Anal. Mach. Intell. **31**(6), 1140–1146 (2009)
15. Rejer, I.: Genetic algorithms for feature selection for brain computer interface. Int. J. Pattern Recogn. Artif. Intell. **29**(5), 1559008 (2015)
16. Ruta, D., Gabrys, B.: Classifier selection for majority voting. Inf. Fusion **6**(1), 63–81 (2005)
17. Trawiński, B., Smętek, M., Telec, Z., Lasota, T.: Nonparametric statistical analysis for multiple comparison of machine learning regression algorithms. Int. J. Appl. Math. Comput. Sci. **22**(4), 867–881 (2012)
18. Tulyakov, S., Jaeger, S., Govindaraju, V., Doermann, D.: Review of classifier combination methods. In: Marinai, S., Fujisawa, H. (eds.) Machine Learning in Document Analysis and Recognition. SCI, vol. 90, pp. 361–386. Springer, Heidelberg (2008)

Springer

http://www.springer.com/978-3-319-59059-2

Artificial Intelligence and Soft Computing
16th International Conference, ICAISC 2017, Zakopane, Poland, June 11-15, 2017, Proceedings, Part II
Rutkowski, L.; Korytkowski, M.; Scherer, R.; Tadeusiewicz, R.; Zadeh, L.A.; Zurada, J.M. (Eds.)
2017, XXIV, 742 p. 247 illus., Softcover
ISBN: 978-3-319-59059-2