Volume 3, Issue 3, March 2013 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

A Comparative Study of Meta Classifier Algorithms on Multiple Datasets

1.P.KALAISELVI
PG Student Bharath University
Department of Computer Science and Engineering, India.

2.Dr.C.NALINI
Department of Computer Science and Engineering Bharath University
India.

Abstract- Data mining is the process of analyzing data from different perspectives and summarizing it into useful information. While the procedures for comparison of a pair of classifiers on a single dataset have been proposed almost a decade ago, comparative studies with more classifiers and/or more data sets still employ partial and unsatisfactory solutions. In this paper, Robot Navigation dataset is analyzed using WEKA data mining tool to explore the efficient classifier to find the high accuracy. Various classification algorithms are applied over multiple datasets and finally, based on high accuracy the best classifier is identified.

Keywords: Data Mining, MetaClassifier, MultiClass, Efficient, Accuracy.

1. Introduction

Data Mining is a process of Semi-Automatically Analyzing large database to find the patterns and data. Data Mining is closely related to Knowledge Discovery. Data Mining basic techniques are Clustering, Association rule discovery, Classification , Sequential pattern discovery and the Distributed data mining Techniques are Distributed Classifier learning , collective data mining , distributed clustering and others. The main goal of data mining is Predictive and Descriptive. Descriptive data mining provides information to understand what it is happening inside the data without a predetermined idea. Predictive data mining its provides the user to submit records with unknown field values. Classification and Prediction are prediction model and clustering and association rules and frequent item set is Descriptive model. The objective of this paper is to apply meta learning techniques and to provide a comprehensive review of different classification techniques in Meta classification. The number of cases classified correctly provides us with an estimate of the accuracy of the model. our aim is to find highly accurate models that are easy to understand and achieve efficiency when dealing with large Datasets and multiple dataset[1].

In the second section are described about related work, and the section III described about different techniques used in this paper. The IV section described about Performance Evaluation and the V section described about Experiment setup and result and

The last section about conclusion and future enhancement

2. Related Work

A supervised machine learning task involves constructing a mapping from input data to the appropriate outputs. In a classification learning task, each output is one or more classes to which the input belongs. The goal of classification learning is to develop a model that separates the data into the different classes, with the aim of classifying new examples in the future. For example, a credit card company may develop a model that separates people who defaulted on their credit cards from those who did not based on other known information such as annual income. The goal would be to predict whether a new credit card applicant is likely to default on his credit card and thereby decide whether to approve or deny this applicant a new card. In a regression learning task, each output is a continuous value to be predicted .Many traditional machine learning algorithms generate a single model (e.g., a decision tree or neural network). Ensemble learning methods instead generate multiple models. Given a new example, the ensemble passes it to each of its multiple base models, obtains their predictions, and then combines them in some appropriate manner (e.g., averaging or voting). As mentioned earlier, it is important to have base models that are competent but also complementary. To further motivate this point, consider This figure depicts a classification problem in which the goal is to separate the points marked with plus signs from points marked with minus signs. None of the three individual linear classifiers (marked A, B, and C) is able to separate the two classes of points. However, a majority vote over all three linear classifiers yields the piecewise linear classifier shown as a thick line. This classifier is able to separate the two classes perfectly. For example, the plusses at the top of the figure are correctly classified by A and B, but are misclassified by C. The majority vote over these correctly classifies these points as plusses. This happens because A and B are very different from C. If our ensemble instead consisted of three copies of C, then all three classifiers would misclassify the plusses at the top of the figure, and so would a majority vote over these classifiers.

3. Various Techniques

The following methods are used to improve the high accuracy .Each model combines a series of K learning models. The models are used in this paper Bagging, Dagging, Decorate,
MultiClassClassifier, MultiboostAB.

3.1Bagging
Bootstrap Aggregating Bagging generates multiple bootstrap training sets from the original training set using sampling with replacement and uses each of them to generate a classifier for inclusion in the ensemble. Given a set, D, of tuples, bagging works as follows. For iteration I (i = 1,2…k), a training set, Di, of d tuples is sampled with replacement from the original set of tuples, D. Note that the term bagging stands for bootstrap aggregation. Each training set is a bootstrap sample. Because sampling with replacement is used, some of the original tuples of D may not be included in Di, whereas others may occur more than once. A classifier model, Mi., is learned for each training set, Di. To classify an unknown tuple, X, each classifier, Mi, returns its class prediction, which counts as one vote. The bagging can be applied to the prediction of continuous values by taking the average value of each prediction for a give test tuple. The bagged classifier often has significantly greater accuracy than a single classifier derived from D, the original training data. It will not be considerably worse and is more robust to the effects of noisy data. The increased accuracy occurs because the composite model reduces the variance of the individual classifiers. For prediction, it was theoretically proven that a bagged predictor will always have improved accuracy over a single predictor derived from [7].

3.2 Boosting
Bootstrap Aggregating Bagging generates multiple bootstrap training sets from the original training set using sampling with replacement and uses each of them to generate a classifier for inclusion in the ensemble. In boosting, weights are assigned to each training tuple. A series of k classifiers is iteratively learned. After a classifier Mi is learned, the weights are updated to allow the subsequent classifier, Mi+1, to pay more attention to the training tuples that were misclassified by Mi. The final boosted classifier, M*, combines the votes of each individual classifier, where the weight of each classifier's vote is a function of its accuracy. The boosting algorithm can be extended for the prediction of continuous values.

3.3 Dagging
This Meta classifier creates a number of disjoint, stratified folds out of the data and feeds each chunk of data to a copy of the supplied base classifier. Predictions are made via majority vote, since all the generated base classifiers are put into the Vote meta classifier. Useful for base classifiers that are quadratic or worse in time behavior, regarding number of instances in the training data.

3.4 Decorate
DECORATE (Diverse Ensemble Creation by Oppositional, Relabeling of Artificial Training Examples) is presented that uses a learner (one that provides high accuracy on the training data) to build a diverse committee. This is accomplished by adding different randomly constructed examples to the training set when building new committee members. These
Artificially constructed examples are given category labels that disagree with the current decision of the committee, thereby directly increasing diversity when a new classifier is trained on the augmented data and added to the committee.

3.5 MultiBoosting
MultiBoosting is another classifier method with same category that can be considered as Wagging. Wagging is a variant of Bagging. Bagging uses resampling to get the datasets for training and producing weak hypothesis whereas Wagging uses reweighting for each training example, pursuing the effect of bagging in a different way.

3.6 Adaboost
Adaboost is a popular boosting algorithm. Suppose We would like to boost the accuracy of some learning method. WE are given D, a data set of d class –labeled tuples,(x1,y1),(x2,y2),…..,(xd,yd), where yi is the class label of tuple xi. Initially, adaboost assigned each training tuple an weight of 1/d . Generating k classifiers for the ensemble requires K rounds through the rest of the algorithm .In round i, the tuple from D are sampled to form a triangle set ,Di of size d. sampling with replacement is used –the same tuple may be selected more ,Mi is derived from the training tuple are then Adjusted according to how they were classified. If a tuple was incorrectly classified its weight is increase. If a tuple was correctly classified its weight is decrease .A tuple weight reflect how hard it is to classify-the higher the weight, the more often it has been misclassified, These weights will be used to generated of the next round. The basic idea is that when we build a classifier, we

want it to focus more on the misclassified tuples of the previous round. Some classifiers may be better at classifying some hard tuples than others. In this way, we build a series of classifiers that complement each other. Once the weights of all of the correctly classified tuples

are updated, the weights for all tuples are normalized so that their sum remains the same as it was before. To normalize a weight, we multiply it by the sum of the old weights, divided by the sum of the new weights. As a result, the weights of misclassified tuples are increased and the weights of correctly classified tuples are decreased, as described above. "Once boosting is complete, the ensemble of classifiers used to predict the class label of a tuple, X Unlike bagging, where each classifier was assigned an equal vote, For each class, c, we sum the weights of each classifier that assigned class c to X. The class with the highest sum is the "winner" and is returned as the class prediction for tuple X

4. Performance Evaluation

4.1DATASET:

4.1.1Data set Description

The Wall Following Robot Navigation dataset collected from UCI repository. The data were collected as the SCITOS G5 robot navigates through the room following the wall in a clockwise direction, for 4 rounds, using 24 ultrasound sensors arranged circularly around its 'waist'. The provided files comprise three different data sets. The first one contains the raw values of the measurements of all 24 ultrasound sensors and the corresponding class label Sensor readings are sampled at a rate of 9 samples per second. The second one contains four sensor readings named 'simplified distances' and the corresponding class label. These simplified distances are referred to as the 'front distance', 'left distance', 'right distance' and 'back distance'. They consist, respectively, of the minimum sensor readings among those within 60 degree arcs located at the front, left, right and back parts of the robot. The third one contains only the front and left simplified distances and the corresponding class label. Files with different number of sensor readings were built in order to evaluate the performance of the classifiers with respect to the number of inputs.

Attribute Information:

4.2 NumberofAttributes

1. Sensor readings 24.data: 24numeric attributes and the class.
2. Sensor readings 4.data: 4numeric attributes and the class.
3. Sensor readings 2.data: 2numeric attributes and the class.

In the first dataset have 24 numeric attributes, each attribute represent in angle and it has four classes Move Forward, Slight Right Turn, Sharp Right Turn, Slight Left Turn .In the second dataset have 4 numeric attributes, these attributes selection is based on 60 degree

1. SD front: minimum sensor reading within a 60 degree arc located at the front of the robot - (numeric: real)
2. SD left: minimum sensor reading within a 60 degree arc located at the left of the robot - (numeric: real)
3. SD right: minimum sensor reading within a 60 degree arc located at the right of the robot - (numeric: real)
4. SD back: minimum sensor reading within a 60 degree arc located at the back of the robot - (numeric: real) and the class are used Move Forward, Slight Right Turn, Sharp Right Turn, Slight Left Turn .In the third dataset two numeric attributes are selected,

1. SD front: minimum sensor reading within a 60 degree arc located at the front of the robot - (numeric: real)
2. SD left: minimum sensor reading within a 60 degree arc located at the left of the robot - (numeric: real) and class: Move-Forward, Slight-Right-Turn Sharp-Right-Turn Slight-Left-Turn

The performance of the classifiers depends on the characteristics of the data to be classified. The Random sub-sampling, k-fold cross validation and bootstrap method. In our study, we have selected k-fold cross validation for evaluating the classifiers. In k-fold cross validation, the initial data are randomly partitioned into k mutually exclusive subset or folds d1,d2,…,dk, each approximately equal in size. The training and testing is performed k times. In the first iteration, subsets d2… dk collectively serve as the training set in order to obtain a first model, which is tested on d1; the second iteration is trained in subsets d1, d3,…, dk and tested on d2; and so no[2].

WEKA 3.6.5 tool is used to study the Performance of the chosen algorithms and the results are used to measure the Accuracy and Time from the confusion matrix 2x2 obtained[1].

5. Experiment Setup And Result

Original Attribute:

L={A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y}

List of usage of Attributes 90%---100%.A-X represents the degree of angle and Y represent the class of these angles.

.

Table 1: Shows the Selected Attribute

S.no	Algorithm Name	Selected Attributed	Subset Name
1	BestFirst	M,N,O,R,S,T	S1
2	ExhaustiveSearch	M,N,O,R,S,T	S2
3	GeneticSearch	M,N,O,R,S,T	S2

Table 2: Shows the Selected Attributes with Classification algorithms for subset S1

Sno	Algorithm name	Nominal class classifier	Accuracy	Time to built
1.	Bagging	REPTree	99.2852 %	0.68 seconds
2.	Dagging	SMO	57.313 %	2.77 seconds
3.	Decorate	J48	69.8864 %	0.11 seconds
4.	MultiClassClassifier	Logistic	63.2148 %	0.98 seconds
5.	MultiBoostAB	Decision Stump	69.8864 %	0.1 seconds

Figure 1 shows the graphical representation of difference in Accuracy

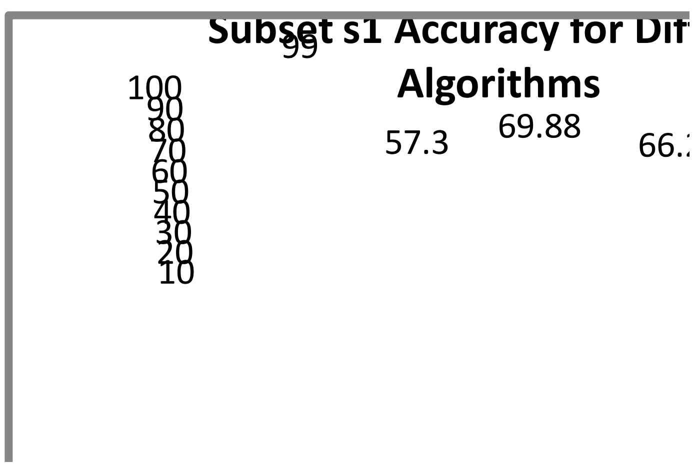

Table 3: Shows the Selected Attributes with Classification algorithms for subset S2

Sno	Algorithm name	Nominal class classifier	Accuracy	Time to built
1.	Bagging	REPTree	99.2852 %	0.47 seconds
2.	Dagging	SMO	57.313	1.93 seconds
3.	Decorate	J48	99.0652	10.22
4.	MultiClassClassifier	Logistic	66.2148	0.76
5.	MultiBoostAB	Decision Stump	69.8864	0.06

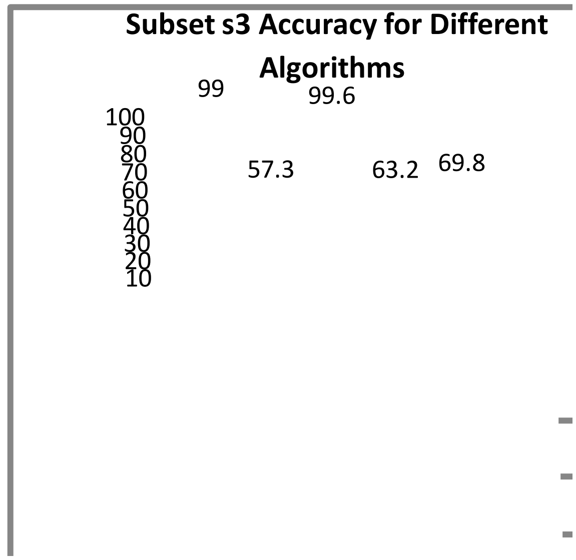

Figure2 shows the graphical representation of difference in Accuracy

Table 4: Shows the Selected Attributes with Classification algorithms for subset S3

Sno	Algorithm name	Nominal class classifier	Accuracy	Time to built
1.	Bagging	REPTree	99.2852 %	0.47 seconds
2.	Dagging	SMO	57.313	1.93 seconds
3.	Decorate	J48	99.0652	10.22
4.	MultiClassClassifier	Logistic	66.2148	0.76
5.	MultiBoostAB	Decision Stump	69.8864	0.06

Figure 3 shows the graphical representation of difference in Accuracy

6. Conclusion And Future Enchanesment

In this paper, the performance of the different classifier methods like Bagging, Dagging, Decorate, MultiClassClassifier, and MultiboostAB Are compared. In this Bagging is best algorithm to finding the accuracy. In these Experiment Robot Navigation datasets are used. Classification Accuracy and Time is Calculated by 10-fold validation method. In future we may conduct the same experiments with different datasets instead of multiple dataset, MULTICLASS and combine few ensemblers with the different base classifier to study how the ensemblers combined with the base classifiers boost the performance accuracy could.

ACKNOWLEDGEMENT

The authors would like to thank the management of Bharath University for the support and encouragement for this research work

REFERENCES

[1] R.MahaLakshmi, Research Scholar, Manonmanium University, Tirunelveli, Tamil Nadu, India
[2] Suresh Subramanian,Research Scholar , Karpagam University Coimbatore, Tamil Nadu, India Spam Email classification based on machine learning algorithms
[3] W.Frawley and G. Piatetsky-Shapiro andC.Matheus, Knowledge Discovery in Databases: An Overview. Magazine, Fall 1992, pgs 213-228
[4] .Thair Nu PhyuProceedings of the International MultiConference of Engineers and Computer Scientists2009 Vol I IMECS 2009, March 18 - 20, 2009, Hong Kong

[5] E. Alpaydin, Introduction to machine learning. London: The MIT Press, 2004.
[6] R. O. Duda, P. E. Hart, and D. G. Stork, Pattern classification, 2 ed. New York: John Wiley & Sons, Inc.,2001.
[7] Group-based Meta-classification Noor A. Samsudin, Andrew P. Bradley,School of Information Technology and Electrical Engineering The University of Queensland, Australia
[8] S A Robust Meta-Classification Strategy for Cance Diagnosis from Gene Expression Data Gabriela Alexe, Gyan Bhanot, IBM Computational Biology Center, IBM T.J.Watson Research.
[9] Ms. Bhoomi Trivedi,Ms. Neha Kapadia,INDUS institute of Eng. & Tech,TCET, Kandivali(E),Ahmedabad Modified Stacked generalization withSequential Learning. TCET2012 on IJCA
.[10] .J.R. Quinlan, “Induction of decision trees,” In Jude W.Shavlik, Thomas G. Dietterich, (Eds.), Readings in Machine Learning. Morgan Kaufmann, 1990. Originally published in Machine Learning, vol. 1, 1986, pp 81–106.
[11] Salvatore Ruggieri,”Efficient C4.5 Proceedings of IEEE transactions on knowledge and data engineering”, Vo1. 14, 2, No.2, PP.438-444, 20025
[12] P. Domingos, M. Pazzani, On the optimality of the simple Bayesian classifier under Zero-one loss, Machine learning 29(2-3)(1997) 103-130.11

