Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 36 (2014) 328 – 334

Procedia Computer Science

Complex Adaptive Systems, Publication 4
Cihan H. Dagli, Editor in Chief
Conference Organized by Missouri University of Science and Technology
2014- Philadelphia, PA

AHP Based Classification Algorithm Selection for Clinical Decision Support System Development

Sina Khanmohammadi[a]*, Mandana Rezaeiahari[a]

[a]*Department of Systems Science and Industrial Engineering, Watson School of Engineering, State University of New York at Binghamton, Binghamton, New York, 13902-6000, U.S.A*

Abstract

Supervised classification algorithms have become very popular because of their potential application in developing intelligent data analytic software. These algorithms are known to be sensitive to the characteristic and structure of input datasets, therefore, researchers use different algorithm selection methods to select the most suitable classification algorithm for specific dataset. These methods do not consider the uncertainty about input dataset, and relative importance of different performance measurements (such as speed, accuracy, and memory usage) in the target application domain. Therefore, these methods are not appropriate for software development. This is especially true in medical field where various high dimensional noisy data might be used with the software. Hence, software developers need to select one supervised classification algorithm that has the highest potential to provide good performance in wide variety of datasets. In this regard, an Analytic Hierarchy Process (AHP) based meta-learning algorithm is proposed to identify the most suitable supervised classification algorithm for developing clinical decision support system (CDSS). The results from ten publicly available medical datasets indicate that Support Vector Machine (SVM) has the highest potential to perform well on variety of medical datasets.

Peer-review under responsibility of scientific committee of Missouri University of Science and Technology

Keywords: Clinical Decision Support System (CDSS); Machine Learning; Medical Informatics; Algorithm Selection; AHP, Meta-Learning

1. Introduction

With the advancement of computers and increasing usage of electronic patient record systems, huge amounts of medical data have become available. This huge data have brought various opportunities such as developing clinical decision support systems (CDSS) that can improve the diagnosis and prognosis of patients [1, 2]. For example, many algorithms have been developed for automatic breast cancer diagnosis using patient information [3, 4]. One of the main methods used for developing CDSS software is supervised classification algorithms. Supervised classification algorithms are one of the most popular machine learning techniques where the objective is to predict the category of new observation based on identified patterns in the labeled dataset. For this purpose, the predictive variables (called features) and category of each observation (called class) is given, and supervised classification algorithm uses some part of this data (training set) to identify the patterns in dataset, and use the other part of the data (testing set) to validate the extracted patterns. Fig.1 shows the general framework of supervised classification algorithms.

Fig.1. The general framework of supervised classification algorithms

* Corresponding author.
E-mail address: skhanmo1@binghamton.edu

Peer-review under responsibility of scientific committee of Missouri University of Science and Technology
doi:10.1016/j.procs.2014.09.101

Each step of the supervised classification framework is described as follows. For convenience, we will refer to supervised classification algorithms as classification algorithms (classifiers) in the reminder of this paper. However, it should be noted that there is another group of classification algorithms called unsupervised classification algorithms (clustering) which do not consider the class label information in the dataset [5].

Step one of classification framework consists of gathering the data. If the investigator wants to collect the data by him/herself, an expert should be consulted to understand which features are most relevant to the output being investigated. If consulting an expert is not possible, then the data should be collected in brute-force format [6], meaning all the possible features should be collected.

Step two is about preprocessing the data and making it ready for classification algorithm [7]. Three major aspects of data should be addressed here. First, every data has some sort of noise which should be considered and if possible eliminated from the dataset. The second issue is missing values, which is usually caused by human related or equipment related errors. These missing values should be dealt with to improve the performance of classification algorithms. Two simple methods to deal with missing data are removing the observations that contain missing values, or replacing the missing values with average value of the feature they belong to. These simple methods have some limitations, therefore, other sophisticated algorithms have been developed to deal with missing values [7]. The last issue that needs to be addressed in this step is feature selection. Feature selection is the act of removing redundant and irrelevant features by selecting the ones that are most related to the class label [6, 8]. This will improve the performance of classification algorithms.

One drawback of classification algorithms is the fact that their performance (such as accuracy and speed) depends on the characteristic of datasets (such as size, sparsity, and dimensionality) [9]. This results in selective superiority characteristic meaning that each classification algorithm performs well on certain type of dataset [10]; therefore, the investigator should select the right classification algorithm for the specific dataset being analyzed [11]. This process is called classification algorithm selection and is the third step in classification algorithm framework. Two main methods for classification algorithm selection include performance comparison and meta-learning, which will be discussed in section 2.

Step four consists of fitting the selected classification algorithm to the training dataset. Training dataset is a portion of overall dataset used for identifying patterns that can describe the relationship between observations and class label.

After fitting the model, the model needs to be validated in step five. Two main methods exist for this purpose. In the first method data is divided into three section (for example 70%-15%-15%). The first section is training dataset used for fitting the model. The second section is validation dataset used to test the model and tune different parameters of the classification algorithm (such as number of iterations in Support Vector Machine), and the third section is testing dataset used to compare two different classification algorithms. The second method is called k-fold cross validation, where dataset is divided into k groups each containing same number of observations. $k-1$ folds are used for training and the k^{th} fold is used for testing. This process is repeated until each of the k folds has been selected as testing dataset, then the overall accuracy is calculated as the average accuracy of all the test processes [6].

Finally, in step six the validated model is retrained using whole dataset and used to predict the class label of new observations.

In this paper, we propose a new algorithm selection model called Analytic Hierarchy Process Algorithm Selection (AHP-AS) which considers the uncertainty about input dataset and relative importance of different performance measurements. These properties make the AHP-AS method suitable for software development. To illustrate this, we have applied the AHP-AS model to sample medical datasets to identify the most desirable classification algorithm for developing CDSS.

The reminder of this paper is organized as follows. Section 2 includes a brief background of different classification algorithm selection methods followed by details of the AHP model. Section 3 provides detailed description of the proposed AHP-AS method and the experimental results, followed by a brief conclusion in section 4.

2. Background and Related Work

2.1. Classification Algorithm Selection

Algorithm selection is defined as “learning a mapping from feature space to algorithm performance space, and acknowledged the importance of selecting the right features to characterize the hardness of problem instances” [12]. Simplest classification algorithm selection method random selection. Clearly, this is not a good strategy as it does not consider any information about the dataset.

A popular method among researchers is performance comparison, where the best performing classification algorithm is selected using several experiments on the target dataset. In this regards, different approximate statistical tests have been proposed to rank the performance of classification algorithms. Brazdil and Soares [13] have proposed three different ranking techniques including Average Rank (AR), Success Rate Ratios (SRR) and Significant Wins (SW). Their experiments show that SRR and AR perform better than SW. Along the same lines, Demsar [14] has proposed nonparametric tests of Wilcoxon signed rank test and Friedman test for comparison of different classifiers. They claim that nonparametric tests are more suitable for classifier comparison. Testing different classification algorithms on large datasets can be computationally intensive and slow, therefore, some researchers such as Petrak [15] has proposed sub sampling methods to evaluate the performance of different classification algorithms.

An alternative to performance comparison method is meta-learning approach, where classification algorithms are selected based on the knowledge acquired from datasets that are similar to the target dataset. In other words, meta-learning consists of learning about the performance of learning algorithms without directly testing them on the target dataset [16]. Meta-learning algorithms provide automatic and systematic guidelines about model selection and method combination. Knowledge acquisition in meta-learning can be done theoretically or experimentally.

Rendell and Cho [17] were the pioneers who used this method to examine the impact of features such as size and concentration of the classes on the classification algorithms. After Rendell and Cho, Aha [18] used various rule based learning algorithms to extract rules that can guide the user in classification algorithm selection. A much larger scale project regarding meta-learning was called Comparative Testing of Statistical and Logical Learning (StatLog), where different machine learning algorithms were tested to provide a benchmark [19]. Along the same lines, Lim and his colleagues [20] have evaluated different classification algorithms in terms of accuracy and CPU time. According to their results, the Polyclass Algorithm provided the best accuracy, but it was not statistically significantly different from other algorithms. Regarding to CPU time, the C4.5 decision tree provided the lowest CPU time and it was statistically significant when compared to other algorithms. Tanwani and his colleagues [11] have focused on classification algorithm selection for biomedical datasets, and provided a comprehensive evaluation of diverse machine learning algorithms on biomedical datasets. They also provided set of guidelines for selecting the appropriate classifier based on the properties of target biomedical dataset. Other researchers such as Kalousis and Hilario [21] have proposed building algorithm profiles to address the algorithm selection problem. The profiles consist of meta-level feature values that describe the learning algorithm in terms of functionality, robustness, and so on. Kalousis and Hilario used expert knowledge and previous empirical studies to identify the profiles. Brodley [10] considered the algorithm selection problem as an optimization problem and used knowledge about the representational biases of a set of classification algorithms to conduct an exhaustive search across a feature space defined by candidate classification algorithms. More recent research in this area include Ali and Smith's [9] comprehensive study where they employed C5.0 decision tree to generate general rules for selecting classification algorithms based on the type of target dataset.

Considering the objective of developing clinical decision support systems (CDSS), meta-learning algorithms in literature have the following limitations:

- Most of them consider each performance measure (such as accuracy, CPU time, and memory usage) separately without taking into account all these measures together.
- Most of them require a lot of computational resources.
- Few of them are focused on medical datasets.
- Some of them have been implemented using diverse set of programming languages, which makes the performance measure comparisons (such as CPU time) biased.
- They don't consider domain expert knowledge about the cost of each performance measure. For example, in medical field the accuracy has much higher priority than speed or memory usage.
- They don't consider the uncertainty about characteristic of dataset that is going to be used with the classifier.

In this research, we address these issues by proposing a new meta-learning framework called AHP-AS. This framework uses the AHP algorithm which is discussed in next section.

2.2. AHP

AHP is a multiple criteria decision-making tool which has been used for wide variety of applications [22]. The Analytic Hierarchy Process (AHP) was developed by Saaty in the 1970s to aid decision makers rank information based on pair-wise comparison of a number of criteria [23]. The criteria can be either qualitative, quantitative or both. Pair-wise comparisons are made among elements of each hierarchy using a nominal scale [24]. More specifically, AHP provides a procedure to calibrate the numeric scale for the measurement of quantitative and qualitative performances. The scale ranges from 1/9 for "least valued than" to 9/9 for "absolutely more important than", covering the entire spectrum of the comparison [22]. Generally, AHP follows four major steps [25]. First, the decision analyst should break down the decision problem into a hierarchy of interrelated decision elements. Second, a comparison matrix is formed for attributes pair-wise comparison. The result of this step is an eigenvector (also called a priority vector) that represents the relative ranking of importance (or preference) attached to the attributes. Third, a paired comparison of each criteria and alternative choices within each criterion is arranged. This step is repeated for all of the attributes. Fourth, the weighted evaluation is calculated which is the multiplication of priority weights and matrices of evaluation ratings of alternatives for each attribute. For all mentioned steps, AHP uses matrix algebra to sort out the factors and arrive at a mathematically optimal solution. Saaty allowed some measures of inconsistency for the logic of preferences [23]. Inconsistencies may happen when comparing three items, A, B, and C. For instance, if item A is more preferred over item B, and item B is more preferred over item C, then by the transitive property, item A should be more preferred over item C. If not, then the comparisons are not consistent. This measure of inconsistency is called Consistency Ratio (CR) which should be less than 0.1 to obtain priority weights that are acceptable. The flowchart of decision process for AHP is shown in Fig 2.

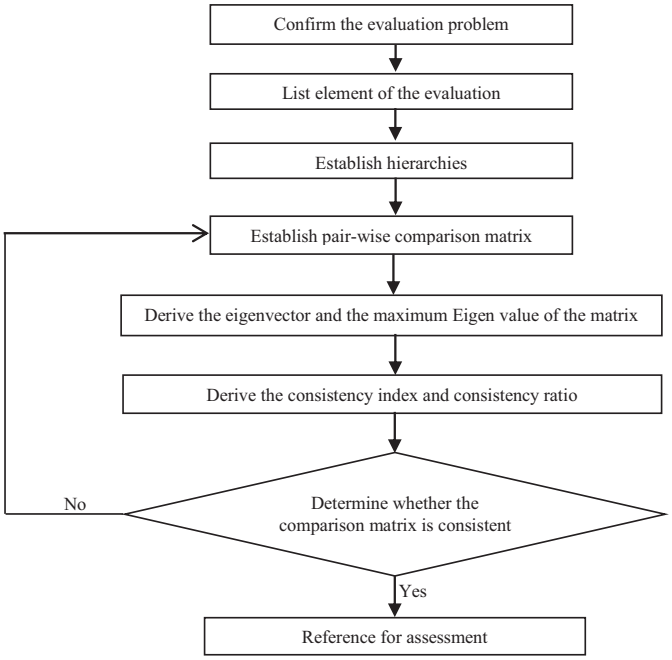

Fig. 2. Flowchart of the decision process of AHP [24]

3. Methodology and Results

The overall framework for the proposed classification algorithm selection model consists of six steps and is illustrated in Fig 3.

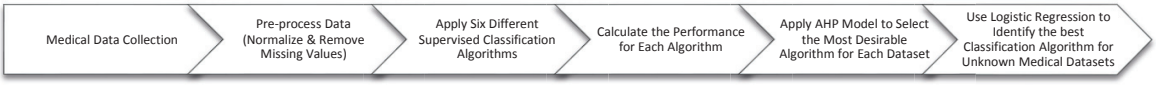

Fig. 3. Overall decision making framework for identifying the most suitable classification machine learning algorithm for medical data analysis software

At step one, ten different medical datasets from UCI repository [26] were retrieved to represent sample medical datasets. High dimensional datasets such as gene expression are not considered in this study as these types of datasets are usually used by bioengineering researches who can employ existing classification algorithm selection methods in the literature. The focus of this study is more on diagnostic datasets that might be used in CDSS by healthcare professionals who do not have the necessary knowledge about machine learning algorithms. The datasets used in this study include Breast Cancer Wisconsin (Original, Diagnostic and Prognostic), Cardiotocography, Diabetes, Echocardiogram, Heart Disease, Parkinson's, Hepatitis and Urinary Disease. Table.1 describes the different properties of these datasets.

Step two is the preprocessing step, where each sample dataset is preprocessed by removing the observations that have missing values and normalizing the values of each feature to a value between 0 and 1.

In step three and four, six common classification algorithms including Support Vector Machine (SVM), Logistic Regression (LR), K-Nearest Neighbors (KNN), Naïve Bayes (NB), Decision Tree (DT) and Discriminant Analysis (DA) were applied to each of these datasets. For each algorithm three performance measurements were calculated including average classification accuracy (using 5 fold cross validation), time required for training (in seconds) and memory usage (megabytes).

Table 1. Properties of different medical dataset obtained from UCI repository

Dataset	Number of Features	Number of Observations	Dataset Characteristics	Attribute Characteristics
Breast cancer Wisconsin (Original)	10	699	Multivariate	Integer
Breast Cancer Wisconsin (Diagnostic)	32	569	Multivariate	Real
Breast Cancer Wisconsin (Prognostic)	34	198	Multivariate	Real
Cardiotocography Data	23	2126	Multivariate	Real
Diabetes Data	20	768	Multivariate, Time-Series	Categorical, Integer
Echocardiogram Data	12	132	Multivariate	Categorical, Integer, Real
Heart Disease Data	14	270	Multivariate	Categorical, Real
Parkinson's Data	23	197	Multivariate	Real
Hepatitis Data	19	155	Multivariate	Categorical, Integer, Real
Urinary Disease Data	6	120	Multivariate	Categorical, Integer

All the experiments were implemented in MATLAB. The summary of applying machine learning algorithms to the sample medical datasets is given in Table 2.

Table 2. Summary of the classification algorithms applied to 10 sample medical datasets

	Average Speed of Training (Seconds)	Average Accuracy (%)	Average Memory Usage (MB)
SVM	0.0090	86.2933	0.1074
LR	0.0255	84.6301	0.8066
KNN	0.0317	84.3888	0.6584
NB	0.0086	80.8345	0.0444
DT	0.0263	80.3296	0.3861
DA	0.0161	82.0821	0.3900

Using the results in Table 2, an AHP model was developed to rank the best machine learning algorithms for each specific dataset. AHP provides a way to incorporate expert's knowledge (i.e. healthcare providers' opinions) in classification algorithm selection, which has hardly been captured in previous studies. Considering expert's knowledge, pair-wise comparisons for attributes are placed in a matrix and the priority weights are calculated. Additionally, local consistency ratio of the attributes is obtained to assure the consistency of the subjective judgments. The first step to calculate local consistency ratio is to multiply the matrix of comparisons (matrix A) by the priority weights (matrix B) (see Eq.1). Next, "D" is calculated by dividing matrix C by matrix B (see Eq. 2). Then, the maximum eigenvalue (λ_{max}) is calculated by taking the average of the elements of matrix D (see Eq. 3). Next, the consistency Index (CI) is obtained by the formula shown in Eq. 4; parameter n denotes the number of attributes. Finally, the Consistency Ratio (CR) is calculated by dividing CI to Random Index (RI). RI is developed by Saaty [23] for different values of n. Here, RI equals to 0.58. Consistency Ratio (CR) of the attributes is calculated as 0.07 which is less than Saaty's proposed empirical limit (0.1). Thus, it can be concluded that the subjective judgments are consistent.

$$\begin{matrix}[A] & & & [B] & [C]\end{matrix}$$
$$\begin{vmatrix} 1 & 7 & 9 \\ 0.143 & 1 & 3 \\ 0.111 & 0.333 & 1 \end{vmatrix} \times \begin{vmatrix} 0.776 \\ 0.155 \\ 0.068 \end{vmatrix} = \begin{vmatrix} 2.477 \\ 0.471 \\ 0.206 \end{vmatrix} \quad (1)$$

$$D = \left| \frac{2.477}{0.776} \quad \frac{0.471}{0.155} \quad \frac{0.206}{0.068} \right| = |3.19 \quad 3.043 \quad 3.013| \quad (2)$$

$$\lambda_{max} = \frac{3.19 + 3.043 + 3.013}{3} = 9.246 \quad (3)$$

$$CI = \frac{\lambda_{max} - n}{n - 1} = \frac{9.246 - 3}{2} = 0.041 \quad (4)$$

$$C.R. = \frac{CI}{RI} = \frac{0.041}{0.58} = 0.07 \quad (5)$$

The next step is to perform the calculations for alternative comparisons with respect to each of the attributes. The attributes including average accuracy, time required for training, and memory usage are quantified attributes so the priority weights for the alternatives can be achieved objectively. Therefore, the local CR of the quantified attributes equal to zero. Once the priority weights of the alternatives are determined, the overall priority weights of alternatives are calculated by multiplying the matrix of evaluation ratings to the vector of priority weights and summing all over the attributes. For calculating the global consistency ratio (C.R.H), the vectors of the third level consistency and ratio indices should be present. As mentioned earlier, CI for the quantified values are set to zero. Thus, the global consistency of the decision hierarchy is equal to 0.022 (see Eqs. 6-8). The final results of the AHP model are the ranked classification methods for 10 different medical datasets shown in Table 3.

$$M = second\ level\ CI + \left|\begin{matrix} Vector\ of\ second\ level \\ priority\ weights \end{matrix}\right| \times |Vector\ of\ third\ level\ CIs| = 0.041 + |0.776 \quad 0.155 \quad 0.068| \times \left|\begin{matrix} 0 \\ 0 \\ 0 \end{matrix}\right| = 0.04 \tag{6}$$

$$\bar{M} = second\ level\ RI + \left|\begin{matrix} Vector\ of\ second\ level \\ priority\ weights \end{matrix}\right| \times |Vector\ of\ third\ level\ RIs| = 0.58 + |0.776 \quad 0.155 \quad 0.068| \times \left|\begin{matrix} 1.24 \\ 1.24 \\ 1.24 \end{matrix}\right| = 1.82 \tag{7}$$

$$C.R.H = \frac{M}{\bar{M}} = \frac{0.041}{1.82} = 0.022 \tag{8}$$

Table 3. Summary of the AHP results for selecting the best classification algorithm for each of the 10 sample medical datasets

	First Choice	Second Choice
Breast cancer Wisconsin (Original)	SVM	NB
Breast Cancer Wisconsin (Diagnostic)	NB	SVM
Breast Cancer Wisconsin (Prognostic)	SVM	NB
Cardiotocography Data	NB	SVM
Diabetes Data	NB	SVM
Echocardiogram Data	SVM	NB
Heart Disease Data	NB	SVM
Parkinson's Data	NB	SVM
Hepatitis Data	SVM	LR
Urinary Disease Data	SVM	NB

The results indicate that SVM and NB are the dominating machine learning algorithms for analyzing the 10 sample medical datasets. Therefore, logistic regression was employed to identify the most desirable machine learning classification algorithm for the unknown medical dataset that will be used in CDSS. For this purpose, two main medical dataset properties (number of observations and number of features) were randomly sampled from their respective distribution (obtained from the properties of 10 sample medical datasets). Next, logistic regression was trained using the datasets in Table 3 and the corresponding first choice classification algorithm. Finally, the trained logistic regression model was tested on the sampled dataset properties. The results of logistic regression using 1000 sampled dataset properties indicate that SVM has a higher potential of providing best results for unknown medical datasets. The results are consistent with previous studies that show SVM provides a very good classification result in wide variety of datasets [9]. The logistic regression was employed here because it also provides the probability of the likelihood of one observation belonging to specific class. Fig 4 shows a bar chart of the results.

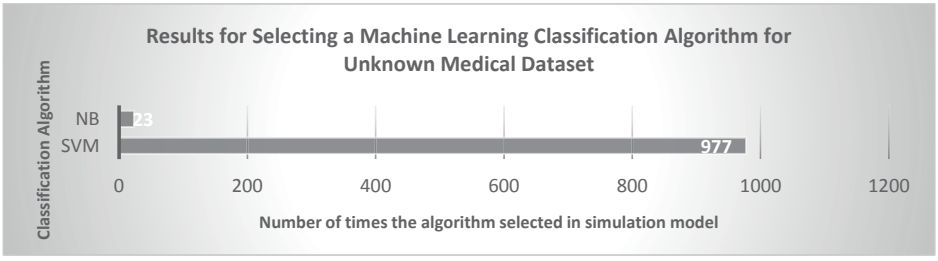

Fig. 4. Results for selecting a machine learning algorithm for unknown medical dataset

4. Conclusions

In this study, a meta-learning algorithm is proposed to choose a machine learning classification algorithm that can be used for developing clinical decision support system (CDSS). Using ten sample medical datasets, the proposed model suggest SVM as the most desirable classification algorithm for developing CDSS. The aim of this research was not to identify a classification algorithm that has the best performance in all medical datasets; rather, the aim was to identify an algorithm that has the potential to perform well in most medical datasets. The reliability of the model can be improved using more sample datasets; however, good quality accessible datasets that can be used for this purpose are limited. In our future work, we will consider more characteristics of the datasets such as sparsity, noisiness and etc. to improve the reliability of the model.

References

1. Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in medicine* 2001; 23:89-109.
2. Bocklitz T, Melanie P, Carsten S, Josef K, Axel N, Petra R, and Jürgen P. A comprehensive study of classification methods for medical diagnosis. *Journal of Raman Spectroscopy* 2009; 40:1759-1765.
3. Tan AC, and David G. Ensemble machine learning on gene expression data for cancer classification. *Appl. Bioinformatics* 2003; 2:75-83.
4. Zheng B, Sang WY, and Sarah SL. Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. *Expert Systems with Applications* 2014; 41:1476-1482.
5. Jain AK, Murty MN, and Flynn PJ. Data clustering: a review. *ACM computing surveys (CSUR)* 1999; 31:264-323.
6. Kotsiantis SB. Supervised machine learning: A review of classification techniques. Informatica 2007; 31:249-268.
7. Kotsiantis SB, Kanellopoulos D, and Pintelas PE. Data preprocessing for supervised leaning. *International Journal of Computer Science* 2006; 1:111-117.
8. Guyon I, and André E. An introduction to variable and feature selection. *The Journal of Machine Learning Research* 2003; 3:1157-1182.
9. Shawkat A, and Smith KA. On learning algorithm selection for classification. *Applied Soft Computing* 2006; 6:119-138.
10. Brodley CE. Addressing the selective superiority problem: Automatic algorithm/model class selection. *Proceedings of the Tenth International Conference on Machine Learning* 1993.
11. Tanwani AK, Afridi MJ, Shafiq MZ, and Farooq M. Guidelines to select machine learning scheme for classification of biomedical datasets. *Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics* 2009; 5483:128-139.
12. Rice JR. The algorithm selection problem. *Advances in Computers* 1975; 15:65-118.
13. Brazdil PB, and Soares C. A comparison of ranking methods for classification algorithm selection. *Machine Learning: ECML 2000* 2000; 1810:63-75.
14. Demšar J. Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine Learning Research* 2006; 7:1-30.
15. Petrak J. Fast subsampling performance estimates for classification algorithm selection. *Proceedings of the ECML-00 Workshop on Meta-Learning: Building Automatic Advice Strategies for Model Selection and Method Combination* 2000; 3-14.
16. Smith-Miles KA. Cross-disciplinary perspectives on meta-learning for algorithm selection. *ACM Computing Surveys (CSUR)* 2008; 41:6(1):6(25)
17. Rendell L, and Cho H. Empirical learning as a function of concept character. *Machine Learning* 1990; 5:267-298.
18. Aha DW. Generalizing from Case studies: A Case Study. *ML*; 1-10.
19. Michie D, Spiegelhalter DJ, and Taylor CC. Machine learning, neural and statistical classification. In:*Machine Learning*. Ellis Horwood; 1994. p. 1297-1300.
20. Lim TS, Loh WY, and Shih YS. A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms. *Machine learning* 2000; 40:203-228.
21. Hilario M, and Kalousis A. Building algorithm profiles for prior model selection in knowledge discovery systems. *International journal of engineering intelligent systems for electrical engineering and communications* 2000; 8:77-88.
22. Vaidya OS, and Kumar S. Analytic hierarchy process: An overview of applications. *European Journal of operational research* 2006; 48:1-29.
23. Saaty T. How to make a decision: the analytic hierarchy process. *European journal of operational research* 1990; 48:9-26.
24. Lin ZC, and Yang CB. Evaluation of machine selection by the AHP method. *Journal of Materials Processing Technology* 1996; 57:253-258.
25. Zahedi F. The analytic hierarchy process-a survey of the method and its applications. *interfaces* 1986; 16:96-108.
26. UCI machine learning repository. Available at: http://archive.ics.uci.edu/ml/