Meta learning and Base learning on various data sets, A Comparative Study

1. **Sanjay Kumar Sen,**
Asst. Professor, Computer Sc.& Engg., Orissa Engg. College,Bhubaneswar.
2. **Prof. Dr. Sujata Dash,**
Reader, North Odisha University,Baripada.

Abstract: There are varieties of data mining algorithms and techniques available for the extraction of hidden knowledge from large data base. Classification is one of the widely used techniques in data mining. Different machine learning algorithms have been proposed for data classification. Several base classifiers are combined by meta learner whose individual performances in some way contribute to the overall classification. The premise is that meta-learning enhances the data mining task with the ability to learn and adapt from previous experience. This paper comprises empirical evaluation on the performance of two meta learning approaches on several base learning algorithm.

Keywords: Meta-learning, classification, meta-learner, base classifiers, data mining.

---

1. Introduction

A huge amount of data is being created from normal business operations in organizations which is accumulated in modern data repositories regularly. Data mining combines statistics, machine learning and database techniques to extract useful information in order to achieve the competitive advantage in business and in many organizational functions. Classification is one of the important and widely used techniques in data mining applications. Meta-learning associates multiple learning algorithms as base learners and combines their features to accumulate meta-knowledge that is useful to understand and predict the accuracy of a meta-learner (Vilalta and Drissi, 2002; Saso and Bernard, 2004; Todorovski, and Dˇzeroski, 2002). It is said that meta-learning aims at improving the performance of a learning system through previous experience. Empirical evidence claim that a meta-learner results better accuracy of base learners while removing biasness of an individual learner and retaining comprehensibility (Domingos, 1998). Several meta-learning approaches such as bagging, boosting, stacking and voting have been proposed in machine learning literature (Ting, and Witten, 1999; Bauer, and Kohavi, 1999; Bouziane, Messabih, and Chouarfia, 2011).

Understanding the performance of a learning algorithm is important to apply it to real-world problems and has been an interesting topic of research for data mining community (Dietterich, 1997). Predictive power and classification accuracy of a learning algorithm depend on several characteristics of a dataset (Neslihan and Zuhal, 2010). Consequently, meta-learning framework is also related to these characteristics. We investigate this aspect of meta-learning and report how dataset characteristics affect the classification performance of meta-learning approaches.

2. Objective

The objective of the study is aimed at performance of learning algorithms and meta-learning approaches and how they are affected by characteristics of datasets.

3. Methodology

3.1 Datasets

The datasets used in the study are taken from publicly available UCI Machine Learning Repository (Web-11, 2014). The sample is carefully selected to reflect diversity with a mixed category of statistical characteristics spread over several application domains. Table-5.1 summarizes the characteristics of the datasets.

Table-1: Dataset Characteristics

Dataset	#Instances	#Attributes	#Missing Values (%)
Iris	150	5	0
Breast cancer	286	10	0
bank	600	11	0
Credit rating	690	16	12

3.2 Algorithms

A set of five popular classification algorithms are chosen for base-level learning. They are Naïve Bayes (NB), Multi-Layer Perceptron (MLP), Sequential Minimal Optimization (SMO), IBk and J48. Similarly, two popular meta-learning approaches such as Logitech and Bagging which combine classifiers are considered for evaluation. A brief description of each of the selected algorithms is provided here.

- The Naïve Bayes (NB) is a Bayesian Network model that uses statistical approach to classify instances based on probabilities. It is robust to isolated noise points, irrelevant attributes and handles missing values.
- Multi-Layer Perceptron (MLP) is a type of Artificial Neural Network algorithm that models human brain to learn. It has high tolerance to noisy data. It is capable of handling large number of instances and different attribute types.

- Sequential Minimal Optimization (SMO) is an improved algorithm based on Support Vector Machine and is very efficient for optimization problems. It performs well with sparse data.
- J48 is an implementation of superior C4.5 algorithm. It builds decision tree from a set of labeled training data using the concept of entropy. It is capable of handling both numeric and nominal attributes as well as attributes with missing values. It performs well with nominal attributes.
- **bagging**, is a **machine learning** ensemble meta-algorithm designed to improve the stability and accuracy of**machine learning** algorithms used in statistical classification and regression. It also reduces variance and helps to avoid over fitting.
- **LogitBoost** is a boosting algorithm formulated by Jerome Friedman, Trevor Hastie, and Robert Tibshirani. The original paper casts the AdaBoost algorithm into a statistical framework.

.

4. Experiment

Weka (Web-12, 2014) has been used to conduct the experiment. We use 10-fold cross validation as the test mode to record classification accuracy. This approach is suitable to avoid biased results and provide robustness to the classification. Also, the parameters of a classification algorithm are chosen to their default values.

The following steps have been applied to generate experimental data in order to answer the stated research questions.

Step-1: Run all the selected four base learning algorithms on a dataset one after another to record each one's classification accuracy.

Step-2: Run the meta-learning algorithm on the data set to record its classification accuracy.

5. Result Analysis

Classification accuracy of different classifiers on the sample datasets is provided in Table-1. Based on the data captured from UCI data repository. It is clear that the same classifier is not always the best one for all datasets. Different classifiers perform differently and the best one for each dataset is different.

Table-2 (META-LEARNING Vs Base Learning) Classification Accuracy

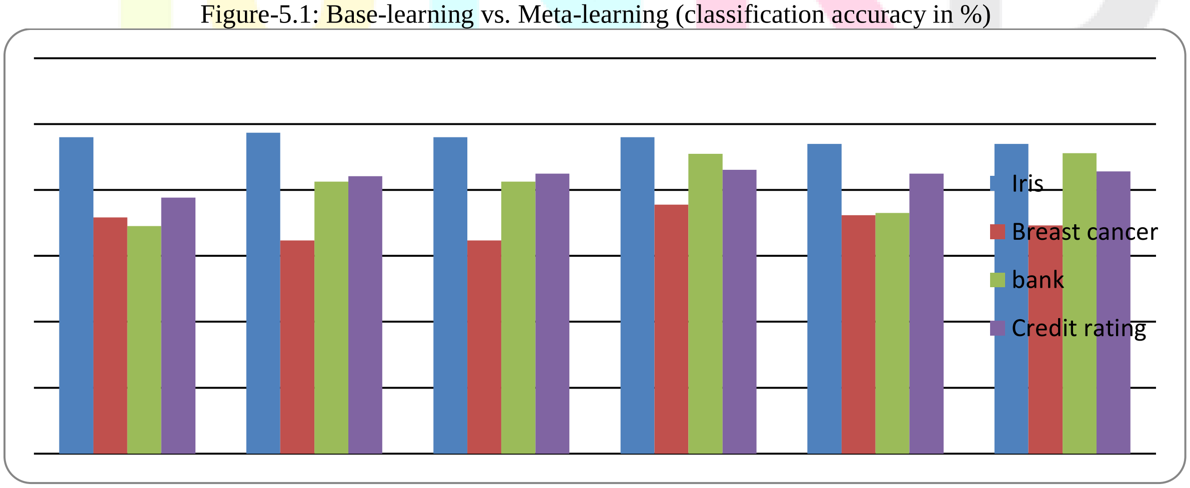
Data-sets	**Base Classifiers**				**Meta-classifier**	
	NB	**MLP**	**SMO**	**J48**	**Logitboost**	**Bagging**
Iris	96	97.3333	96	96	94	94
Breast cancer	71.6783	*64.6853*	64.6853	75.5245	72.3776	69.2308
bank	69	82.5	82.5	91	73	91.1667
Credit rating	77.6813	84.2029	84.9275	86.087	84.9275	85.6522

Table-3: Dataset and its Best Classifier

Dataset	**Best Classifier**
Iris	MLP
Breast cancer	J48
bank	bagging
Credit rating	J48

Table-2 and its graphical presentation in Figure-1 show that classification performed by meta-learning approaches is poor than or at best the same as the classification performed by the best base-classifier for a given dataset. In most of the cases, the best base-classifier provides better classification accuracy. So it is concluded that meta-learning does not provide better accuracy that that of the best base-classifier. However, meta-learning provides improved accuracy for the meta-classifier as compared to its performance as a base-classifier.

Figure-5.1: Base-learning vs. Meta-learning (classification accuracy in %)



It is found from Table-2 and its graphical presentation in Figure-5.1 that meta-learning offers satisfactory performance over base-learning when attributes of the dataset are of nominal type. Compared to stacking, voting offers slightly better performance with this characteristic of a dataset.

Figure-5.2: Screen shot of Meta-Learning process

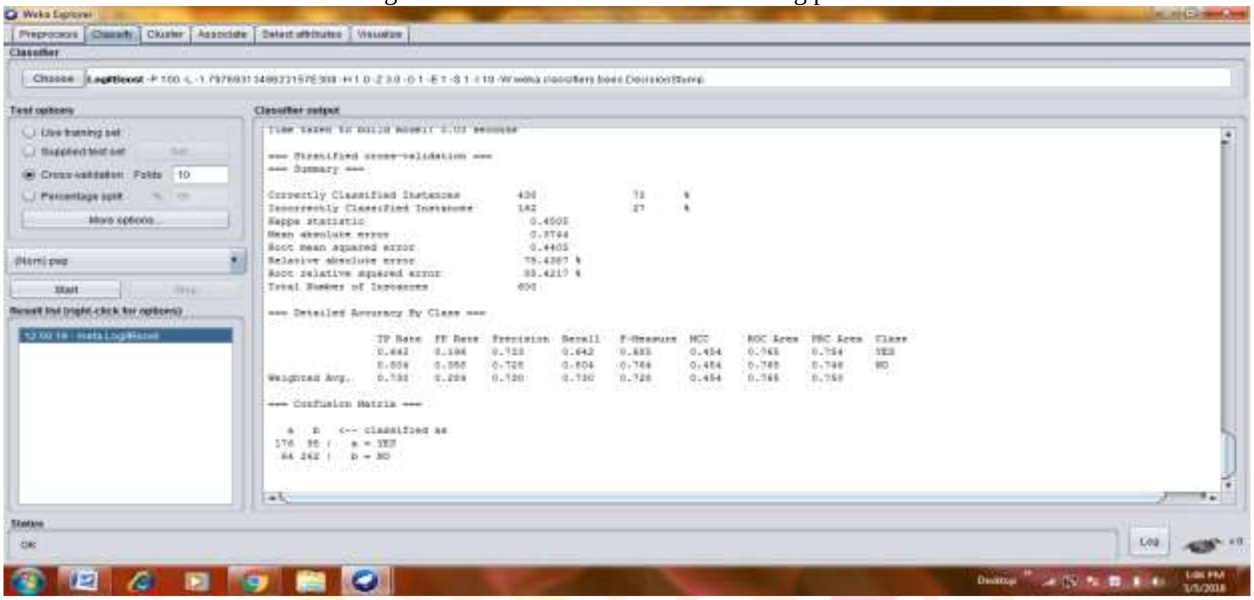
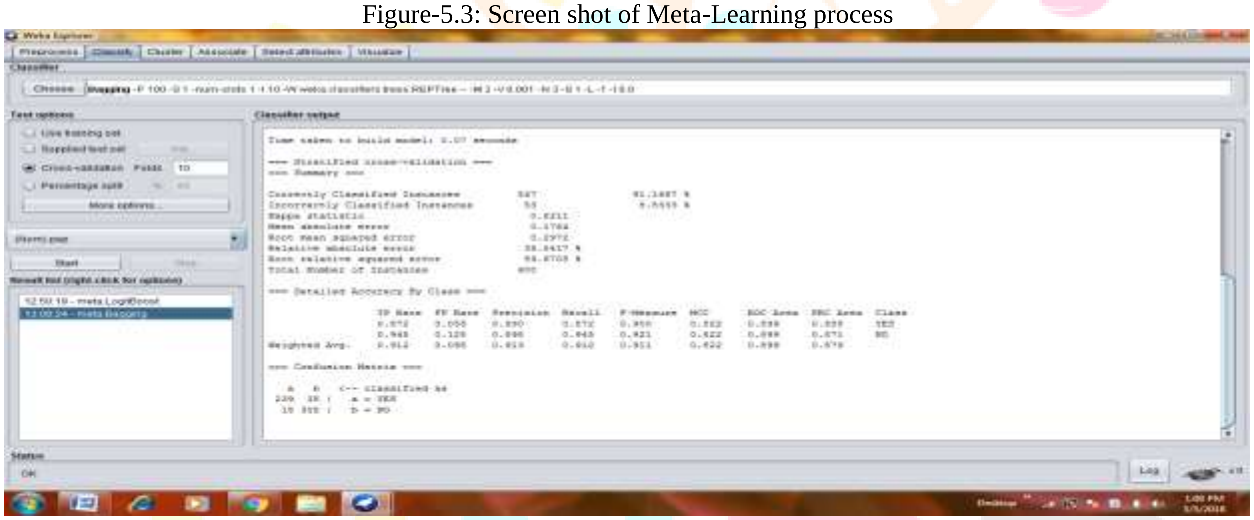


Figure-5.3: Screen shot of Meta-Learning process



6. Conclusion

In this study, four base classifiers and two meta-learning approaches are considered. Their classification performance data on four publicly available datasets is recorded to answer the research questions formulated. Based on the finding, it can be said that characteristics such as attribute type and class size of a dataset do influence the classification accuracy. However, there is no strong relationship between them. Again the relationship is not completely linear as evident from the correlation measures. It is also observed that the classification accuracy values deviate across different datasets. Consequently, neither base-learning algorithm nor meta-learning approach is the best choice for all datasets. It can be further said that while meta-learning improves accuracy performance over base-learning, the empirical data obtained in this study could not provide a strong support to it.

7. Reference

[1] Vilalta R. and Drissi Y. (2002), "A Perspective View and Survey of Meta-Learning", *Journal of Artificial Intelligence Review*, 18 (2), pp.77-95.

[2] Domingos Pedro (1998), "Knowledge Discovery via Multiple Models", *Intelligent Data Analysis*, 2, pp.187-202.

[3] Ting, K. M., and Witten, I. H. (1999), "Issues in stacked generalization", *Journal of Artificial Intelligence Research, 10*, pp.271–289.

[4] Dietterich, T. G. (1997), "Machine-learning research: Four current directions", *AI Magazine, 18(4)*, 97–136.

[5] Neslihan D, and Zuhal T (2010), "A Comparative Framework for Evaluating Classification Algorithms", in Proceedings of the World Congress on Engineering (WCE 2010) , Vol-I, June 30- July 2, London, U.K.

[6] Neslihan D, and Zuhal T (2010), "A Comparative Framework for Evaluating Classification Algorithms", in Proceedings of the World Congress on Engineering (WCE 2010) , Vol-I, June 30- July 2, London, U.K.