Ranking of Classifiers based on Dataset Characteristics using Active Meta Learning

Nikita Bhatt	Amit Thakkar	Amit Ganatra	Nirav Bhatt
Asst. Prof in Computer Engineering, CHARUSAT, Changa, Gujarat	Asso. Prof in Information Technology, CHARUSAT, Changa, Gujarat	Professor in Computer Engineering, CHARUSAT, Changa, Gujarat	Asst. Prof in Information Technology, CHARUSAT, Changa, Gujarat

ABSTRACT

Classification is a machine learning technique which is used to categorize the different input patterns into different classes. To select the best classifier for a given dataset is one of the critical issues in Classification. Using cross-validation approach, it is possible to apply candidate algorithms on a given dataset and best classifier is selected by considering various evaluation measures of Classification. But computational cost is significant. Meta Learning automates this process by acquiring knowledge in form of Meta-features and performance information of candidate algorithm on datasets and creates a Meta Knowledge Base. Once Meta Knowledge Base is generated, system uses k-Nearest Neighbor as a Meta Learner that identifies the most similar datasets to new dataset. But generation of Meta Example is a costly process due to a large number of candidate algorithms and datasets with different characteristics involved. So Active Learning is incorporated into Meta Learning System that reduces generation of Meta example and at the same time maintaining performance of candidate algorithms. Once the training phase is completed based on Active Meta Learning approach, ranking is provided based on Success Rate Ratio (SRR) method that considers accuracy as a performance evaluation measure.

Keywords

Classification, k-NN, Meta Learning, SRR

1. INTRODUCTION

In 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].

Out of available classification algorithms, selection of algorithm is a time consuming task which involves experimentation with different classifiers and analyzing the performance of those classifiers [8]. The 'no free lunch' theorem of Wolpert and Macready [3], states that "any two algorithms are equivalent when their performance is averaged across all possible problems". Traditional approaches to predicting the performance of algorithms, involve costly trial and error procedures or expert knowledge which is not always easy to acquire [10]. In addition to it, today's successful application of models in real-world scenarios requires continuous adaptation to new needs. Rather than starting afresh on new tasks, one would expect the learning mechanism itself to re-learn, taking into account previous experience to get profit from the repetitive use of a predictive model [16] [17].Apart from that, Predictive factors such as available amount of training data, the spatial variability of the effective average distance between the samples, and the type and amount of noise in the data set influence classifiers to significant degree [11].

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]. So Meta learning framework is developed in supervised machine learning which relates the performance of machine learning algorithms to the characteristics of the problem (i.e., characteristics of its datasets) [3] [21].In Meta learning, knowledge is acquired by meta-examples that store (a) the features that describe the dataset (problem) and (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). But generation of Meta example is a costly process that involves empirical evaluation of the candidate algorithms on a dataset. So Active Meta learning technique is used, that reduces the cost of generating Meta examples by selecting relevant Meta examples [10].

2. APPROACHES FOR ALGORITHM SELECTION

Table 1 shows various approaches to algorithm selection problem.

Table 1. 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 requires O(nm) according to graph theory which is a costly process.
Expert Advice	On arrival of new dataset, 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 [7].
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.

3. META LEARNING

Meta Learning 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 [14].

3.1 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. Upon arrival of each dataset, features are extracted which is either simple features (Attribute Count, Nominal Count, Numeric Count, Incomplete Instance Count, Instance Count)or Statistical features (Mean, Standard Deviation, Skewness, kurtosis). Once features are extracted, candidate algorithms are applied on it. Meta features and performance information of candidate algorithms are stored into database known as Meta Knowledge Base. So output of Acquisition Mode is Meta Knowledge Base. Fig 1 shows the general structure of the acquisition mode.

Advisory Mode:

In the advisory mode, meta-knowledge acquired in the Acquisition 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 ranking. Fig 2 shows the general structure of the advisory mode.

3.2 Weakness of Meta Learning

An important issue in the development of Meta Learning systems for algorithm recommendation is the computational cost of generating Meta data[18][25]. This implies running the candidate algorithms on all the training datasets. Generating a whole set of meta-example may be expensive, depending for instance number of candidate algorithms, complexity of algorithms, number of available problems, efficiency of the procedure used for algorithm evaluation. In most Meta learning approach in automatic algorithm selection, some features may not be directly relevant and some features may be redundant or irrelevant [6]. In Meta Learning, Space Complexity of Meta Knowledge Base is increased as Meta Learning space is vast and getting larger with invention of new algorithms.

4. ACTIVE LEARNING

Active Leaning is a paradigm of Machine Learning in which the learning algorithm has some control over the inputs on which it trains [3].

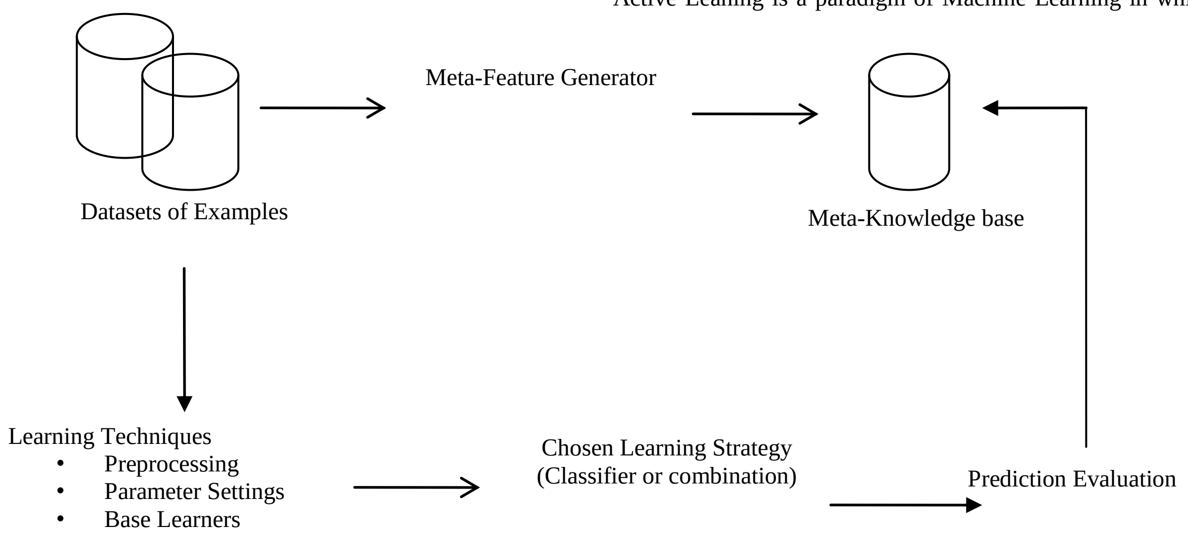

Fig 1: Acquisition Mode of Meta Learning

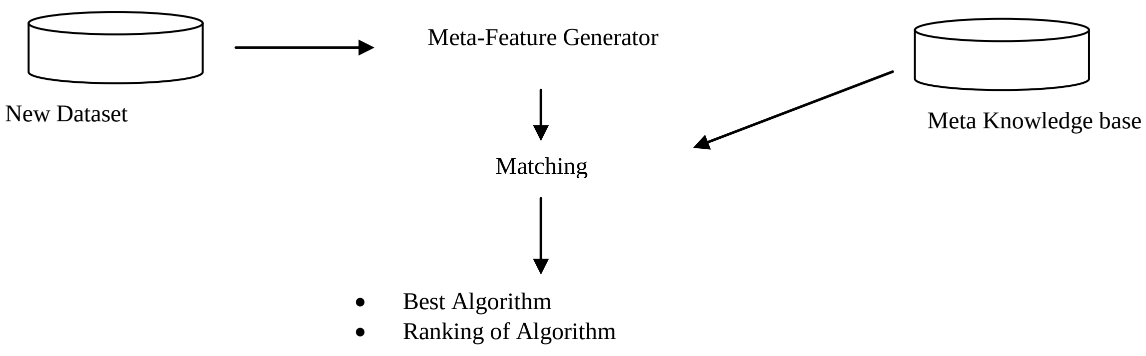

Fig 2: Advisory Mode of Meta Learning

The main objective of this paradigm is to reduce the number of training examples, at same time maintaining, or even improving, the performance of the learning algorithm. Active Learning is ideal for learning domains in which the acquisition of labeled examples is a costly process. Active Learning starts with a small set of one or more labeled examples and a large set of unlabeled ones. Labeled examples in the context of meta-learning are generated from datasets for which the candidate algorithms were already evaluated. The unlabeled examples in turn correspond to datasets which are only candidates for meta-example generation. An active learning module receives these two sets as input and selects, from the later, the next example to be labeled. The selection of unlabeled meta-examples is performed based on pre-defined criteria which take into account the meta-features of the problems and current set of labeled examples. Labeling is done by evaluating the candidate algorithms on the selected problems and the best algorithm on each of them becomes the label of the corresponding meta-example.

4.1 Approaches to Active Learning

Table 2 shows approaches for Active Learning.

Table 2. Active Meta Learning Approaches

Approach	**Description**
Membership Query	Learner artificially creates informative examples in the input domain and asks to annotate it. According to [22], this approach is limited in practice since it is likely to produce examples that do not have any sense in the domain of the application.
Selective Sampling Approach	The learning algorithm has access to a set of unlabeled examples and, at each moment, selects the most informative ones. (a) Uncertainty-based Method: In uncertainty-based methods for selective sampling, in order to select unlabeled examples, the learner initially uses the currently labeled examples to generate a prediction for each unlabeled example. Following, a degree of uncertainty of the provided prediction is assigned for each unlabeled example. Finally, the active method selects the example with highest uncertainty. (b) Version space reduction methods (committee-based methods): A subset of the version space (i.e. a committee of hypotheses consistent with the current labeled examples) is generated and then applied to make predictions for the unlabeled examples. The method then selects the unlabeled example on which the members of the committee most disagree. (c) Error reduction methods: The selected unlabeled example is the one that minimizes the expected error rate of the learner, once labeled and included in the training set. Since the true label of an unlabeled example is not known a priori, the expected error rate is an average rate over the possible labels that the example could be assigned to.

5. PROPOSED ALGORITHM

Input: Meta example (meta features and performance information of initial training datasets)

Output: Ranking of classifiers for a test dataset

Method:

Step 1: Creation of Meta Knowledge Base which contains meta features and performance information of initial training datasets.

Step 2: Meta Features Extraction from a set of datasets (sample datasets) to generate Unlabeled Meta Example

Step 3: Selection of Unlabeled Meta Example with highest uncertainty by sampling method of Active Learning

$$\delta(V_{x,}d_i, V_{x,}d_j) = \frac{|V_x\,, d_i - V_x, d_j\,|}{\mathrm{Max}_{k\neq i}(V_x, d_k) - \mathrm{Min}_{k\neq i}(V_x, d_k)}$$

Where $\mathrm{Max}_{k\neq i}\,(V_x,d_k)$ calculates the maximum value of meta-attribute x for all datasets except d_i and $\mathrm{Min}_{k\neq i}\,(V_x,d_k)$ calculates the corresponding minimum.

Step 4: Labeling of Selected Unlabeled Meta Example having highest uncertainty.

Step 5: Meta Feature extraction for the new problem.

Step 6: Find relevant datasets from the Meta knowledge base for the new problem using distance function.

Step 7: Apply Ranking method

- Find Success Rate Ratio of algorithm j and k on dataset i.

$$SRR_{j,k}^{i} = \frac{(1 - ER_j^i)}{(1 - ER_k^i)}$$

Where ER_j^i is the measured error rate of algorithm j on dataset i.

- Calculate a pair wise mean Success Rate Ratio for each pair of algorithms.

$$SRR_{j,k} = \frac{\sum_i SRR_{j,k}^{i}}{n}$$

Where j and k are algorithms and n is number of datasets.

- Find overall mean Success Rate Ratio for each algorithm.

$$SRR_j = \frac{\sum_k SRR_{j,k}}{(m-1)}$$

Where m is number of algorithms.

The classification uncertainty of the k-NN algorithm is defined as the ratio of: (1) the distance between the unlabeled example and its nearest labeled neighbor; and (2) the sum of the distances between the unlabeled example and its nearest labeled neighbors of different classes.

6. EXPERIMENTS

In order to study the performance, the algorithm has been tested on an Intel based machine with a 2.40 GHz processor and 3 GB of main memory. All the Performances related to work is implemented under the java 1.6. The Weka tool developed by the Waikato University is used to simulate the classification process. WekaMetal 0.11 is used to simulate the meta-learning process.

6.1 Dataset Characteristics

Table 3 shows that each dataset has different characteristics.

6.2 k-NN Meta Learner

In proposed work, active learning approach uses uncertainty sampling method for k-NN algorithm. Here method selects unlabeled examples for which the current k-NN learner has the highest uncertainty in its prediction. A high value of uncertainty indicates that the unlabeled example has nearest neighbors with similar distances but conflicting labeling. So once an uncertain unlabeled example is labeled, it is expected that the uncertainty in its neighborhood is reduced [20]. k-NN is an instance based learning approach. IBL has advantage that system is extensible. Once a new experimental result becomes available; it can be easily integrated into existing results without the need to reinitialize complex re-learning. This property is relevant for algorithm selection because, the user starts with a small set of meta-data but these set increases steadily with time.

Experiment 1: Table 4 shows ranking of the classifiers for the test dataset “anneal” using SRR based on Meta learning and Active Meta Learning (Using Uncertainty Sampling Method) approach. Success Rate Ratio (SRR) ranking method works on accuracy measure where the accuracy of the classification model is determined by comparing true class labels in the testing set with those assigned by the model [15].For Meta learning approach, all the datasets with different characteristics shown in table 3 is selected and based on that ranking is provided. Here the classifier whose SRR is highest has been assign the first rank, classifier whose SRR is second highest has been assign a second rank and so on. For Active Meta Learning approach, first 8 datasets from the table 3 is selected initially. After applying uncertainty sampling method, one dataset out of remaining training datasets is selected and ranking is provided. Success Rate Ratio can be seen as a measure of the advantage of one algorithm over other algorithm. Fig 3 shows ranking of classifiers based on SRR using Meta Learning and Active Meta Learning approach.

Datasets	Attribute Count	Nominal Count	Numeric Count	Class Count	Default Accuracy	Incomplete Instance Count	Instance Count	Missing Values	Mean	Mean StdDev	Mean Kurt-osis	Mean Skewn-ess
balance-scale	5	0	4	3	0.46	0	625	0	3	1.41	-1.3	0
contact-lenses	5	4	0	3	0.62	0	24	0	0	0	0	0
diabetes	9	0	8	2	0.65	0	768	0	44.98	25.73	2.74	0.52
glass	10	0	9	7	0.35	0	214	0	11.26	0.68	9.59	1.63
labor	17	8	8	2	0.64	56	57	326	9.41	2.23	1.53	0.24
sonar	61	0	60	2	0.53	0	208	0	0.28	0.14	2.16	0.95
vehicle	19	0	18	4	0.25	0	846	0	117.63	22.77	5.1	1.03
vowel	14	3	10	11	0.09	0	990	0	-0.1	0.69	-0.39	0.09
iris	5	0	4	3	0.33	0	150	0	3.46	0.94	-0.78	0.06
weather	5	2	2	2	0.64	0	14	0	77.6	8.42	-1.29	0.08
zoo	18	16	1	7	0.4	0	101	0	2.84	2.03	-0.65	0.13
kr-vs-kp	37	36	0	2	0.52	0	3196	0	0	0	0	0
lymph	19	15	3	4	0.54	0	148	0	2.045	1.018	9.427	2.295
wisconsin	33	0	32	1	0	0	194	0	87.54	33.36	2.7	1.12
ionosphere	35	0	34	2	0.64	0	351	0	0.247	0.51	0.3197	-0.564
spect_test	23	22	0	2	0.55	0	187	0	0	0	0	0
heart-statlog	14	0	13	2	0.55	0	270	0	46.036	8.44	0.079	0.423

Table 3. Dataset with different characteristics

Table 4. Ranking of the classifiers for the test dataset "anneal" using SRR

Mode	Datasets in Meta Knowledge Base	Min.Dist.	NN	J48	P	Ibk	IB1	DT	NB	OneR	DS	ZeroR
Meta Learning	17	1 lymph - 4.1257 2 vehicle - 4.3198 3 Diabetes - 4.5451	1.14	1.06	1.05	1.02	1.01	0.97	0.9	0.88	0.81	0.6
Rank			1	2	3	4	5	6	7	8	9	10
			NN	**J48**	**P**	**Ibk**	**DT**	**IB1**	**NB**	**OneR**	**DS**	**ZeroR**
Proposed	9	1 vehicle - 4.4642 2 Diabetes - 4.4964 3 kr-vs-kp - 4.8651	1.15	1.09	1.08	1.06	1.05	1.05	0.88	0.83	0.75	0.57
Rank			1	2	3	4	5	6	7	8	9	10
NN = Neural Network, DT = Decision Table, NB = Naïve Bayes, DS = Decision Stump, P = PART												

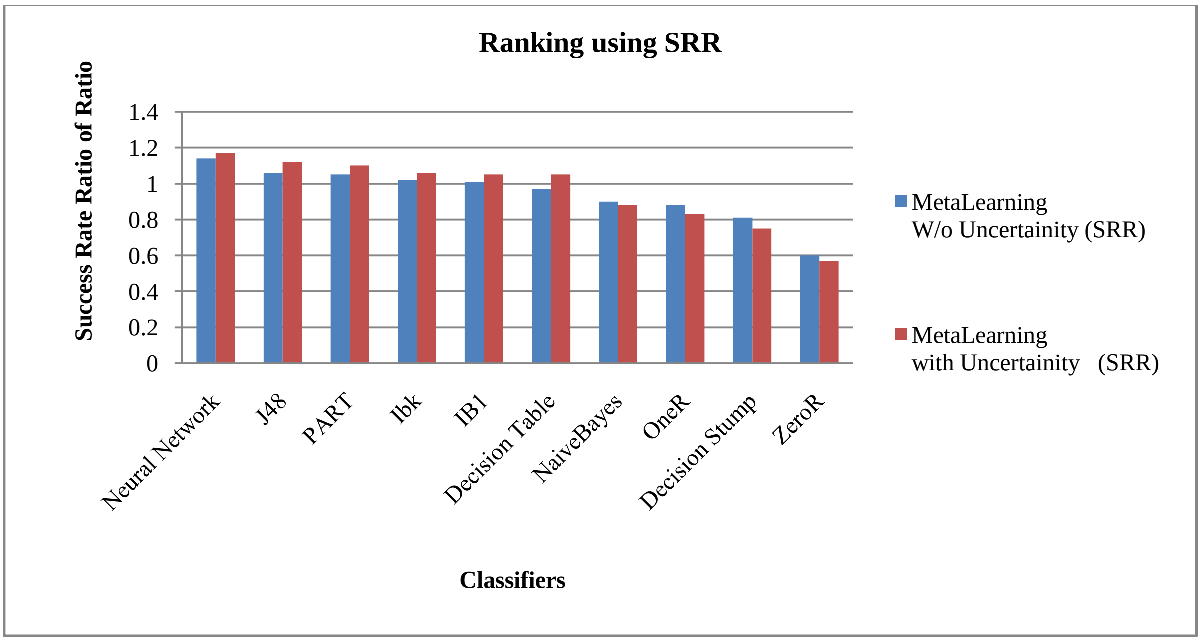

Fig 3: Ranking of classifiers based on SRR using Meta Learning and Active Meta Learning approach

Experiment 2: Table 5 shows time (in sec.) required to perform ranking using Meta learning and proposed method. Here experiment is carried out with 16 datasets for Meta Learning approach and 8 datasets for proposed approach.

Table 5. Performance of Meta learning and proposed system with time

Method	Datasets	Time (sec)
Meta Learning	17	88
Proposed	9	27

7. FUTURE WORK

Future Extension is to implement other ranking techniques like ARR (Adjusted Ratio of Ratio) and compare the performance of SRR with ARR. Uncertainty sampling method is sensitive to the presence of outliers, so Outlier Detection techniques can be used to handle outlier and then ranking is provided.

8. CONCLUSION

The presented work, "Ranking of Classifiers based on Dataset Characteristics using Active Meta Learning", provides a system that automatically provides ranking of the classifiers by considering different characteristics of datasets and classifiers. Here for improving the performance of Meta learning, Uncertainty Sampling Method of Active Meta Learning is used to reduce the set of meta-examples by selecting only the most relevant meta-examples. After generation of the Meta Knowledge Base, Ranking is provided based on Success Rate Ratio (SRR). Experiment shows that ranking of the datasets using reduced Meta Knowledge base maintains with the ranking using Meta Learning.

9. REFERENCES

[1] Niklas Lavesson and Paul Davidsson, 2008, Analysis of Multi-Criteria Metrics for Classifier and Algorithm Evaluation, scientific commons.

[2] Iain Paterson, Helmut Barrer and Jorg Keller, 2002, The Focused Multi-Criteria Ranking Approach to Machine Learning Algorithm Selection - An Incremental Meta Learning Assistant for Data Mining Tasks, IDDM Institute for advanced studies, Austria, CiteSeer.

[3] Marcilio C.P.de Souto, Ricardo B. C. Prudencio, Rodrigo G. F. Soares, 2008, Ranking and Selecting Clustering Algorithms Using a Meta-Learning Approach.

[4] Christophe Giraud-Carrier, Chair, Dan Ventura, Yiu-Kai Dennis Ng Eric Mercer, Sean Warnick, 2011, Relationships among Learning Algorithms and Tasks, Proceedings of the International Conference on Machine Learning and Applications.

[5] Ajay Kumar Tanwari, Jamal Afridi, M.ZubairShafiq and Muddassar Farooq, 2009, Guidelines to Select Machine Learning Scheme for Classification of Biomedical

[6] Datasets, nexginrc, Evolutionary Computation, Machine Learning Scheme for Classification of Biomedical Datasets, Springer.

[7] Mykola Pechenizkiy, 2003, Data Mining Strategy Selection via Empirical and Constructive Induction, Finland.

[8] Stuart Moran,YulanHe, Kecheng Liu, 2009, An Empirical Framework for Automatically Selecting the Best Bayesian Classifier, Proceedings of the World Congress Engineering 2009 Vol I WCE 2009 London July 1-3,U.K.

[9] Silviu Cacoveanu, Camelia Vidrighin, Rodica Potolea, 2005, Evolution Meta-Learning Framework For automatic Classifier Selection.

[10] Shawkat Ali, Kate A. Smith, January 2006, On learning algorithm selection for classification, Applied Soft Computing Volume 6, Issue 2, 119-138.

[11] Ricardo B. C. Orudencio and Teresa B. Ludermir, 2008, Selective Generation of training examples in active meta-learning, International Journal of Hybrid Intelligent Systems.

[12] C.M. van der walt and E.Barnard, 2008, Data Characteristics that determines classifier performance.

[13] Myra Spiliopoulou, Alexis Kalousis, Lukas C. Faulstich and Theoharis, 2000, NOEMON: An Intelligent Assistant for Classifier Selection, Citeseer.

[14] Alexandros Kalousis And Melanie Hilario, 2002, Algorithm selection via meta learning.

[15] Ricardo Vilalta , Christophe Giraud-Carrier, Pavel Brazdil, Carlos Soares, 2004, Using Meta Learning to support Data Mining , 32-45.

[16] S.Appavu alias Balamurugan, Dr.R.Rajaram, G.Athiappan, M.Muthupandian, 2007, Data Mining Techniques for suspicious Email Detection: A Comparative Study, IADIS European Conference Data Ming, Madurai.

[17] C. Giraud-Carrier, R.Vilalta and P. Brazdil, 2004, Introduction to the special issue on meta-learning, Machine Learning 54, 187–193.

[18] Carlos Soares and Pavel B. Brazdil, 2002, Zoomed Ranking: Selection of Classification Algorithms Based on Relevant Performance Information, Principles of Data Mining and Knowledge Discovery, Springer Link.

[19] Ricardo B.C. Prudencio and Teresa B. Ludermin, 2009, Combining Uncertainty Sampling Methods for Active Meta Learning, Ninth International Conference on Intelligent Systems Design and Applications, 220-225.

[20] MihaGrcar, BlazFortun, Dunja Mladeni, 2002, k-NN Versus SVM in the Collaborative Filtering Framework, Data Science and Classification.

[21] Ricardo B.C. Prudencio and Teresa B. Ludermin, 2010, Active Meta-Learning with Uncertainty Sampling and Outlier Detection, IEEE World Congress on Computational Intelligence.

[22] D. Cohn, L. Atlas, and R. Ladner, 1994, Improving generalization with active learning, Machine Learning, vol. 15, pp. 201–221.

[23] G. Riccardi and D. Hakkani-Tur, 2005, Active learning: theory and applications to automatic speech recognition, IEEE Transactions on Speech and Audio Processing 13, 504–511.

[24] D. Angluin, 1998, Queries and concept learning, Machine Learning 2, 319–342.

[25] I. Muslea, S. Minton and C. Knobrock, 2006, Active learning with multiple views, Journal of Artificial Intelligence Research 27, 203–233.

[26] Ricardo B.C. Prudencio and Teresa B. Ludermin, 2011, 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.

Pavel B. Brazdil, Carlos Soares, 2003, Ranking Learning Algorithms: Using IBL and Meta-Learning on Accuracy and Time Results.