Effectiveness of Rotation Forest in Meta-learning Based Gene Expression Classification

Gregor Stiglic, Peter Kokol
University of Maribor, Slovenia
{gregor.stiglic, kokol}@uni-mb.si

Abstract

A lot of research has been done in the field of assembling classifiers in ensembles and on the other hand selecting the most appropriate single classifiers for a given problem which was solved by meta-learning techniques. This paper presents application of recently proposed ensemble of classifiers called Rotation Forest to Grading meta-learning scheme, where it is used as one of the base classifiers and meta-level classifier at the same time. Our proposed Grading variation is compared to four widely used classifiers on 14 datasets from the domain of gene expression classification problems. Experimental evaluations show that using Rotation Forest at meta-level most significantly impacts the accuracy of Grading scheme and confirms that it can be used for estimation of classifiers regions of strong and weak classification.

1. Introduction

There has been a lot of research done in the field of gene expression classification where supervised machine learning techniques have been adapted to a problem of extremely high dimensional data with only a few samples available. Machine learning techniques like Nearest Neighbors (kNN) [1,2], SVM [3], Decision Trees [4] and ensembles of different base classifiers [5], to name just a few, were proposed to solve this problem. As already stated in "No free lunch theorem" [6] there is no ultimate method and therefore searching for the method that would solve the most of gene expression classification problems at the lowest possible error rates still continues.

One of the recently proposed classifiers that is based on combining diverse base classifiers into an ensemble is called Rotation Forest and was presented by Rodriguez et al. in [7]. Rodriguez empirically showed that Rotation Forest outperforms other ensemble building methods on majority of tested datasets. One of the very useful characteristics of the proposed method is that it can be used in conjunction with practically any base classifier that is used in the ensemble creation process. Therefore there are still a lot of possible improvements of Rotation Forest that should be evaluated.

Another technique that aims to improve the accuracy of classification was proposed in [8] by Seewald and Furnkranz and is called Grading. It was inspired by idea of meta-classifiers where a so called meta-classifier is used to select the most appropriate classifier or group of classifiers for a given problem. Grading even goes a step further and is using meta-classifiers to describe regions of a classifier where it might be wrong and regions where the outcome of the classifier can be trusted with high confidence. Grading scheme can therefore be used for improvement of classification and also to evaluate ability of describing regions of weak and strong classification for a given meta-classifier.

Next section presents basics of Grading meta-classification scheme. Section 3 explains details of experimental settings that include short description of used classifiers, datasets and

Twentieth IEEE International Symposium on Computer-Based Medical Systems (CBMS'07)
0-7695-2905-4/07 $20.00

evaluation procedure. It is followed by Results section where different variants of Grading are compared. Section 5 containing discussion and future work plans concludes the paper.

2. Grading Classifiers

A meta-learning scheme consisting of grading classifiers was introduced by Seewald and Furnkranz in [8]. It was inspired by ideas from one of the first meta-classification schemes called Stacking [9] and research done by Bay and Pazzani [10] who introduced classifier error characterization. Initial step of Grading algorithm is basically the same as in Stacking, where all examples are classified by each base classifier using cross-validation. In classical Stacking meta-learning scheme those classification results would be used to create a new dataset where a meta-classifier would be trained. The final decision of Stacking is therefore based on classification results of all the base learners that are used as an input for meta-learning classifier that can predict the final class of a new example. Main difference between Stacking and Grading is the fact that in Grading a separate meta-classifier is built for each base classifier. Aim of each meta-classifier is to learn where a base learner is wrong and where it is producing accurate results. This is done by creation of separate dataset for each of the base classifiers with identical examples to the original training dataset, only that the class value represents if the base classifier classified the example correctly or not. Each of those datasets is used to train meta-learner which will be used at classification stage. At classification stage both base and meta-classifiers are used. Next, each of base classifiers classifies a given example. If the outcome of base classifier will be included in final prediction depends on the outcome of meta-classifier belonging to the base classifier. If meta-classifier is confident in outcome of base classifier for the given example it will be included in the final decision based on the confidence of meta-classifier. This is the reason why Grading works best with meta-classifiers that are able to predict class probability distribution (i.e. degree of confidence) and not just the class of their final decision.

3. Experimental Settings

This section describes datasets and experiments that try to focus on effectiveness of Rotation Forests when used as a meta-classifier in Grading scheme and at the same time evaluate the ability of describing weak and strong areas of different base classifiers. Another interesting aspect of this research is the application of Grading to gene expression classification datasets which present a completely different structure as most of usually used UCI Repository [11] datasets where most of datasets contain less than 20 numerical features.

3.1 Datasets

To empirically compare different variants of Grading and compare their performance to single base classifiers we chose 14 datasets from gene expression classification domain. Main characteristics of datasets are explained in Table 1. All of the datasets presented in Table 1 were obtained from Kent Ridge Biomedical Data Set Repository [12] and represent different gene expression classification problems. The same repository also contains detailed information on source of datasets and is available at http://sdmc.i2r.a-star.edu.sg/rp/.

Twentieth IEEE International Symposium on Computer-Based Medical Systems (CBMS'07)
0-7695-2905-4/07 $20.00 © 2007 IEEE

Table 1. Details of gene expression datasets

Dataset	Source	Genes	Patients	Classes
ALL	Yeoh et al. (2002)	12558	327	7
ALLAML	Golub et al. (1999)	7129	72	2
Breast	Van't Veer (2002)	24481	97	2
CNS	Mukherjee et al. (2002)	7129	60	2
Colon	Alon et al. (1999)	2000	62	2
DLBCL	Alizadeh et al. (2000)	4026	47	2
DLBCL-NIH	Rosenwald et al. (2002)	7399	240	2
DLBCL-Tumor	Shipp et al. (2002)	6817	77	2
Lung	Gordon et al. (2002)	12533	181	2
Lung-Harvard	Bhattacharjee et al. (2001)	12600	203	5
Lung-Michigan	Beer et al. (2002)	7129	96	2
MLL	Armstrong et al. (2001)	12582	72	3
Ovarian	Petricoin et al. (2002)	15154	253	2
Prostate	Singh et al. (2002)	12600	102	2

3.2 Rotation Forest

One of the integral parts for the success of Grading meta-learning classification scheme is a choice of meta-classifier. In this paper different combinations of base and meta-learning classifiers are evaluated with focus on effectiveness of Rotation Forest classifier. It is a recently proposed machine learning method for construction of classification ensembles. It was proposed by Rodriguez et al. [7] and is based on Principal Components Analysis (PCA) which is used for axis rotation of previously selected sets of features. This way a high number of accurate and diverse classifiers can be created that are assembled in an ensemble of classifiers. Because of their sensitivity to axis rotations, the most appropriate choice of base classifier for Rotation Forest is Decision Tree. In our experiments within Weka [13] machine learning environment a variant of C4.5 Decision Tree called J48 was used.

3.3 Base Classifiers

Four different base classifiers were used for empirical testing of Grading performance. Nearest Neighbors (NN) [14] and Support Vector Machines (SVM) [15] represent single classifiers, while Random [16] and Rotation Forests were used as representatives of classifier ensembles. Again Weka based implementations of the above classifiers were used – i.e. IBk representing NN method and SMO representing SVM. All experiments were done using 5 x 10-fold cross validation for each of 14 available datasets. All tests were designed to focus on impact of the right meta-classifier selection. For each fold the accuracies of all four base classifiers and all four Grading variants were measured. Four Grading variants consisted of all four base classifiers as base learners and one of them chosen as meta-classifier in each of variants. The settings used for training the base classifiers differed from default Weka setting at IBk, where weight with 1/distance was used, and Random Forests where 100 instead of 10 decision trees were built.

Another important pre-processing step that was necessary due to computational limitations was integration of feature selection method that reduced the initial datasets to 100 most relevant features. According to a recent extensive feature selection evaluation on gene expression datasets by Symons [17], where ReliefF feature selection method [18] achieved constantly accurate results, the same method was chosen for our pre-processing step. Even higher accuracy could be achieved by selection of features

0-7695-2905-4/07 $20.00

using SVM based Recursive Feature Elimination [19], but this would result in a high bias of SVM classifier as already described by Ambroise and McLachlan in [20]. On the other hand Symons confirmed that ReliefF can be used even with biased feature selection when comparison of classification performance is in question.

4. Results

As mentioned in section 3 all accuracies for classifiers were measured for base classifiers and their applications as meta-learners for Grading. Table 2 presents the results for four single classifiers on all datasets. It can be seen that both ensemble methods performed better than single classifiers but not on all datasets. This fact should be exploited by Grading where an optimal balance between classification outcomes of base learners is aimed to be found.

Table 2. Classification accuracy and standard deviation for base classifiers

	Random For.	Rotation For.	SMO	IBk
Colon	85.05 ± 15.79	86.43 ± 15.26	85.38 ± 15.01	85.1 ± 16.26
DLBCLTumor	94.46 ± 7.36	96.32 ± 6.23	97.11 ± 6.01	97.36 ± 5.65
DLBCL	95.8 ± 8.43	97.3 ± 6.33	97.8 ± 6.6	100 ± 0
Breast	78.04 ± 12.67	79.51 ± 12.57	77.56 ± 10.37	82.47 ± 12.06
Lung	98.79 ± 2.47	99.23 ± 1.88	98.9 ± 2.2	98.12 ± 3.08
MLL	94.71 ± 8.47	92.71 ± 9.22	97.21 ± 7.04	94.14 ± 9.59
Prostate	94.13 ± 6.78	94.73 ± 6.54	94.29 ± 6.77	93.16 ± 7.52
Ovarian	99.45 ± 1.65	99.61 ± 1.18	100 ± 0	99.13 ± 1.62
Lung-Harvard	94.19 ± 4.41	93.1 ± 5.23	94.19 ± 4.6	94.36 ± 4.5
DLBCL-NIH	71 ± 7.9	68.17 ± 8.21	61.58 ± 9.08	57.5 ± 1.67
ALL	89.29 ± 5.25	90.35 ± 5.35	90.08 ± 5.27	89.66 ± 5.11
AMLALL	98.68 ± 3.96	95.82 ± 6.3	97.54 ± 5.22	97.82 ± 5.26
CNS	77.67 ± 15.43	76.33 ± 15.52	74.33 ± 17.22	77 ± 15.64
Lung-Michigan	99.6 ± 1.2	100 ± 0	100 ± 0	100 ± 0
Average	**90.78 ± 7.27**	**90.69 ± 7.13**	**90.43 ± 6.81**	**90.42 ± 6.28**

Comparing Rotation and Random Forest it can be observed that Random Forest achieved higher accuracy, but overall Rotation Forest won on 9 datasets, while Random Forest managed 5 wins as it is shown in Figure 1.

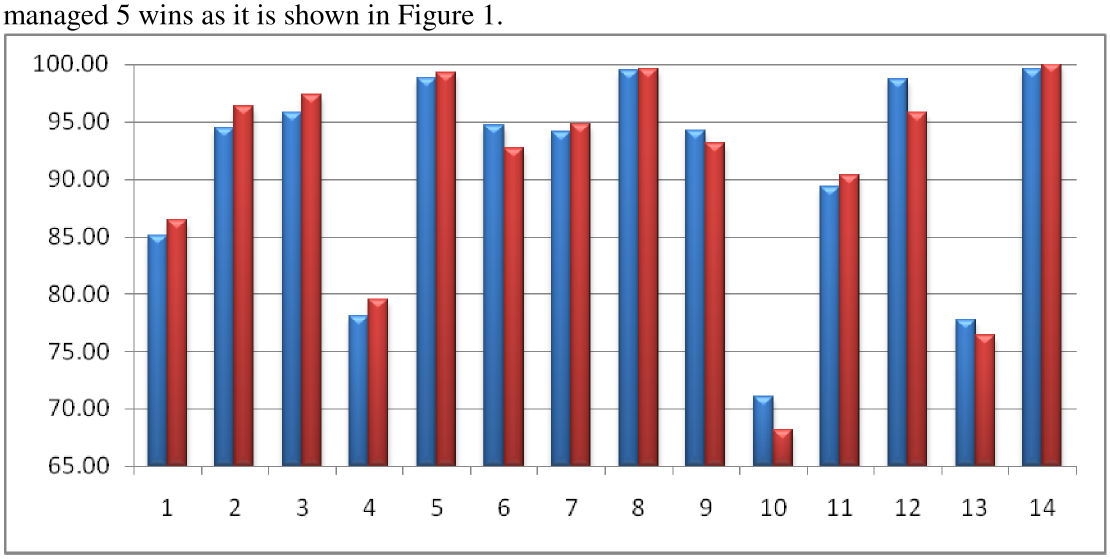

Figure 1. Comparison of classification accuracy for Random (left columns) and Rotation Forest (right columns) on all 14 datasets (ordered as in Table 2).

When comparing Grading schemes (Table 3) it is evident that Rotation Forest and SVM based Grading classifiers achieved the most notable improvement in comparison to accuracy of the base classifier itself (Figure 2).

Table 3. Classification accuracy and standard deviation for Grading with different meta-classifiers

	Grading based on			
	Random For.	Rotation For.	SMO	IBk
Colon	83.81 ± 13.98	84.43 ± 13.98	83.14 ± 14.73	82.81 ± 14.86
DLBCLTumor	97.39 ± 5.12	97.57 ± 5.12	96.54 ± 5.81	96.79 ± 5.52
DLBCL	99.1 ± 2.7	99.5 ± 1.5	99.5 ± 1.5	99.5 ± 1.5
Breast	80.64 ± 11.08	81.07 ± 12.96	83.27 ± 10.42	82.11 ± 11.75
Lung	98.9 ± 2.2	98.9 ± 2.2	98.9 ± 2.2	98.9 ± 2.2
MLL	95.79 ± 8.34	96.04 ± 7.72	95.79 ± 8.18	95.54 ± 8.26
Prostate	94.13 ± 6.78	94.13 ± 6.73	94.33 ± 6.36	93.73 ± 7.12
Ovarian	99.61 ± 1.18	99.37 ± 1.42	99.21 ± 1.68	99.21 ± 1.7
Lung-Harvard	94.18 ± 4.67	94.19 ± 3.92	93.9 ± 4.32	93.88 ± 4.6
DLBCL-NIH	68.08 ± 8.63	67.83 ± 7.76	65.08 ± 6.93	64.92 ± 7.36
ALL	91.19 ± 4.64	90.39 ± 4.85	90.33 ± 5.47	89.53 ± 5.04
AMLALL	98.11 ± 4.57	98.39 ± 4.29	98.39 ± 4.29	98.39 ± 4.29
CNS	77.67 ± 15.59	79.33 ± 15.17	79.33 ± 14.5	77.33 ± 14.61
Lung-Michigan	100 ± 0	100 ± 0	100 ± 0	100 ± 0
Average	**91.33 ± 6.39**	**91.51 ± 6.26**	**91.26 ± 6.17**	**90.9 ± 6.34**

Another interesting observation, when comparing performance of Rotation and Random Forests to that of Grading using each of them as meta-classifier is, that Grading improves the accuracy and at the same time decreases standard deviation comparing to base ensemble learners.

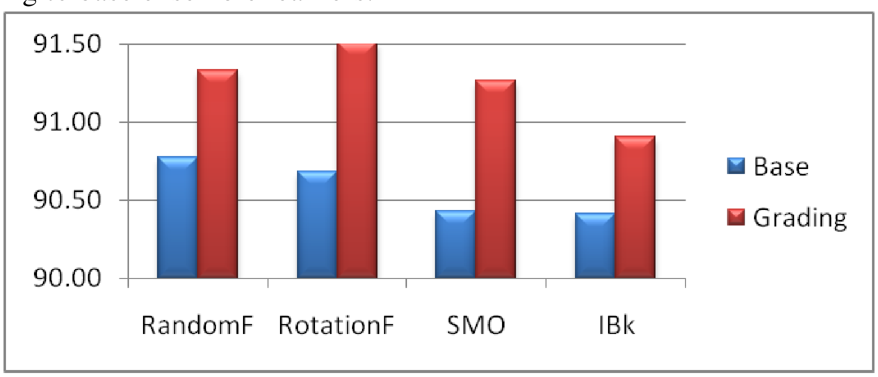

Figure 2. Comparison of classification accuracy for base classifiers in direct comparison to the same classifiers used as meta-classifiers for Grading

As expected there is no clear winner for all datasets. Even the best performing Grading scheme using Rotation Forest managed to win only in 4 and tie in 5 out of 14 datasets when compared to other three Grading variants.

5. Discussion and Future Work

In this paper we empirically evaluated effectiveness of Rotation Forest classifier when it is used as a meta-classifier for Grading meta-learning scheme. The results of Rotation Forest based Grading confirm that this classifier shows the ability to describe weak and strong regions of the base classifier with a high confidence. Additional to improvement of the accuracy it was evident that Grading returned more stable predictions as the standard deviation dropped at both ensemble based methods.

Twentieth IEEE International Symposium on Computer-Based Medical Systems (CBMS'07)
0-7695-2905-4/07 $20.00 © 2007 IEEE

However it would be relevant to test the performance of the proposed combination for Grading using different classification performance metrics or executing even more cross-fold validations because of random nature in some of the proposed meta-classifiers to gain more accurate results. One of the possible improvements would be addition of other ensemble building methods like Bagging or Boosting and using the best characteristics of each ensemble creation technique to improve overall performance of the final classifier.

6. Acknowledgement

Authors would like to thank Juan J. Rodriguez for providing valuable comments and his implementation of Rotation Forest in Weka.

7. References

[1] W. Wu, E. Xing, I. Mian, and M. Bissell, "Evaluation of normalization methods for cdna microarray data by k-nn classification",BMC Bioinformatics, vol. 6, issue 191, pp. 1-21, 2005.

[2] S. Dudoit, and J. Fridlyand, "Classification in microarray experiments", Statistical Analysis of Gene Expression Microarray Data, Interdisciplinary Statistics, Chapman & Hall/CRC, pp. 93– 158, 2003.

[3] T. Furey, N. Cristianini, N. Duffy, D. Bednarski, M. Schummer, and D. Haussler, "Support vector machines classification and validation of cancer tissue samples using microarray expression data", Bioinformatics, vol. 16, pp. 906-914, 2000.

[4] H. Liu, L. Jinyan, and W. Limsoon, "A Comparative Study on Feature Selection and Classification Methods Using Gene Expression Profiles and Proteomic Patterns", Genome Informatics, vol. 13, pp. 51-60, 2002.

[5] J. Lee, J. Lee, M. Park, and S. Song, "An extensive comparison of recent classification tools applied to microarray data", Computational Statistics & Data Analysis, vol. 48(4), pp. 869-885.

[6] D.H. Wolpert, and W.G. Macready, "No Free Lunch Theorems for Optimization", IEEE Transactions on Evolutionary Computation, vol. 1(1), pp. 67-82, 1997.

[7] J. J. Rodríguez, L. I. Kuncheva, and C. J. Alonso, "Rotation forest: A new classifier ensemble method," IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(10), pp. 1619-1630, 2006.

[8] A.K. Seewald, and J. Furnkranz, "An Evaluation of Grading Classifiers," Advances in Intelligent Data Analysis, 4th International Conference, IDA 2001, Proceedings, Springer, Berlin, pp. 115-124, 2001.

[9] D.H. Wolpert, "Stacked generalization," Neural Networks, vol. 5(2), pp. 241-259, 1992.

[10] S. D. Bay, and M. J. Pazzani, "Characterizing model errors and differences," 17th International Conference on Machine Learning, Morgan Kaufmann, San Francisco, CA, pp. 49-56, 2000.

[11] C.L. Blake, and C.J. Merz, "UCI Repository of machine learning databases," Irvine, CA: University of California, Department of Information and Computer Science.

[12] J. Li and H. Liu, "Ensembles of cascading trees," in Proc. of IEEE ICDM 2003 Conference, pp. 585, 2003.

[13] I.H. Witten, and E. Frank, Data Mining: Practical machine learning tools with Java implementations. Morgan Kaufmann, San Francisco, 2005.

[14] T.M. Cover, and P.E. Hart, "Nearest Neighbor Pattern Classification," IEEE Transactions on Information Theory, vol. IT-13(1), pp.21-27, 1967.

[15] V. Vapnik, Statistical learning theory. New York: John Wiley and Sons, 1998.

[16] L. Breiman, "Random forests," Machine Learning, vol. 45, pp. 5-32, 2001.

[17] S. Symons, "Machine Learning Algorithms for Microarray Data", Diploma Thesis, Eberhard Karls University, Tubingen, Germany.

[18] I. Kononenko, "Estimating attributes: analysis and extensions of Relief", Machine Learning: ECML-94, Springer Verlag, pp. 171–182, 1994.

[19] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene Selection for Cancer Classification using Support Vector Machines", Machine Learning, vol. 46(1–3), pp. 389-422, 2002.

[20] C. Ambroise, and G.J. McLachlan, "Selection bias in gene extraction on the basis of microarray gene-expression data", Proc Natl Acad Sci USA 2002, 99:6562-6566.