A Survey & Current Research Challenges in Meta Learning Approaches based on Dataset Characteristics

Nikita Bhatt, Amit Thakkar, Amit Ganatra

Abstract— Classification is a process that predicts class of objects whose class label is unknown. According to No Free Lunch (NFL) theorem, there is no single classifier that performs better on all datasets. Meta learning is one of the approaches that acquired knowledge based on the past experience. The knowledge in Meta-Learning is acquired from a set of meta-examples which stores the features of the problem and the performance obtained by executing a set of candidate algorithms on Meta Features. Based on the experience acquired by the system during training phase, ranking of the classifiers is provided based on considering various measures of classifiers.

Index Terms—***Classification, Meta Learning, Ranking***

I. INTRODUCTION

Data Mining is a process that extracts patterns from the large datasets. There are major research areas in Data Mining including association mining, clustering, classification, web mining, text mining, etc. Classification is one of the techniques in Data Mining that solves various problems like algorithm selection, model comparison, division of training and testing data, preprocessing, etc. It is a two- step process.1) Build classification model using training data. Every object of the data must be pre-classified i.e. its class label must be known.2) The model generated in the preceding step is tested by assigning class labels to data objects in a test dataset. The test data is different from the training data. Every element of test data is also pre classified in advance. The accuracy of the classification model is determined by comparing true class labels in the testing set with those assigned by the model [15].

Classification is important when a data repository contains samples that can be used as the basis for future decision making. Machine learning researchers have proposed many different types of classification algorithms, including nearest-neighbor methods, decision tree induction, error back propagation, reinforcement learning, lazy learning, rule-based learning, statistical learning, etc. [9].The selection most adequate algorithmfor a new problem is a difficult task as classification algorithms are originated from different-areas like statistics, machine learning, and neural network with considerably different performance [26].

Algorithm selection is a time consuming task which involves experimentation with different classifiers and analyzing the performance of those classifiers [8].Apart from that,NFL (No Free Lunch) Theorem states that if algorithm A outperforms algorithm B on some cost functions, then there must exist exactly as many other functions where B outperforms A [23]. In other words, it is impossible to build an algorithm that performs optimally for all tasks [25]. As a consequence, it becomes important for researchers and practitioners to discover and implement mechanisms that may determine which machine learning algorithm perform best on which tasks [4].

In deciding which classifier will work best for a given dataset, there are two options. The first is to put all the trust in an expert's opinion based on knowledge and experience. Thesecond is to run through every possible classifier that could work on the dataset, identifying rationally the one which performs best. The latter option, while being the most thorough, would take time and require a significant amount of resources, especially with larger datasets, and as such is impractical. If the expert consistently chooses an ineffective classifier, the most effective classification rules will never be learned, and resources will be wasted. Neither methods, provides an effective solution and as a result it would be extremely helpful to both users and experts, if it were known explicitly which classifier, of the multitude available, is most effective for a particular type of the dataset [7].

Meta learning is a framework developed in the field of supervised machine learning with the aim of automatically predicting algorithm performance thus assisting users in the process of algorithm selection [2, 16].In Meta learning, knowledge is acquired by the meta-examples that store,
(a) The features that describe the dataset (problem).
(b)Performance information obtained by executing candidate algorithms on training datasets. After generation of meta-examples, Meta learner (learning algorithm) is applied to acquire knowledge that relates performance of candidate algorithms to the features of the datasets (problems). As it is usually difficult to identify a single best algorithm reliably, it is good alternative to provide ranking [2].

II. CLASSIFICATION MEASURES

There are various measures to evaluate the performance of the classifier. Table I shows various measures of the classifiers.

Manuscript Received February 2012

Nikita Bhatt, Department of Computer Engineering, Charotar Institute of Technology (Faculty of Technology and Engineering), Charotar UniversityofTechnologyChanga,Anand, Gujarat, India, 388421 nikitabhatt.ce@ecchanga.ac.in

Amit Thakkar, Department of Information Technology, Charotar Institute of Technology (Faculty of Technology and Engineering), Charotar UniversityofTechnologyChanga,Anand, Gujarat, India, 388421 amitthakkar.it@ecchanga.ac.in

Amit Ganatra, Department of Computer Engineering, Charotar Institute of Technology (Faculty of Technology and Engineering), Charotar UniversityofTechnologyChanga,Anand, Gujarat, India, 388421 amitganatra.ce@ecchanga.ac.in

Table I Classification Measures

Measures	Description	Importance
Precision	Precision can be seen as a measure of exactness or quality.	Precision is Used to retrieved fraction of instances those are relevant.
Recall	Recall is a measure of completeness or quantity.	Recall is used to retrieve fraction of relevant instances that are retrieved.
Accuracy	The accuracyis the proportion of the total number of predictions that were correct. Accuracy is related to the degree of bias in the measurements	Accuracy is used to represent the correct answer or percentage of accurate classification.
ROC (Receive Operating Curve)	A ROC graph is a plot with the false positive rate on the X axis and the true positive rate on the Y axis. An ROC curve or point is independent of class distribution or error costs.	ROC curves is used to provide a visual tool for examining the tradeoff between the ability of a classifier to correctly identify positive cases and the number of negative cases that are incorrectly classified.
ARR (Adjusted Ratio of Ratio)	ARR is a method based on success rate ratio and an adjusted time ratio.	ARR is used to find the overall ratio of success rate.
MeanAbsolute Error	It indicates how close prediction matches to eventual outcomes	It is used to find the closeness of prediction with outcomes

III. APPROACHES FOR THE ALGORITHM SELECTION PROBLEM

Algorithm selection is one of the difficult problems in classification. Table II shows the approaches to solve the problem along with its description.

IV. WORKING OF META LEARNING

In classification, to solve algorithm selection problem various approaches are available and a lot of research work is carried outin that direction. Meta-learning is currently hot research topic in machine learning, which has emerged from the need to improve the generalization ability and stability of the learned models and support data mining automation in issues related to algorithm and parameter selection [8]. It is the process of generating knowledge that relates the performance of machine learning algorithms to the characteristics of the problem (i.e., characteristics of its datasets) [3].

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. Meta-learning studies how to choose the right bias dynamically,as opposed to base-learning where the bias is fixed a priori, or user parameterized [14].

Table II Approaches for algorithm selection

Approach	Description
Trial and Error Approach	Available classifiers are applied on datasets. Suppose we have n classifiers and m datasets, this procedure require O (nm) according to graph theory which is costly process.
Expert Advice	When any new dataset comes, we take advice from the expert which is not always easy to acquire.
Proposed Framework	Authors have proposed framework which is restricted to model of classifier. Performance of classifier is evaluated for limited number of datasets and classifiers.
Meta Learning	Meta Learning is the study of principled methods that exploit Meta knowledge to obtain efficient models and solutions by adapting machine learning and data mining processes.

A. Architecture of Meta Learning

Meta Learning System can be divided into two modes:
1) Acquisition Mode
2) Advisory Mode

Acquisition Mode:

During the knowledge acquisition mode, the main goal is to learn about the learning process itself. We assume that the input to the system consists of datasets of examples. Upon arrival of each dataset, the meta-learning system invokes a component responsible for extracting dataset characteristics or Meta features. The goal of this component is to gather information that transcends a particular domain of application. During the knowledge acquisition mode, the learning techniques do not exploit knowledge of previous results. Statistics derived from different learning strategies (e.g., a classifier or combination of classifiers, and their performance) may be used as a form of characterizing the task under analysis. Information derived from the meta-feature generator and the performance evaluation module can be combinedinto a meta-knowledge base. This knowledge base is the main result of the knowledge acquisition phase [14]; it reflects experience accumulated across different tasks. Fig 1(a) shows the general structure of the acquisition mode.

Advisory Mode:

In the advisory mode, meta-knowledge acquired in the exploratory mode is used to configure the learning system in a manner that exploits the characteristics of the new data. Meta-features extracted from the dataset are “matched” with the meta-knowledge base to produce a recommendation regarding the best available learning strategy [14]. Fig 1(b) shows the general structure of the advisory mode.

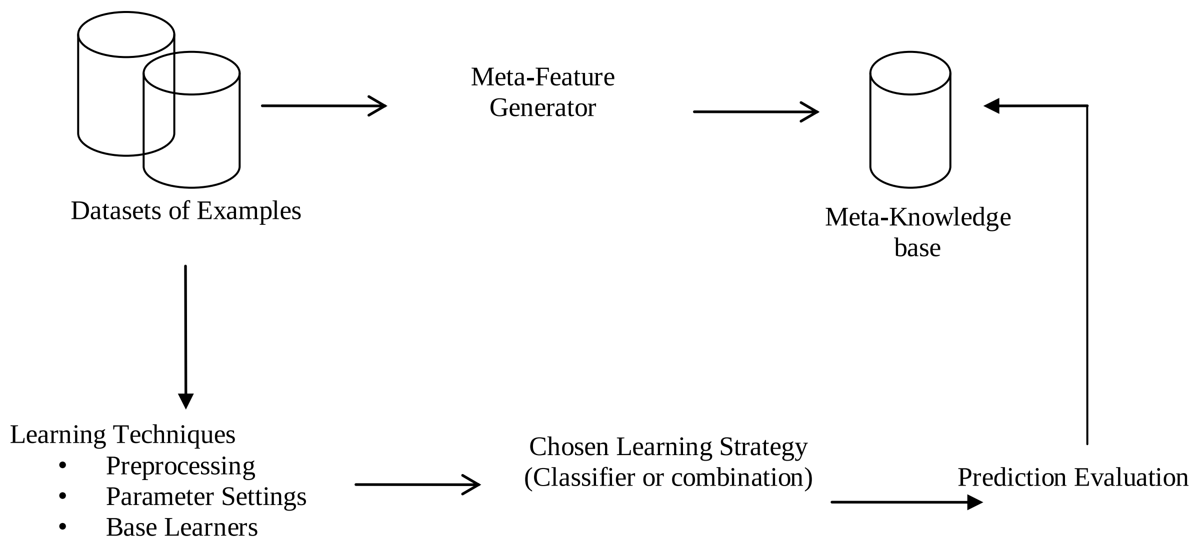


Fig 1(a) Acquisition Mode

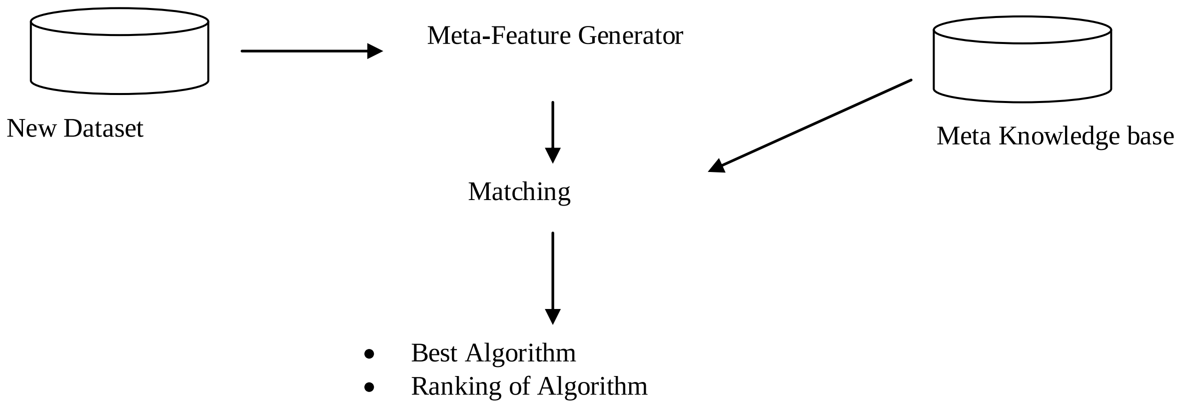


Fig 1(b) Advisory Mode

B .Process of Meta Learning

The meta-learner is a learning system that receives as input a set of such meta-examples and then acquires knowledge used to predict the algorithms performance for new problems being solved. The meta-features are, in general, statistics describing the training dataset of the problem, such as number of training examples, number of attributes, correlation between attributes, class entropy, among others [24].

In Meta-Learning, each meta-example stores, as performance information, a class attribute which indicates the best algorithm for the problem, among a set of candidates. In this case, the class label for each meta-example is defined by performing a cross-validation experiment using the available dataset. The meta-learner is simply a classifier which predicts the best algorithm based on the meta-features of the problem [20].

Meta Learning process is specified in fig 2. A database is created with meta-data descriptions of a set of datasets. These meta-data contain estimates of the performance of a set of candidate algorithms on those datasets as well as some Meta features describing their characteristics. A machine learning algorithm is applied to this database to induce the model that relates the value of the Meta features to the performance of the candidate algorithms [21].

C .Meta Features

The goal of Meta learning is to relate the performance of learning algorithms to data characteristics, i.e. Meta features. Therefore, it is necessary to compute measures from the data that are good predictors of the relative performance of algorithms.

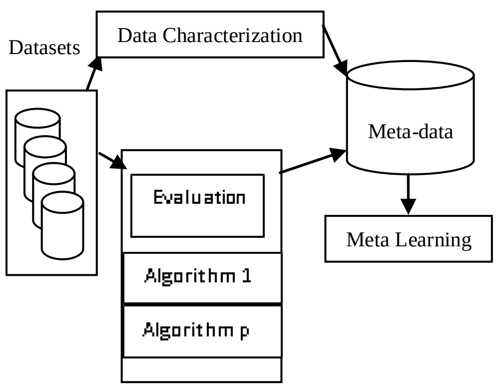


Fig. 2. Process of Meta Learning

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

- **Simple, Statistical and Information-theoretic**

Three different approaches to data characterization can be identified, namely simple, statistical and information-theoretic measures, Land markers and model-based measures.

There are two main directions used so far in order to characterize a dataset for providing suggestion as to which classification algorithm(s) is(are) more appropriate for a specific dataset. In the first one, measures that describe statistical and information based properties of the datasets used. In the second one a dataset is described using the performance of very simple learners. In a very successful metaphor the first category of measures is described as the genotype of the datasets ,i.e. the inner structure of the dataset, and the second category as the phenotype of the datasets ,i.e. the visible properties of the dataset produced by the interaction of its genotype with the environment in case the simple learners [13].

(1) Genotype of dataset

Simple Characteristics

i. No of Instances
ii. No of attributes
iii. No of classes
iv. No of binary attributes

Statistical Characteristics

v. Standard Deviation Ratio
vi. Mean absolute correlation attributes
vii. First canonical correlation
viii. Mean skwness of attributes
ix. Mean Kurtosis of attributes

(2) Phenotype of dataset

i. Entropy of class
ii. Mean entropy of attributes
iii. Equivalent number of attributes
iv. Noise-signal ratio

- **Model based**

In Model based approach, a model is induced from the data and the Meta features are based on properties of that model. An example of a model-based data characteristic is the number of leaf nodes in a decision tree. Meta features obtained using this approach is only useful for algorithm recommendation if the induction of the model is sufficiently fast.

In the first approach, consisting of simple, statistical and information-theoretic measures, the Meta features are computed directly on the dataset. In model-based data characterization, they are obtained indirectly through a model. If this model can be related to the candidate algorithms, then these approaches provide useful Meta features.

- **Land markers**

Land markers are quick estimates of algorithm performance on a given dataset. They can be obtained in two different ways. The estimates can be obtained by running simplified versions of the algorithms. An alternative way of obtaining quick performance estimates is to run the algorithms whose performance we wish to estimate on a sample of the data, obtaining the so-called sub sampling Land markers.

Like model-based meta-features, Land markers characterize the dataset indirectly. But they go one further, by representing the performance of a model on a sample of the data, rather than representing properties of the model. If the performance of the Land markers is, in fact, related to the performance of the base-algorithms, we can expect this approach to be more successful than the previous ones.

V. APPROACHES OF META LEARNING

1. In [22] for labeling meta-examples, initially 20 algorithms were evaluated through cross-validation on 22 classification problems. For each algorithm, the authors generated a set of meta-examples, each one associated either to the class label applicable or to the class label non-applicable. The class label applicable was assigned when the classification error obtained by the algorithm fell within a pre-defined confidence interval, and non-applicable was assigned otherwise. Each problem was described by a set of 16 meta-features and, finally, a decision tree was induced to predict the applicability of the candidate algorithms.

2. In [13], the authors performed the labeling of Meta-examples by deploying clustering algorithms. For labeling of meta-examples, initially the error rates of 10 algorithms were estimated for 80 classification problems. From this evaluation, a matrix of dimension 80 X 10 is generated, in which each row stored the ranks obtained by the algorithms in a single problem. The matrix was given as input to a clustering algorithm, aiming to identify groups (clusters) of problems in which the algorithms obtained specific patterns of performance. The meta-examples were then associated to the class labels corresponding to the identified clusters. Hence, instead of only predicting the best algorithm or the applicability of algorithms, the metal earner can predict more complex patterns of relative performance.

3. In the Zoomed-Ranking approach [17], instance-based learning is used to produce rankings of algorithms taking into account accuracy and execution time. In this approach, each meta-example stores the meta-features describing a learning problem, as well as the accuracy and execution timeobtained by each candidate algorithm in the problem. Given a new learning problem, the Zoomed-Ranking retrieves the most similar past problem based on the similarity of meta-features. The ranking of algorithms is then recommended for the new problem by deploying a multi-criteria measure that aggregates the total accuracy and execution time obtained by the algorithms in the similar problems.

4. The Land marking approach [26] tries to relate the performance of the candidate algorithms to the performance obtained by simpler and faster designed learners, called Land markers. This approach claims that some widely used meta-features are very time consuming, and hence, land marking would be an economic approach to the characterization of learning problems and to provide useful information for the Meta-Learning process.

5. In [12], a set of different meta-learners is used not only to predict a class label associated to algorithm performance, but also to recommend a ranking of algorithms. In this approach, a strict meta-learner is built for each different pair (X, Y) of algorithms. Given a new learning problem, the outputs of the meta-learners are collected and then pints are credited to the algorithms according to

the output. For instance if 'X' is the output of meta-learner (X,Y) then the algorithm X is credited with one point. The ranking of algorithms is recommended for the new problem directly from the member of points assigned to the algorithms.

Apart from above approaches, some other approaches are there in which research work is carried out. Table III shows various approaches of Meta learning for the algorithm selection problem.

Table III Meta Learning Approaches

YEAR	APPROACH	DESCRIPTION
2000	A1	Intelligent assistant, NOEMON for collection of dataset
2002	A2	Zooming and Ranking of Meta Learning
2002	A3	Meta-learning with incremental learning using K-nn
2003	A4	Decision Support System approach
2005	A5	Combing dataset characterization with land marking
2008	A6	Meta-learning for unsupervised learning
2008	A7	Software engineering concepts of quality attributes and metrics
2009	A8	Resampling based ensemble methods and meta-learning
2009	A9	Empirical framework that quantitatively assesses the accuracy of selection of best Bayesian classifiers
2010	A10	Active Meta Learning
2011	A11	Clustering-based meta learning
2011	A12	Active Meta Learning based on Uncertainty Sampling Method

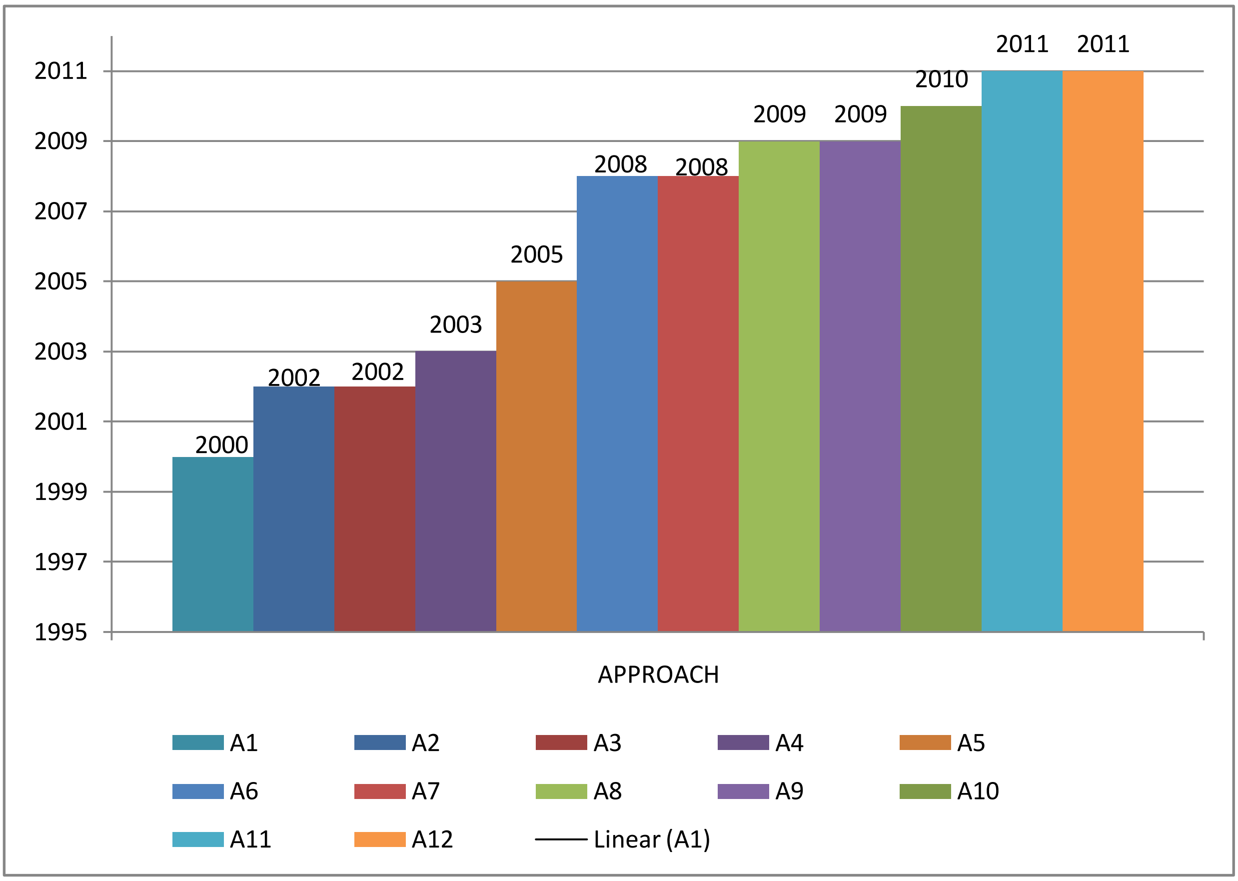


Fig. 4. Approaches of existing work

Table IVDataset Characteristic

Datasets	Meta Features										
	No of Instance	No of attribute	Nominal Count	Numerical count	Class Count	Missing Value	Mean	Mean Std.dev	Correlated Attribute	Outlier	Noisy Data
Contact-lenses	24	5	4	0	3	0	0	0	×	×	×
Diabetes	768	9	0	8	2	0	44.98	25.73	×	×	√
Super market	4627	217	216	0	2	3580	0	0	×	×	×
Glass	214	10	0	9	7	0	11.26	0.68	×	×	×
segment-test	1500	20	0	19	7	0	23.73	21.79	×	×	×
Vote	898	39	32	6	6	0	348.5	405.17	×	×	×
soybean	683	35	35	0	19	233	0	0	×	×	√

VI. PARAMETER FOR SELECTION OF CLASSIFICATION MODEL

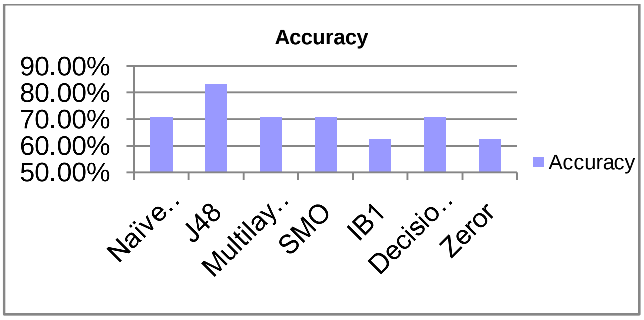
For the selection of classification model, various parameters need to be selected like dataset characteristics and classifier characteristics.

A. Dataset characteristics

Each dataset has different characteristics and classifier performance depends on the dataset characteristics. Someexperiment is performed by taking dataset from UCI machine repository. Table IV shows the dataset with different characteristics.

B. Impact of dataset on classifier performance

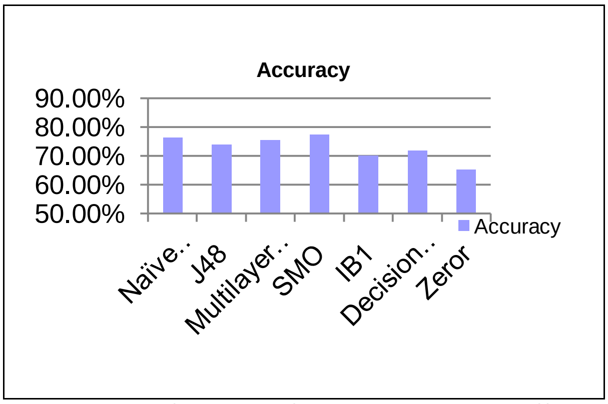
According to "The ‘no free lunch’ theorem of Wolpertand Macready, any two algorithms are equivalent when their performance is averaged across all possible problems Some experiment is carried out in Weka 3.6 with default settings. Here dataset is selected with different characteristics and classifier is selected from different class. For evaluating performance of the classifier, accuracy and mean absolute error is considered as measures. Graph 1 shows impact of contact_lense dataset on different classifier.



Graph 1. Performance of contact_lense dataset on different classifier

For contact_lense dataset, J48 gives better performance. Here accuracy is considered as parameter to measure the performance.

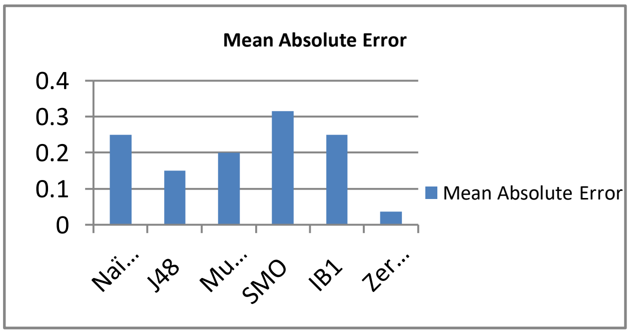
Graph 2 shows impact of diabetes dataset on different classifier.



Graph 2. Performance of diabetes dataset on different classifier

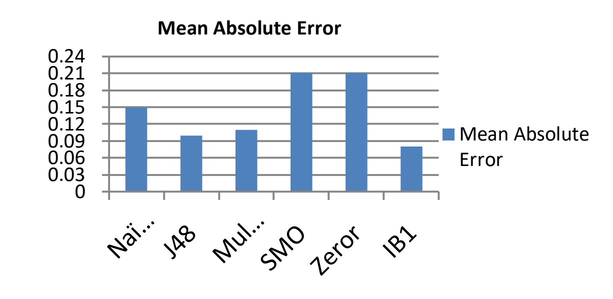
In graph 2, SMO gives better performance. Here accuracy is considered as parameter to measure the performance.

Graph 3 shows impact of contact_lense dataset on different classifiers.



Graph 3. Performance of contact_lense dataset

For contact_lense dataset, Zeror gives better performance than other classifiers as Mean absolute error of Zeror is less than other classifiers.



Graph 4. Performance of glass dataset

Graph 4 shows performance of glass dataset on different classifier.

For glass dataset, IB1 gives betterperformance than other classifiers. Here mean absolute error is taken as a measure for the evaluation.

C. Impact of classifier characteristics on performance

Each Classifier of different model has different characteristics. Table V shows comparative study of classifier characteristics. Table VI shows the performance of classifier based on its characteristics.

Table V Comparative Study of Classifier Characteristics

Classifier	Characteristics
k-Nearest Neighbor	• Instance of instance-based learning. • If no of instances are more kNN gives less misclassification error [27]. • On datasets with high to extremely high level of sparsity, kNN starts failing as it is unable to form reliable neighborhoods [19].
Naïve Bayes	• Robust to isolated noise points [28]. • Handles missing values [28]. • Robust to irrelevant attributes [28]. • Correlated attributes degrade the performance of NB classifier [28].
SMO	• Sequential minimal optimization (SMO) is an algorithm for efficiently solving the optimization problem [29]. • On datasets with high to extremely high level of sparsity, it gives best performance [30].
Neural Network	• Neural networks are used when the exact nature of the relationship between inputs and output is not known [30]. • NN is more powerful than the linear perceptron, as it can distinguish data that is not linearly separable [30].
J48	• Handling both continuous and discrete attributes. • J48 gives better performance for categorical attribute [31]. • Handling training data with missing attribute values.

Table VIPerformance of Classifier based on its Characteristics

Datasets	NB	J48	SMO	k-NN	Justification
Supermarket	**64.00%**	63.71%	63.71%	37.84%	NB gives better performance for missing values. Here missing values are 3570
KDD-Train	90.23%	94.62%	92.00%	**99.67%**	NB does not give better performance if attributives are correlated. If no of instances are more k-NN gives less misclassification error
Breast cancer	**75.52%**	71.67%	69.58%	72.37%	J48 gives better performance for categorical attribute.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Credit ratings	77.68%	**86.087%**	84.92%	81.15%	J48 better handles continuous attributes

VII. GENERATION OF RANKING BASED ON ADJUSTED RATIO OF RATIO

Considering the NFL theorem, we cannot expect that a single best algorithm could be found and be valid for all datasets. We address this issue by dividing the entire process into two distinct phases. In the first phase, we identify subset of relevant datasets and should be taken into account later. In the second phase, we proceed to construct a ranking on the basis of the datasets identified.

Following steps are performed for the generation of ranking for a new dataset.

Step 1: Dataset collection (Here dataset with different characteristics are selected)
Step 2: Meta Feature Extraction for training data
Step 3: Learning Strategy Selection
Step 4: Selection of Performance Measures
Step 5: Generation of Meta Knowledge Base
Step 6: Meta Feature Extraction for testing data
Step 7:Find relevant datasets from the Meta knowledge base for the new problem

$$\delta(V_x, d_i, V_x, d_j) = \frac{|\, V_x, d_i - V_x, d_j \,|}{Max_{k \neq i}(V_x, d_k) - Min_{k \neq i}(V_x, d_k)}$$

Step 8: Selects the first three dataset which has minimum distance
Step 9:Find pair wise mean adjusted ratio of ratio for each pair of algorithm.

$$\mathrm{ARR}_{a_p,a_q} = \frac{\left(\sum d_i \mathrm{ARR}^{d_i}_{a_p,a_q}\right)}{n}$$

Where n is number of datasets.
Step 10:Find overall mean adjusted ratio of ratio for each algorithm.

$$\mathrm{ARR}_{a_p} = \frac{\left(\sum_{a_q} \mathrm{ARR}_{a_p,a_q}\right)}{(m-1)}$$

Where m is number of algorithms.

VIII. CURRENT RESEARCH CHALLENGES

1. Traditional approaches to predicting the performance of algorithms involve, in general costly trial-and-error procedures, or require expert knowledge, which is not always easy to acquire [16].One of the issues in Meta Learning is generation of Meta Examples [10].

2. In order to produce a single meta-example, it is necessary to perform an empirical evaluation (e.g. cross-validation) of the candidate algorithms on a problem. Hence, the cost of generating a whole set of meta-examples may be high,Depending, for instance, on the number and complexity of the candidate algorithms, the methodology of empirical evaluation and the amount of available problems [18].

3. Existing work performs evaluation of classifier by considering single criterion.

a) Predictable factors such as the available amount of training data (relative to the dimensionality of the feature space), the spatial variability of the effective average distance between data samples, and the type and amount of noise in the data set influence such classifiers to a significant degree [11].

b) Authors have developed framework in which accuracy was measured achieved on a limited number of datasets, and a limited number of classifiers and their parameter settings [7].

c) Various applications like Biomedical datasets pose a unique challenge to machine learning and data mining algorithms for classification because of their high dimensionality, multiple classes, noisy data and missing values [5].

d) Meta Learning approaches for automatic algorithm selection assume that the features used to represent meta-instances are sufficiently relevant But some features may not be directly relevant, and some features may be redundant or irrelevant [6].

IX. CONCLUSION

The different approaches of Meta learning based on dataset characteristics provides a system that automatically provides ranking of the classifiers by considering different characteristics of datasets and different characteristics of classifiers After the generation of the Meta Knowledge Base, Ranking is provided based on Adjusted Ration of Ratio (ARR) or accuracy or time that helps non-experts in algorithm selection task.

REFERENCES

1. Pavel B. brazdil and Carlos Soares ,"A Comparision of Ranking Methods for Classification Algorithm Selection,2000.
2. R. Vilalta and Y. Drissi. A perspective view and survey of meta-learning. Journal of Artificial Intelligence Review, 18(2):77–95, 2002.
3. .Marcilio C.P.de Souto,RicardoB.C.Prudencio,RodrigoG.F.Soares, "Ranking and Selecting Clustering Algorithms Using a Meta-Learning Approach",2008.
4. Christophe Giraud-Carrier, Chair,DanVentura,Yiu-Kai Dennis Ng Eric Mercer,SeanWarnick, "Relationships among Learning Algorithms and Tasks", Proceedings of the International Conference on Machine Learning and Applications,2011.
5. Ajay Kumar Tanwari,JamalAfridi,M.ZubairShafiq and MuddassarFarooq, "Guidelines to Select Machine Learning Scheme for Classification of Biomedical Datasets",nexginrc, Evolutionary Computation, Machine Learning Scheme for Classification of Biomedical Datasets,Springer,2009
6. MykolaPechenizkiy, "Data Mining Strategy Selection via Empirical and Constructive Induction", Finland, 2003.
7. Stuart Moran,YulanHe,Kecheng Liu, "An Empirical Framework for Automatically Selecting the Best Bayesian Classifier", Proceedings of the World Congress Engineering 2009 Vol I WCE 2009, July 1-3,London,U.K,2009.
8. SilviuCacoveanu,CameliaVidrighin,RodicaPotolea, "Evolution Meta-Learning Framework For automatic Classifier Selection",2005.
9. ShawkatAli,Kate A. Smith, "On learning algorithm selection for classification",Applied Soft Computing Volume 6,Issue 2,119-138,January 2006.
10. Ricardo B.C.Orudencio and Teresa B. Ludermir, "Selective Generation of training examples in active

meta-learning", International Journal of Hybrid Intelligent Systems,2008.
11. C.M. van der walt and E.Barnard, "Data Characteristics that determines classifier performance",2008.
12. Myra Spiliopoulou, Alexis Kalousis, Lukas C. Faulstich and Theoharis, "NOEMON: An Intelligent Assistant for Classifier Selection", Citeseer,2000.
13. AlexandrosKalousis and Melanie Hilario, "Algorithm selection via meta learning",2002.
14. Ricardo Vilalta, Christophe Giraud-Carrier, PavelBrazdil, Carlos Soares, "Using Meta Learning to support Data Mining", 32-45,2004.
15. S.Appavu alias Balamurugan, Dr.R.Rajaram, G.Athiappan, M.Muthupandian, "Data Mining Techniques for suspicious Email Detection: A Comparative Study", IADIS European Conference Data Ming, Madurai,2007.
16. C. Giraud-Carrier, R.Vilalta and P. Brazdil, "Introduction to the special issue on meta-learning", Machine Learning 54, 187–193, 2004.
17. Carlos Soares and Pavel B. Brazdil, "Zoomed Ranking: Selection of Classification Algorithms Based on Relevant Performance Information",Principles of Data Mining and Knowledge Discovery,SpringerLink,2002.
18. Ricardo B.C. Prudencio and Teresa B. Ludermin, "Combining Uncertainity Sampling Methods for Active Meta Learning", Ninth International Conference on Intelligent Systems Design and Applications, 220-225,2009.
19. MihaGrcar, BlazFortun, DunjaMladeni"kNN Versus SVM in the Collaborative Filtering Framework", Data Science and Classification,2002.
20. Ricardo B.C. Prudencio and Teresa B. Ludermin, "Active Meta-Learning with Uncertainty Sampling and Outlier Detection", IEEE World Congress on Computational Intelligence, 2010.
21. Ricardo B.C. Prudencio and Teresa B. Ludermin, "Uncertainty Sampling Methods for Selecting Datasets in Active Meta Learning", Proceedings of International joint Conference on Neural Networks, San Jose, California, USA, July 31-August 5, 1082-1089, 2011.
22. D. Michie, D. Spiegelhalter and D. Taylor, Machine Learning,
23. Neural and Statistical Classification, Ellis Horwood, New York, 1994.
24. D.H.Wolpert and W.G.Macready. No Free Lunch Theorems for search. Technical Report SFI-TR-95-02-010, The santa Re Institute,1996.
25. P Brazdil ,C.Soares and J.dacosta, "Ranking Learning algorithms: Using IBL and meta-learning on accuracy and time results," Machine Learning,vol.50,np-8,pp.911-921,2004.
26. D.H.Wolpert. The supervised learning no-free-lunch theorems. In proceedings of the sixth on-line word conference on soft computing in Industrial applcations, pages 325-330,2001.
27. B.Pfahringer, H.Bensusan and C.Girand carrier, Meta learning by landmarking various learning algorithms, in proceedings of the 17th international conference on Machine Learning (ICML-2000),2000,743-750.

WEB SITE

28. "The bias-variance tradeoff", http://nlp.stanford.edu/IR-book/html/htmledition/the-bias-variance-tradeoff-1.html
29. cs-people.bu.edu/evimaria/cs565-11/lect11.pptx.
30. http://en.wikipedia.org/wiki/Sequential_minimal_optimization
31. "Semantic Assisted, Multiresolution Image Retrieval in 3D Brain MR Volumes", AzharQuddus, Waterloo, Ontario, Canada, 2010.
32. http://www.statsoft.com/textbook/neural-networks
33. "Classification via Decision Trees in WEKA", DEPAUL UNIVERSITY,http://maya.cs.depaul.edu/classes/ect584/weka/classify.html

AUTHOR PROFILE

Nikita Bhatt has received her B.E degree in Computer Engineering from Sardar Patel University, India in 2006 and currently pursuing her master degree in Computer Engineering from Charotar University of Science and Technology, Changa, Gujarat. Her current research interest includes Data Mining, Meta Learning and Active Meta Learning.

AmitThakkar has received his B.E degree in Information Technology from Gujarat University, Gujarat, India in 2002 and master Degree from Dharmsinh Desai University, Gujarat, India in 2007. He has joined his Ph.D in the area of Multi relational Classification at KadiSarvaVishvidhalayaUniversity, Gandhinagar, India in June 2010. Since 2002 he has been with faculty of Engineering and Technology, Charotar University of Science and Technology, Changa, Gujarat, Where he is currently working as an Associate Professor in the Department of Information Technology. He has published more than 20 research papers in the field of data mining and web technology. His current research interest includes Multi relational Data Mining, Relational Classification and Associate Classification.

Amit P. Ganatra (B.E.-'00-M.E. '04-Ph.D.* '11) has received his B.Tech. and M.Tech. degrees in 2000 and 2004 respectively from Dept. of Computer Engineering, DDIT-Nadiad from Gujarat University and Dharmsinh Desai University, Gujarat and he is pursuing Ph.D. in Information Fusion Techniques in Data Mining from KSV University, Gandhinagar, Gujarat, India and working closely with Dr.Y.P.Kosta (Guide). He is a member of IEEE and CSI. His areas of interest include Database and Data Mining, Artificial Intelligence, System software, soft computing and software engineering. He has 11 years of teaching experience at UG level and concurrently 7 years of teaching and research experience at PG level, having good teaching and research interests. In addition he has been involved in various consultancy projects for various industries.

After spending almost a year in C.U.Shahcollege of Engineering, Wadhwan, Gujarat, he joined CITC as a faculty member in 2001. His general research includes Data Warehousing, Data Mining and Business Intelligence, Artificial Intelligence and Soft Computing. In these areas, he is having good research record and published and contributed over 70 papers (Author and Co-author) published in referred journals and presented in various international conferences. He has guided more than 90 industry projects at under graduate level and 47 dissertations at Post Graduate level.

He is concurrently holding Associate Professor (Jan 2010 till date), Headship in computer Engineering Department (since 2001 to till date) at CSPIT, CHARUSAT and Deanship in Faculty of Technology-CHARUSAT (since Jan 2011 to till date), Gujarat. He is a member of Board of Studies (BOS), Faculty Board and Academic Council for CHARUSAT and member of BOS for Gujarat Technological University (GTU). He was the founder head of CE and IT departments of CITC (now CSPIT).