Construction of Sequential Classifier Using Confusion Matrix

Robert Burduk, Pawel Trajdos

Construction of Sequential Classifier Using Confusion Matrix

Robert Burduk, Pawel Trajdos

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

Abstract. This paper presents the problem of building the decision scheme in the multistage pattern recognition task. This task can be presented using a decision tree. This decision tree is built in the learning phase of classification. This paper proposes a split criterion based on the analysis of the confusion matrix. Specifically, we propose the division associated with an incorrect classification. The obtained results were verified on the data sets form UCI Machine Learning Repository and one real-life data set of the computer-aided medical diagnosis.

Keywords: Multistage classifier, sequential classifier, confusion matrix

1 Introduction

The classification task may be divided according to its complexity. There are two groups here: one-step and multistage approach. In the one-step approach there is no division into smaller tasks classification. However, a multistage (sequential) approach breaks up a complex decision into a collection of several simpler decisions [1–4]. Many algorithm build a tree structure in the learning process [5, 6]. In other approaches the decision tree structure is fixed before the learning process [7]. Generally, the synthesis of the multistage classifier is a complex problem. It involves a specification of the following components [4, 8]:

- design of a decision tree structure [9],
- selection of features used at each non-terminal node of the decision tree [10–13],
- the choice of decision rules for performing the classification [14].

In particular, this paper discuses a way to design a decision tree structure. The split criterion is based on the confusion matrix. The potential division of the node is associated with the analysis of misclassification in the learning process. In the experiment decision rules are chosen arbitrarily in the entire tree.

The content of the work is as follows. Section 2 introduces the idea of the hierarchical (sequential) classifier. In Section 3 we describe the proposed split criterion. In the next section we present the results of the experiments verified on data sets form UCI repository and one real-life data set of the computer-aided medical diagnosis. The last section concludes the paper.

2 Hierarchical classifier

The hierarchical classifier contains a sequence of actions [15, 16]. These actions are simple classification tasks executed in the individual nodes of the decision tree. Some specific features are measured on every non-leaf node of the decision tree. At the first nonleaf node features x_0 are measured, at the second features x_1 are considered and so on. Every set of features comes from the whole vector of features. In every node of the decision tree the classification is executed according to the specific rule. The decisions $i_0, i_1, ..., i_N$ are the results of recognition in the suitable node of the tree. The design of a decision tree structure is based on split criterion.

In our task of classification the number of classes is equal to NC. The terminal nodes are labeled with the number of the classes from $M = 1, 2, ..., NC$, where M is the set of labels classes. The non-terminal nodes are labeled by numbers of 0, NC+1, NC+2 reserving 0 for the root-node. Let us introduce the notation for the received model of the multistage recognition [8]:

- $\overline{\mathcal{M}}$ – the set of internal (nonleaf) nodes,
- $\mathcal{M}_i$ – the set of class labels attainable from the i-th node ($i \in \overline{\mathcal{M}}$),
- $\mathcal{M}^i$ – the set of nodes of the immediate descendant node i ($i \in \overline{\mathcal{M}}$),
- m_i – the node of the direct predecessor of the i-th node ($i \neq 0$).

In each interior node the recognition algorithm is used. It maps observation subspace to the set of the immediate descendant nodes of the i-th node [17, 18]:

$$\Psi_i : X_i \to \mathcal{M}^i, \quad i \in \overline{\mathcal{M}} \ . \tag{1}$$

This approach minimizes the misclassification rate for the particular nodes of a tree. The decision rules at each node are mutually independent. In experiment the decision rules are chosen arbitrarily in the entire tree. Each of the classifiers used in the nodes of the tree takes a decision based on the full set of attributes available in the training set.

In our method of induction, the classification tree is a regular binary tree. This means that on each of the tree nodes there is a leaf or a node has two children.

Induction of the decision tree is performed by the top-down method. This means that it is initiated by the classifier located in the root of the tree. Using the proposed criterion the decision is made whether to continue the division. The process is repeated for the subsequent child nodes of the tree, until the state wherein the nodes in the tree can no longer be divided.

3 Split criteria

The division of the internal node will be made on the basis of the multidimensional confusion matrix. Specifically, we propose the division associated with an incorrect classification. This division is binary, which means that the node that

will meet the criterion of the split will have two child nodes. One of them represents a new internal decision tree node. The second one represents the label of a class that met the appropriate condition. The internal node of the decision tree is analysed in detail via the multidimensional confusion matrix. The columns of the confusion matrix correspond to the predicted labels (decisions made by the classifier in the internal node). The rows correspond to the true class labels. In this matrix the diagonal elements represent the overall performance of each label. The off-diagonal elements represent the errors related to each label.

Now we present the split criterion. For every class labels from internal node we create the $L \times L$ dimensional confusion matrix. Now we calculate the set of factors $W(k_l)$, where $l = 1, 2, ..., L$ is the number of class labels, according to the formula:

$$W(k_l) = \sum_{m=1, m \neq l}^{L} w_{l,m} + \sum_{m=1, m \neq l}^{L} w_{m,l}. \tag{2}$$

The example of the confusion matrix is presented in Tab. 1.

Table 1. The confusion matrix for the nonleaf node i

		estimated			
		k_1	k_2	...	k_L
true	k_1	$w_{1,1}$	$w_{1,2}$	...	$w_{1,L}$
	k_2	$w_{2,1}$	$w_{2,2}$	...	$w_{2,L}$
	$\vdots$	$\vdots$	$\vdots$		$\vdots$
	k_L	$w_{L,1}$	$w_{L,2}$	...	$w_{L,L}$

The division of the node occurs when

$$\left| \max_{l \in 1,2,...,L} W(k_l) - \min_{l \in 1,2,...,L} W(k_l) \right| > T, \tag{3}$$

where $T \in [0, 1]$ is a fixed threshold value. At the threshold one does not make a division of node. If we set this value at the beginning of the experiment, it indicates that the classification process is performed in the one-stage approach.

4 Experiments

In the experiential research 10 data sets were tested. Nine data sets come from UCI Machine Learning Repository [19]. The tenth comes from the Surgical Clinic Wroclaw Medical Academy and describes the acute abdominal pain diagnosis problem. 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 [20, 21]. The

numbers of attributes, classes and available examples of the investigated data sets are presented in Tab. 2.

Table 2. Description of data sets selected for the experiments

Data set	example	attribute	class
Acute Abdominal Pain	476	31	8
Breast Tissue	106	10	6
Ecoli	336	7	8
Glass Identification	214	10	6
Irys	150	4	3
Lung Cancer	31	52	3
Seeds	210	7	3
Vertebral Column	310	6	3
Wine	178	13	3
Yeast	1484	8	10

Tab. 3 presents the mean error for $5 - NN$ (5-nearest neighbor) classifier for the selected values of the parameter T. In Tab. 4 we presented the average ranks for all experiments. The average ranks are calculated on the basis of the Friedman test [22]. Each column in this table is attributed to one test. This means that the one-step approach $T = 1$ was compared in order to propose in the work the sequential approach. In Tab. 4 the lowest average ranks in each group are shown in bold.

Table 3. Avarage error for 5-NN classifier

Data set	T=0,02	T=0,04	T=0,02	T=0,4	T=0,5	T=1
Acute	0,161	0,161	0,163	0,165	0,168	0,163
Breast	0,428	0,419	0,438	0,489	0,457	0,412
Ecoli	0,128	0,131	0,124	0,131	0,131	0,131
Glass	0,336	0,344	0,333	0,327	0,34	0,343
Irys	0,03	0,035	0,03	0,03	0,035	0,035
Lung	0,472	0,525	0,537	0,458	0,487	0,557
Seeds	0,112	0,112	0,112	0,112	0,112	0,112
Vert.	0,184	0,181	0,178	0,178	0,181	0,184
Wine	0,302	0,303	0,297	0,322	0,325	0,319
Yeast	0,431	0,437	0,44	0,457	0,444	0,443

All classifiers from the group NN (Nearest Neighbor), with properly chosen values T, may improve the quality of classification. The value of this improvement, however, is not significant from the statistical point of view. For the post-hoc Bonferroni-Dunn [23] test the critical difference (CD) for the 27 values of the parameter T and 10 data sets is equal $CD = 11,7$. This CD is calculated at $\alpha = 0.05$. Although the differences of the average rank do not exceed CD,

Table 4. Average ranks from Friedman test

T	3-NN	5-NN	7-NN	9-NN	SVM
0	15,45	14,1	19	15,85	15,85
0.02	13.95	**7.8**	15.1	16.9	15.85
0.04	**9.95**	13.15	17	15.8	15.85
0.06	14.2	13.05	16.7	16.4	15.85
0.08	11.85	11.15	14	11.6	15.85
0.1	12.1	13.45	15.15	13.95	14.5
0.12	16.85	10.95	19.75	11.2	14.5
0.14	12.65	19.1	11.7	**7.85**	14.9
0.16	12.25	12	14.35	16.15	13.2
0.18	15.75	14.25	15.5	15.1	13.2
0.2	15.4	8.55	11.7	10.95	13.2
0.22	12.85	14.7	**7.45**	16.15	13.2
0.24	11.05	13.55	16.15	13.65	13.2
0.26	18.1	16.5	16.1	14.85	14.55
0.28	16	12.7	14.3	9.85	14.55
0.3	13.55	13.8	15.35	16.5	14.55
0.32	12	12.85	11.5	14.15	13.7
0.34	14	12.75	13.7	16.4	13.15
0.36	15.2	15.35	13.2	15.9	**13.15**
0.38	10.7	13.55	14.3	14.5	**13.15**
0.4	16.05	15.75	14.95	9.65	**13.15**
0.42	13.15	13.55	14.7	14.35	**13.15**
0.44	12.9	18	10.15	12.75	**13.15**
0.46	16.6	15.5	12.75	14.1	**13.15**
0.48	13.3	17.95	11.1	13.65	**13.15**
0.5	16.15	16.95	11.7	14.45	**13.15**
1	16	17	10.65	15.35	**13.15**

the obtained results can be considered promising because they are close to this value (in particular 5-NN).

5 Conclusions

In the paper we propose a split criterion based on analizing the confusion matrix. Specifically, we propose the division associated with an incorrect classification. This criterion is used in the design of a decision tree structure in the multistage classifier. With a fulfilled criteria a binary split of the analyzed decision node is carried out. If we set $T = 1$, it indicates that the classification process is performed in the one-stages approach. Which means that it does not creat a decision tree.

Experiments done in the work show that we have obtain promising results. The proposed approach improves the quality of classification for $k-NN$ group of classifiers. In some cases, the difference of mean ranks obtained by Friedman test is close to the critical difference. In the future work we can use different division criteria in order to design a sequential classifier. For example, the separable linearization [24] or MacArthur's [25] overlapping niches model can be used in the split criterium.

Acknowledgments. The work was supported in part by the statutory funds of the Department of Systems and Computer Networks, Wroclaw University of Technology and by the by The Polish National Science Centre under the grant N N519 650440 which is being realized in years 2011–2014.

References

1. Kołakowska, A., Malina, W.: Fisher Sequential Classifiers, IEEE Transaction on Systems, Man, and Cybernecics – Part B Cybernecics. vol. 35, no. 5, 988–998 (2005)
2. Mui, J., Fu K.S.: Automated classification of nucleated blood cells using a binary tree classifier, IEEE Trans. Pattern Anal. Mach. Intell. vol. PAMI-2, 429–443 (1980)
3. Podolak, I.T.: Hierarchical classifier with overlapping class groups. Expert Syst. Appl., vol. 34(1), 673–682 (2008)
4. Safavian, S.R., Landgrebe, D.: A survey of decision tree classifier methodology. IEEE Trans. Systems, Man Cyber., 21 (3), 660–674 (1991)
5. Penar, W., Woźniak, M.: Experiments on classifiers obtained via decision tree induction methods with different attribute acquisition cost limit. Advances in Soft Computing 45, 371–377 (2007)
6. Quinlan, J.R.: Induction on Decision Tree, Machine Learning, Vol. 1, 81–106 (1986)
7. Manwani, N., Sastry P.S.: Geometric decision tree, Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 42(1), 181–192 (2012)
8. Kurzyński, M.: Decision Rules for a Hierarchical Classifier. Pat. Rec. Let. 1, 305–310 (1983)
9. Woźniak, M.: A hybrid decision tree training method using data streams. Knowledge and Information Systems 29(2), 335–347 (2010)
10. Chorąś, M.: Image feature extraction methods for ear biometrics–A survey. In: Computer Information Systems and Industrial Management Applications, 2007. CISIM'07. 6th International Conference on. IEEE, 2007. p. 261–265 (2007)

11. Chorań, R.S.: Content-based retrieval using color, texture, and shape information. In: Progress in Pattern Recognition, Speech and Image Analysis, 619–626 (2003)
12. Guyon, I., Elisseeff, A.: An introduction to variable and feature selection. The Journal of Machine Learning Research 3, 1157–1182 (2003)
13. Rejer, I.: Genetic Algorithms in EEG Feature Selection for the Classification of Movements of the Left and Right Hand. Advances in Intelligent Systems and Computing 226, 579–589 (2013)
14. Burduk, R., Zmyślony, M.: Decomposition of classification task with selection of classifiers on the medical diagnosis example. Lecture Notes in Computer Science 7209, 569–577 (2012)
15. Burduk, R.: Classification error in Bayes multistage recognition task with fuzzy observations, Pattern Analysis and Applications, 13(1), 85–91 (2010)
16. Kurzyński, M.: On the Multistage Bayes Classifier, Pattern Recognition, 21, 355–365 (1988)
17. Berger, J.: Statistical Decision Theory and Bayesian Analysis. Springer-Verlag, New York (1993)
18. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, John Wiley and Sons (2000)
19. Frank, A., Asuncion, A.: UCI machine learning repository (2010)
20. Burduk, R., Woźniak, M.: Different decision tree induction strategies for a medical decision problem, Central European Journal of Medicine, 7(2), 183–193 (2010)
21. Kurzyński, M.: Diagnosis of acute abdominal pain using three-stage classifier, Computers in Biology and Medicine, vol. 17, no 1, 19–27 (1987)
22. Trawiński, B., Smetek, M., Telec, Z., Lasota, T.: Nonparametric statistical analysis for multiple comparison of machine learning regression algorithms. International Journal of Applied Mathematics and Computer Science 22 (4), 867–881 (2012)
23. Demšar, J.: Statistical comparisons of classifiers over multiple data sets. The Journal of Machine Learning Research 7, 1–30 (2006)
24. Bobrowski, L., Topczewska, M.: Separable Linearization of Learning Sets by Ranked Layer of Radial Binary Classifiers. Advances in Intelligent Systems and Computing 226, 131–140 (2013)
25. MacArthur, R.: On the relative abundance of bird species. Proc. Natl. Acad. Sci. USA 43, 293–295 (1957)