Meta Learning on Small Biomedical Datasets

Turgay Ibrikci, Esra Mahsereci Karabulut and Jean Dieu Uwisengeyimana

Abstract Meta-learning is one of subsections of supervised machine learning that has continuously grown with interests to apply on new data sets in the late years. Meta learning is the process of knowledge that is acquired by the examples. Bagging, dagging, decorate, rotation forest, and filtered classifiers are well known meta-learning algorithms that are performed to compare with these meta-learning algorithms on 8 different biomedical datasets. In these algorithms, the rotation forest had the better results according to F-measurement and ROC Area in most cases.

Keywords Machine learning · Meta learning · Bagging · Dagging · Decorate · Rotation forest · Filtered classifiers

1 Introduction

One of the machine learning algorithms, meta learning is automatic learning algorithms, set by Donald B.Maudsley [1], that are applied on data to understand the interaction between the mechanism of learning and the concrete contexts Meta learning provides one such methodology that allows systems to become more effective through experience. Meta learning differs from base learning in the scope of the level of the adaptation. Learning at the base level is focused on accumulating experience on a specific learning task whereas learning at the meta-level is concerned with accumulating experience on the performance of multiple applications of learning. In the last years, there are many published works that has been done in data mining with meta learning on different biomedical diagnosis problems at many journals and conferences.

T. Ibrikci(✉) · J.D. Uwisengeyimana
Electrical-Electronics Engineering Department, Cukurova University, Adana, Turkey
e-mail: ibrikci@cukurova.edu.tr, uwisenjeado@gmail.com

E.M. Karabulut
Gaziantep Vocational High School, Gaziantep University, Şehitkamil/Gaziantep, Turkey
e-mail: esra.mkarabulut@gmail.com

K.J. Kim and N. Joukov (eds.), *Information Science and Applications (ICISA) 2016*,
Lecture Notes in Electrical Engineering 376,
DOI: 10.1007/978-981-10-0557-2_89

Machine learning algorithms have been proposed many different algorithms for biomedical diagnostic problems, for instances: in breast cancer [2], coronary heart disease [3], and estimation bioinformatics inference systems [4].

These algorithms have used different types of neural network algorithms that are considered to bring several benefits in a machine learning process, such as faster learning, or improved classification performance. Meta-learning algorithms, by supporting studies on this type of algorithms on data, are to ensure that all the successful outcome. The algorithms demonstrate the success rate of between 80% and 95%.

2 Data Sets and Algorithms

2.1 Descriptions of Data Sets

The datasets used in this study are publicly available at "The Data Mining Repository of University of California Irvine (UCI)" [5]. Table 1 summarizes the biomedical datasets with respect to number of instances, attributes and classes.

Table 1 Data sets taken from UCI.

No	Data	Instances	Attributes	Classes
1	Arrhythmia [6]	452	279	2
2	Heart disease(Cleveland) [7]	303	13	5
3	Vertebral column (2C) [8]	310	6	2
4	CTG [9]	2126	21	3
5	Diabetes (Pima Indians) [10]	768	8	2
6	Mammographic mass [11]	961	5	2
7	Parkinson [12]	194	22	2
8	Wisconsin breast cancer [13]	699	9	2

2.2 Algorithms

Bagging: Bagging is a bootstrap method for improving the accuracy of the model by using the multiple random redistribution copies of the training set [14]. The model decreases variance of the base model without changing the bias. Main point on bagging algorithm, average of misclassification errors on divided different subset of the data gives a better estimate of the predictive ability of a learning method. Thus, bagging pursues to reduce the error rate by using a variance of the base classifier.

Dagging: Dagging is one of the most popular ensemble algorithms that creates a number of disjoint, stratified folds out of the data and feeds each chunk of data to a copy of the supplied base learner, while predictions are made via majority vote [15].

Decorate: Diverse Ensemble Creation by Oppositional Relabeling of Artificial Training Examples directly builds diverse hypotheses using additional artificially constructed training examples. It is a simple and general meta-learner that can decide to use any strong learner as a base classifier to build diverse groups [16].

Rotation Forest: Rotation Forest is also one method for generating classifier ensembles based on feature extraction. The idea of the rotation approach is to encourage simultaneously individual accuracy and diversity within the ensemble. Diversity is promoted through the feature extraction for each base classifier. Decision trees are most often chosen because they are sensitive to rotation of the feature axes, hence the name "forest." Accuracy is sought by keeping all principal components and also using the whole data set to train each base classifier [17].

Filtered Classification: Meta-learning algorithms also provide filtering on the data. This filter is generated using the training data, and then applied to the test data. The filter will be processed on the test data without any changing the structure of it [18].

3 Results and Discussion

We evaluate 5 meta-learning algorithms by WEKA data mining tools [19] on 8 datasets using the 10-fold cross-validation accuracy. The results are explained in *recall, false positives rate (FPR), area under ROC curve (AUC)* and *F-Measure (F-m)*.

The calculation of the metrics requires the outcomes of the classifier system to be labeled with four possible states, as true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). TP refers to the number of samples predicted as positive and actually they are positive. FP defines to the number of samples predicted as positive but actually they are negative. FN refers to the number of samples predicted as negative but actually they are positive TN refers to the number of samples predicted as negative and actually they are negative.

Once the output for each sample is labeled, quantitative metrics, as well as AUC, can be calculated for the evaluated algorithm. FPR is the ratio of incorrectly predicted positives to the total of actual negatives, and defined as

$$FPR = \frac{PFP}{FP+TN} \tag{1}$$

The F-measure (F-m) is described as a harmonic mean of two other classification metrics; precision and recall. Precision is the ratio of truly classified positives to all predicted positives defined as

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

Recall, also known as TPR, is the ratio of number of correctly predicted positives to the total of actual positives, and defined as

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$F-m = 2 * \frac{Precision*Recall}{Precision+Recall} \tag{4}$$

AUC measures the classifier's skill in ranking a set of patterns according to the degree to which they belong to the positive class, but without actually assigning patterns to classes. TPR and FPR plot a ROC curve on two axes by using values. Thus, each point on the ROC curve represents a TPR / FPR pair corresponding to a particular decision threshold [20].

Table 2 The results of algorithms experimented on the datasets

Model	Data	Recall	FPR	F-m	AUC	Recall	FPR	F-m	AUC	Model
	Arrhytmia	0.816	0.190	0.816	0.861	0.801	0.204	0.769	0.766	
B	*Cleveland*	0.601	0.245	0.547	0.791	0.561	0.221	0.479	0.679	**R.**
A	*Column2*	0.829	0.218	0.829	0.913	0.858	0.209	0.795	0.843	**F**
G	*CTG*	0.939	0.145	0.937	0.974	0.946	0.119	0.924	0.937	**o**
G	*Diabetes*	0.754	0.317	0.751	0.814	0.762	0.337	0.732	0.781	**r**
I	*Mommog*	0.835	0.169	0.834	0.895	0.826	0.176	0.829	0.868	**e**
N	*Parkinson's*	0.877	0.251	0.874	0.933	0.913	0.169	0.852	0.786	**s**
G	*Wisconsin*	0.960	0.045	0.960	0.990	0.971	0.025	0.933	0.939	**t**
	Average	0.826	0.197	0.818	0.896	0.829	0.182	0.789	0.824	
	Arrhytmia	0.759	0.257	0.756	0.813	0.770	0.239	0.801	0.878	**F**
D	*Cleveland*	0.558	0.237	0.517	0.772	0.525	0.279	0.532	0.783	**i**
A	*Column2*	0.684	0.648	0.578	0.790	0.794	0.255	0.856	0.936	**l**
G	*CTG*	0.866	0.318	0.854	0.879	0.927	0.187	0.945	0.987	**i**
G	*Diabetes*	0.736	0.448	0.701	0.772	0.737	0.345	0.754	0.818	**r**
I	*Mommog*	0.814	0.180	0.814	0.870	0.829	0.174	0.826	0.897	**e**
N	*Parkinson's*	0.846	0.401	0.831	0.768	0.856	0.299	0.912	0.957	**d**
G	*Wisconsin*	0.964	0.046	0.964	0.974	0.933	0.075	0.972	0.988	**C**
	Average	0.778	0.316	0.751	0.829	0.796	0.231	0.824	0.905	
	Arrhytmia	0.810	0.198	0.809	0.873					
D	*Cleveland*	0.548	0.233	0.518	0.74					
E	*Column2*	0.816	0.25	0.815	0.901					
C	*CTG*	0.935	0.124	0.934	0.971					
O	*Diabetes*	0.736	0.346	0.731	0.803					
R	*Mommog*	0.827	0.176	0.827	0.883					
A	*Parkinson's*	0.923	0.165	0.922	0.943					
T	*Wisconsin*	0.963	0.047	0.963	0.992					
E	***Average***	0.819	0.192	0.814	0.888					

According to Table 2, considering the average values, it can be seen that Rotation Forest algorithm outperforms other models with respect to TPR and FPR metrics, with the values of 0.829 and 0.182, respectively, as the highest one in TPR and the lowest one in FPR. This means minimum error is achieved with respect to positive labeled samples. Filtered classification model is the best at

F-measure and area under ROC curve, the metrics that we aim to approach 1. F-measure and AUC trade off precision versus recall, therefore take the negative labeled samples into consideration too. This means that when both of the positive and negative samples are respected Filtered model should be used.

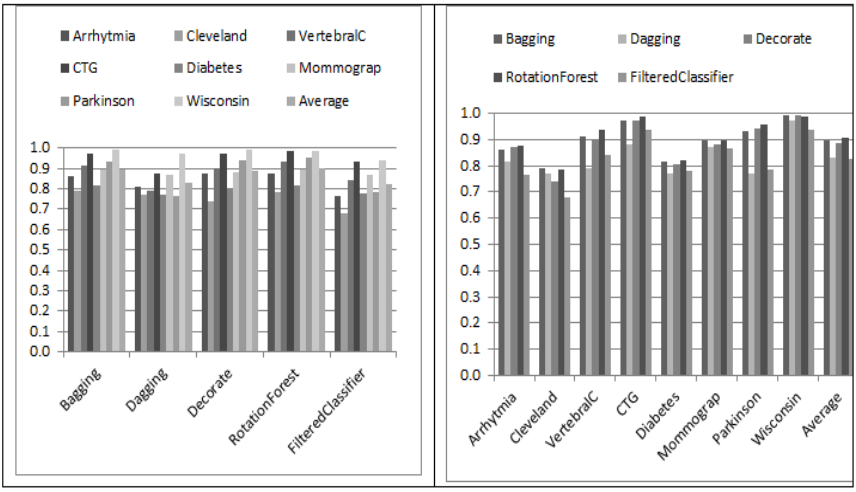

Fig. 1 ***a)*** The ROC Area measure results of the algorithms ***b)*** The results of the algorithms on the datasets with ROC Area

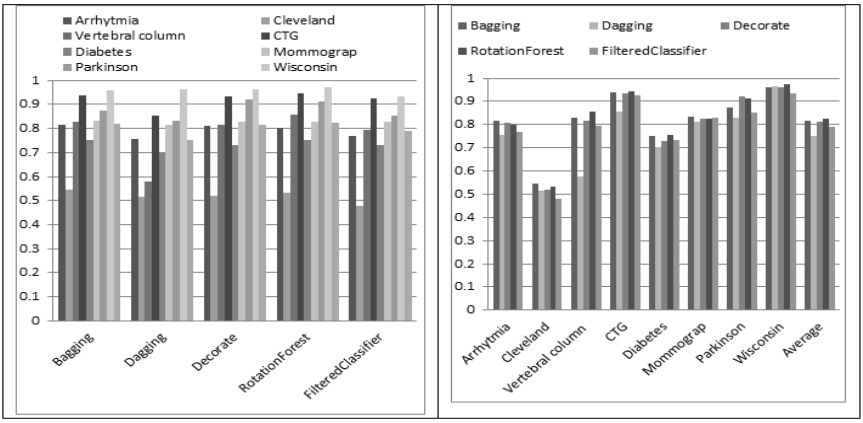

Fig. 2 ***a)*** The F- measure results of the algorithms ***b)*** The results of the algorithms on the datasets with F-Measure

If the class distribution is not balanced in a dataset accurate measure of classifier performance is more important. In this situation one metric may be satisfactory but another may not. Therefore a combined metric should be used such as F-measure and ROC area to determine which model is superior to another. In this study Cleveland, vertebral column, CTG, and diabetes datasets include imbalanced class distribution therefore F-measure and ROC Area results are took into consideration. On these datasets Filtered column produced the best F-measure results of 0.532, 0.856, 0.945 and 0.754 respectively. Furthermore ROC Area results of Filtered column are best on vertebral column; CTG and diabetes are best with the values of 0.936, 0.987, 0.818 and 0.897 respectively.

4 Conclusions

In this paper, we evaluated common meta learning classifiers on small size of different type of biomedical datasets. Meta learning algorithms showed some encouraging results with these datasets. The best results achieve an *F-measure* of 0.972 on the Wisconsin Breast Cancer Data Set with *rotation forest* algorithm and ROC Area of 0.992 on the Wisconsin Breast Cancer Data Set with decorate algorithm. The experimental results demonstrate that especially, the rotation forest classifier method is a suitable method for biomedical data sets with the evaluation of F-Measure and ROC Area results. In future work, we will focus to find the alternative Neural Networks methods to improve the system performances on the biomedical data sets.

We conclude this paper by emphasizing the important role of meta-learning that in particular, meta-learning can serve as a useful algorithm for classifying the medical datasets with exploitation of knowledge. It has a strong potential impact in medical applications.

Acknowledgements This work is supported by Cukurova University Research Fund and BAP Unit(ID#5603). The author, Jean Dieu Uwisengeyimana is supported by TUBITAK – BIDEB 2235.

References

1. Maudsley, D.B.: A Theory of Meta-Learning and Principles of Facilitation: An Organismic Perspective. University of Toronto (1979). 40, 8, 4354-4355-A
2. Muresan, S.: Pre-processing flow for enhancing learning from medical data. In: Int. Computer Comm. and Processing (ICCP), pp. 27–34 (2015)
3. El-Bialy, R., Salamay, M.A., Karam, O.H., Khalifa, M.E.: Feature Analysis of Coronary Artery Heart Disease Data Sets. Procedia Computer Science **65**, 459–468 (2015)
4. Arredondo, T., Ormazabal, W.: Meta-learning framework applied in bioinformatics inference system design. Int. J. Data Min. Bioinform. **11**(2), 139–166 (2015)
5. Lichman, M.: UCI Machine Learning Repository. University of California, School of Information and Computer Science, Irvine (2013). http://archive.ics.uci.edu/ml

6. Guvenir, H.A., Acar, B., Demiroz, G., Cekin, A.: A supervised machine learning algorithm for arrhythmia analysis. In: Proceedings of the Comp. in Cardiology Conference, vol. 24, pp. 433–436 (1997)
7. Detrano, R., Janosi, A., Steinbrunn, W., Pfisterer, M., Schmid, J., Sandhu, S., Guppy, K., Lee, S., Froelicher, V.: International application of a new probability algorithm for the diagnosis of coronary artery disease. A. Journal of Cardiology **64**, 304–310 (1989)
8. RochaNeto, A.R., Barreto, G.A.: On the Application of Ensembles of Classifiers to the Diagnosis of Pathologies of the Vertebral Column: A Comparative Analysis. IEEE Latin America Transactions **7**(4), 487–496 (2009)
9. de Campos, A., et al.: SisPorto 2.0 A Program for Automated Analysis of Cardiotocograms. J. Matern. Fetal Med. **5**, 311–318 (2000)
10. Smith, J.W., Everhart, J.E., Dickson, W.C., Knowler, W.C., Johannes, R.S.: Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. In: Proceedings of the Symposium on Computer Applications and Medical Care, pp. 261–265 (1988)
11. Elter, M., Schulz-Wendtland, R., Wittenberg, T.: The prediction of breast cancer biopsy outcomes using two CAD approaches that both emphasize an intelligible decision process. Medical Physics **34**(11), 4164–4172 (2007)
12. McSharry, P.E., Roberts, S.J., Costello, D.A.E., Moroz, I.M.: Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection. Biomedical Engineering OnLine **6**, 23 (2007)
13. William, H.W., Mangasarian, O.L.: Multisurface method of pattern separation for medical diagnosis applied to breast cytology. In: Proceedings of the National Academy of Sciences, USA, vol. 87, 9193–9196, December 1990
14. Breiman L.: Bias, variance and arcing classifiers. Technical Report TR. 400 (1996)
15. Nithya, R., Manikandan, P., Ramyachitra, D.: Performance Analysis of Meta Classifiers Algorithms using Yeast Dataset. Int. J. of Innovative Research in Comp. and Com. Eng. **3**(9), 8062–8068 (2015)
16. Kotsianti, S.B., Kanellopoulos, D.: Combining bagging, boosting and dagging for classification problems. In: Knowledge-Based Intelligent Information and Engineering Systems. Lecture Notes in Computer Science, vol. 4693, pp. 493–500 (2007)
17. Melville, P., Mooney, R.J.: Constructing diverse classifier ensembles using artificial training examples. In: Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, pp. 505–510 (2003)
18. Rodriguez, J.J., Kuncheva, L.I., Alonso, C.J.: Rotation Forest: A New Classifier Ensemble Method. IEEE Transactions on Pattern Analyses and Machine Intelligence **28**(10), 1619–1630 (2006)
19. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software: an update. SIGKDD Explor. Newsl. **11**, 10–18 (2009)
20. Powers, D.M.: Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation. Journal of Machine Learning Technologies **2**(1), 37–63 (2011)