Ranking and Selecting Clustering Algorithms Using a Meta-Learning Approach

Marcilio C. P. de Souto, Ricardo B. C. Prudencio, Rodrigo G. F. Soares
Daniel A. S. Araujo, Ivan G. Costa, Teresa B. Ludermir, Alexander Schliep

***Abstract*— We present a novel framework that applies a meta-learning approach to clustering algorithms. Given a dataset, our meta-learning approach provides a ranking for the candidate algorithms that could be used with that dataset. This ranking could, among other things, support non-expert users in the algorithm selection task. In order to evaluate the framework proposed, we implement a prototype that employs regression support vector machines as the meta-learner. Our case study is developed in the context of cancer gene expression microarray datasets.**

I. Introduction

In several domains, such as in Machine Learning, there is a variety of algorithms that can be considered as candidates to solve particular problems. One of the most difficulty tasks in these domains is to predict when one algorithm is better than another to solve a given problem [1]. Traditional approaches to predicting the performance of algorithms often involve costly trial-and-error procedures [2]. Other approaches require expert knowledge, which is not always straightforward to acquire.

In the previous context, meta-learning approaches have arisen as effective solutions, able to automatically predicting algorithms performance for a given problem [2], [3], [4]. Thus, such approaches could support non-expert users in the algorithm selection task. As pointed out in [3], there are different interpretations for the term "meta-Learning". In our work, we use "meta-learning" meaning the automatic process of generating knowledge that relates the performance of machine learning algorithms to the characteristics of the problem (i.e., characteristics of its datasets).

So far, in the literature, meta-learning has been used only for selecting/ranking supervised learning algorithms [2], [1], [4]. That is, up to now, there no such an approach for the context of clustering algorithms (i.e., unsupervised learning). Motivated by this, we extend the use of meta-learning approaches for clustering algorithms. We develop our case study in the context of clustering algorithms applied to cancer gene expression data generated by microarray.

Cluster analysis techniques of gene expression microarray data is of increasing interest in the field of functional genomics [5], [6], [7]. One of the reasons for this is the need for molecular-based refinement of broadly defined biological classes, with implications in cancer diagnosis, prognosis and treatment. Although the choice of the clustering method for the analysis of microarray datasets is a very important issue, there are in the literature few guidelines or standard procedures on how these data should be analyzed [8].

The choice of algorithms are basically driven by the familiarity of biological experts to the algorithm rather than the characteristics of the algorithms themselves and of the data [8]. For example, the wide use of hierarchical clustering methods is mostly a consequence of its similarity to phylogenetic methods, which biologists are often acquaint to. Thus, in this context, by using a meta-learning approach, our aim is to provide a framework to support non-expert users in the algorithm selection task.

The remain of this paper is divided into four sections. Section II introduces a brief explanation about meta-learning and some of its techniques. In Section III, we present our meta-learning proposal to rank and select clustering algorithms. Section IV introduce our case study. We describe in Section V the experiments that we developed in order to evaluate the performance of our prototype. Finally, in Section VI, we present some final remarks and further work.

II. Related work

As pointed out before, in this work, we use the term meta-learning meaning the automatic process of obtaining knowledge that relates the performance of learning algorithms to the characteristics of the learning problems [2]. In such a context, each *meta-example* corresponds to a learning task and is composed of: (1) the features describing the problem, called *meta-features* or *meta-attributes*; and (2) the information about the performance of one or more algorithms when applied to the problem.

The *meta-learner* is a learning system that receives as input a set of these meta-examples and, from them, acquires knowledge that will be used to predict the performance of the algorithms for new problems. In the context of selecting supervised learning algorithms, the meta-attributes are, in general, statistics describing the training dataset of the problem. Examples of these statistics are: number of training examples, number of attributes, correlation between attributes, class entropy, among others [9], [1], [10].

In a more strict formulation of meta-learning, each meta-example has, as performance information, a class attribute that indicates the best algorithm for the problem, among a set of candidates [11], [12], [13], [4], [14]. In such a case, the class label for each meta-example is defined by performing a

M. C. P. de Souto and D. S. A. Araujo are with the Department of Informatics and Applied Mathematics, Fed. Univ. of Rio Grande do Norte, Natal, Brazil, email: marcilio@dimap.ufrn.br. R. B. C. Prudêncio, R. G. F. Soares and T. B. Ludermir are with the Center of Informatics, Federal University of Pernambuco, Recife, Brazil. I. G. Costa and Alexander Schliep are with Max Planck Institute for Molecular Genetics, Berlin, Germany.

cross-validation experiment using the available dataset. The meta-learner is simply a classifier which predicts the best algorithm based on the meta-attributes of the problem.

In order to add new functionalities for the meta-learning process, other approaches have been proposed. In [15], [16], for instance, a set of different meta-learners is employed not only to predict a class label associated to the performance of the algorithms, but also to recommend a ranking of the algorithms. In such a framework, a 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, points are given to the algorithms according to the outputs. For example, if "X" is the output of meta-learner (X, Y), then algorithm X is credited with one point. The ranking of algorithms is recommended for the new problem directly from the number of points assigned to the algorithms.

In contrast to the previous approach, in [17], [18] one tries to directly predict the accuracy (or alternatively the error) of each candidate algorithm. The meta-learner in this case can be used either to select the algorithm with the highest predicted accuracy or to provide a ranking of algorithms based on the order of predicted accuracies. In [18], for instance, the authors obtained good results when a linear regression model was used to predict the accuracy of 8 different classification algorithms.

Another interesting approach for meta-learning in the one in [19]. In that work, the performance of the candidate algorithms is related to the performance obtained by simpler and faster designed learners, called *landmarkers*. The authors claim that some widely used meta-attributes are very time consuming. Thus, landmarking would be an economic approach to the characterization of learning problems and to provide useful information for the meta-learning process.

The concepts and techniques of meta-learning have been mainly evaluated in the context of select the best algorithms for classification tasks [2]. However, in recent years, they have been extended to other domains of application, such as in the selection of time series forecasting models [4] and in the design of planning systems [20]. Particulary, in this paper, we extend these concepts for the context of unsupervised learning.

III. Our approach

A. General Architecture

Figure 1 illustrates the general architecture of systems used for, given a dataset, ranking the candidates algorithms. As it is usual in Machine Learning, the system has two phases: training and use. In the training phase, the Meta-Learner (ML) acquires knowledge from the set of meta-examples stored in the Database (DB). This knowledge associates characteristics of the data to the performance of the candidate algorithms. In our case, these are clustering algorithms.

In the phase of use, given a new dataset, the Feature Extractor (FE) generate the values of the meta-attributes that describe these data. According to these values, the Meta-Learner (ML) module produces a ranking of the available

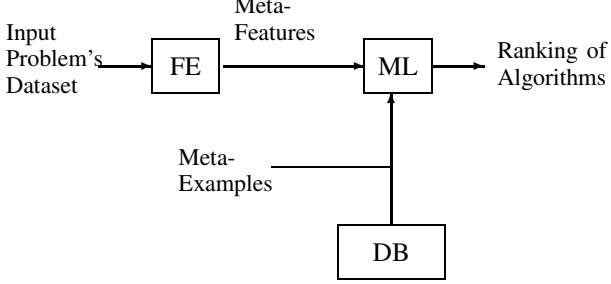

Fig. 1. System's architecture.

candidate algorithms. In order to so, it uses the knowledge previously provided as a result of the training phase.

The DB stores examples of datasets (i.e., meta-examples) used in the training phase. Each meta-example associates a dataset (represented by the chosen set of meta-attributes) to the performance of the candidate algorithms in clustering that data. This set of meta-examples is semi-automatically built: (1) the selection of datasets and algorithms to be considered is a manual task (as usual); (2) the generation of the meta-attributes is automatically performed by the FE module; and (3) the performance of the candidate algorithms in clustering each dataset is empirically obtained by directly applying each algorithm to the data and evaluating the obtained result.

The ML module implements the chosen meta-learning approach to acquiring knowledge (training phase) to be used in the selection or ranking of the candidate algorithms (use phase). As seen in Section II, the meta-learning approaches implement one or more machine learning algorithms to perform these tasks. In this context, we could use a learning technique to suggest one single algorithm from the set of candidate ones. Although this is a valuable approach, a more informative and flexible solution for algorithm selection is to provide a ranking of the candidate algorithms to each dataset under analysis [10]. In such a context, if enough resources are available, more than one algorithm could be used to cluster the data. Also, if the user has some preference for a specific subset of candidate algorithms, he/she can select the algorithms that obtained the best rank among the algorithms of interest.

B. Implementation Issues

In order to implement a system according to the architecture described in the previous section, one has to take into account some important issues. The first issue to be addressed is the type of dataset to be considered, since it will have an impact on all the other aspects in the system's implementation. In the case of this paper, as mentioned before, we consider as case study datasets regarding various types of cancer generated from microarray data.

Next, we need to specify which clustering algorithms will be considered to form the set of candidate algorithms. In this paper, seven clustering algorithms are employed to generate the candidate solutions. These are the single linkage, complete linkage, average linkage, k-means, mixture model

clustering, spectral clustering, and shared nearest neighbors algorithm [21], [22], [23]. These algorithms have been chosen to provide a wide range of recovery effectiveness, as well as to give some generality to the results.

The third issue to be considered is which features will be used by the FE module to describe the datasets. Such a choice depends on the type of dataset being analyzed. For example, in the context of classification problems, we can find standard sets of meta-attributes that have been used in the meta-learning area. This is the case of the *Data Characterization Tool*, developed within the METAL project[1]. In contrast, for cluster analysis, there is no such standard set of attributes, since the application of meta-learning to this domain is new. Nevertheless, we can follow some general guidelines to define them. For instance, one should choose meta-attributes that can be reliably identified, avoiding subjective analysis, such as visual inspection of plots. Subjective feature extraction is time consuming, requires expertise, and has a low degree of reliability [24]. One should also use a manageable number of features in order to avoid a time consuming selection process.

The final issue to be addressed in our work is which meta-learning approach will be used in the ML module. This choice depends upon the user's requirements, since each meta-learning approach has its advantages and limitations. In this work, we will present results with the approach that provides a ranking of the candidate algorithms to each dataset under analysis.

IV. Case Study

We focus on the problem of selecting algorithms for clustering cancer gene expression data. According to what has been defined in Section III-B, as our case study, we consider seven candidate algorithms: single linkage (SL), complete linkage (CL), average linkage (AL), k-means (KM), mixture model clustering (M), spectral clustering (SP), and Shared Nearest Neighbors algorithm (SNN) [21], [22], [23].

We implemented a prototype that according the architecture introduced in Section III-A. In the next sections, we present some relevant details about the three architecture's modules: the Feature Extractor, the Meta-Learner and the Database.

A. The Feature Extractor

We use a set of eight descriptive attributes (meta-attributes). Some of them were first proposed for the case of supervised learning tasks [25].

1) LgE: $\log_{10}$ of the number of examples. A raw indication of the available amount of training data.
2) LgREA: $\log_{10}$ of the ratio of the number of examples by the number of attributes. A rough indicator of the number of examples available to the number of attributes.
3) PMV: percentage of missing values. An indication of the quality of the data.
4) MN: multivariate normality, which is the proportion of T^2 [26](examples transformed via T^2) that are within 50% of a Chi-squared distribution (degree of freedom equals to the number of attributes describing the example). A rough indicator on the approximation of the data distribution to a normal distribution.
5) SK: skewness of the T^2 vector. Same as the previous item.
6) Chip: type of microarray technology used (either cDNA or Affymetrix) - see Section V.
7) PFA: percentage of the attributes that were kept after the application of the attribute selection filter.
8) PO: percentage of outliers. In this case, the values stands for the proportion of T^2 distant more than two standard deviations from the mean. Another indicator of the quality of the data.

As this set is possibly not optimal, in future implementations we will consider new features.

B. Meta-Learner

Our system generates a ranking of algorithms for each dataset given as input. In order to generate a ranking of P candidates (clustering algorithms), we use P regressors, each one responsible for predicting the ranking of a specific algorithm for the input dataset.

For constructing the regressor associated to a given algorithm i, we adopt the following procedure. First, we define a set of meta-examples. Each meta-example corresponds to a dataset, described by a set of meta-attributes, with one of them representing the desired output. The value of the meta-attribute representing the desired output is assigned according to the ranking of the algorithm among all the seven ones used to cluster the dataset. Next, we apply a supervised learning algorithm to each of the P regressors, which will be responsible for associating a dataset to a ranking.

As previously mentioned, we consider seven available clustering algorithms: SL, AL, CL, KM, M, SP and SNN. As a consequence, we build seven regressors, $R_1, \ldots, R_7$, associated to, respectively, SL, AL, CL, KM, M, SP and SNN. Now, suppose that the outputs of the seven regressors for a new dataset are, respectively, 7, 5, 6, 1, 2, 4 and 3. Such an output means, for instance, that model SL is expected to be the worst model (it is the last one in the ranking), AL is fifth best model model, CL the fourth one, KM is supposed to better than all the others, as it is placed as first one in the ranking.

In our implementation, we use the regression Support Vector Machine (SVM) algorithm, implemented in LIBSVM: a library for support vector machines [27]. A reason for this choice is that, in our preliminary results, SVMs showed a better accuracy than models such as neural networks and k-NN.

C. The Database

The Database stores meta-examples regarding cancer gene expression microarray datasets. Each meta-example has two parts: (1) the meta-attributes describing the gene expression

[1] http://www.cs.bris.ac.uk/~cgc/METAL

data, which are those presented in Section IV-A; and (2) a vector with the ranking of each clustering algorithm for that dataset. In order to assign this ranking for a dataset, we run each of the seven clustering algorithms with the non-normalized version of the dataset to produce the respective partitions. The number of clusters is set to be equal to the true number of the classes in the data. The known class labels are not used in any way during the clustering. For all non-deterministic algorithms, we run the algorithm 30 times. Then, for further analysis, we pick the partition with the best corrected Rand index.

In fact, in terms of the index to measure the success of the algorithm in recovering the true partition of the dataset and build the ranking, we also employ the corrected Rand index (cR) [21], [28]. The cR can take values from -1 to 1, with 1 indicating a perfect agreement between the partitions, and the values near 0 or negatives corresponding to cluster agreement found by chance.

Formally, let $U = \{u_1, \ldots, u_r, \ldots, u_R\}$ be the partition given by the clustering solution, and $V = \{v_1, \ldots, v_c, \ldots, v_C\}$ be the partition formed by an a priori information independent of partition U (the gold standard). The corrected Rand is defined as:

$$cR = \frac{\sum_i^R \sum_j^C \binom{n_{ij}}{2} - \binom{n}{2}^{-1} \sum_i^R \binom{n_{i\cdot}}{2} \sum_j^C \binom{n_{\cdot j}}{2}}{\frac{1}{2}[\sum_i^R \binom{n_{i\cdot}}{2} + \sum_j^C \binom{n_{\cdot j}}{2}] - \binom{n}{2}^{-1} \sum_i^R \binom{n_{i\cdot}}{2} \sum_j^C \binom{n_{\cdot j}}{2}}$$

where (1) n_{ij} represents the number of objects in clusters u_i and v_j; (2) $n_{i\cdot}$ indicates the number of objects in cluster u_i; (3) $n_{\cdot j}$ indicates the number of objects in cluster v_j; (4) n is the total number of objects; and (5) $\binom{a}{b}$ is the binomial coefficient $\frac{a!}{b!(a-b)!}$.

Based on the values of the cR, the ranking for the algorithms is generated as follows. The clustering algorithm that presents the highest cR come higher in the ranking. Algorithms that generate partition with the same cR receive the same ranking number, which is the mean of what they would have under ordinal rankings.

V. Experimental Design

A. Description of the Datasets

We describe here the experiments that we developed in order to evaluate the performance of our prototype. Thirty two microarray datasets are included in this analysis (see Table I). These datasets present different values for characteristics such as type of microarray chip (second column), number of patterns (third column), number of classes (fourth column), distribution of patterns within the classes (fifth column), dimensionality (sixth column), and dimensionality after feature selection (last column).

In terms of the datasets, it is important to point out that microarray technology is usually available in two different platforms, cDNA and Affymetrix [5], [6], [7]. Measurements of Affymetrix arrays are estimates on the number of RNA copies found in the cell sample, while cDNA microarrays values are ratios of the number of copies in relation to a control cell sample.

In the case of Affymetrix, following other works, for our datasets, all genes with expression level below to 10 are set to the minimum threshold of 10. The maximum threshold is set at 16,000. Values below or above these thresholds are often not reliable [5], [29], [30]. That is, our analysis is performed on the scaled data to which the ceiling and threshold values have been applied.

Furthermore, in order to remove uninformative genes for the case of Affymetrix arrays, we apply the following procedure. For each gene j (attribute), we compute the mean m_j. But before doing so, in order to get rid of extreme values, we discard the 10% largest and smallest values. Based on this mean, we transform every value x^*_{ij} of example i and attribute j to:

$$y_{ij} = \log_2(x^*_{ij}/m_j)$$

After the previous transformation, we select for further analysis genes whose expression level differed by at least l-fold, in at least c samples, from their mean expression level across samples. With few exceptions, the parameters l and c were chosen in such a way as to yield a filtered dataset with around at least 10% of the original number of genes (features). It is important to point out that the data transformed with the previous equation is only used in the filtering step.

A similar filter procedure was applied for the case of cDNA microarray, but without the need to transform the data. In the case of cDNA microarray datasets, whose attributes (genes) could present missing values, we discard the ones with more than 10% of missing values. The attributes that are kept and still present missing values have the values replaced for the respective mean value of the attribute.

B. System Performance

For a given dataset, in order to generate the ranking, we considered the configuration that obtained the best corrected Rand. We executed the algorithms with Euclidean distance, Pearson correlation and Cosine, but always with the number of cluster set to the real number of classes in the dataset.

We evaluate the performance of the meta-learners using the leave-one-out procedure. At each step, 31 examples are used as the training set, and the remaining example is used to test the generated SVMs. This step is repeated 32 times, using at each time a different test example.

The quality of a suggested ranking for a given dataset is evaluated by measuring the similarity to the ideal ranking, which represents the correct ordering of the models according to the corrected Rand. In our work, we used the Spearman's rank correlation coefficient (see [10]) to measure the similarity between a suggested and the ideal rankings.

Given a dataset i, we calculate the squared difference between the suggested and the ideal rankings for each algorithm j (D^2_{ij}). Then, we compute the sum of these differences for all algorithms:

TABLE I
DATASET DESCRIPTION

Dataset	**Chip**	n	**Nr. Classes**	**Dist. Classes**	d	**Filtered** d
Alizadeh-V1 [31]	cDNA	42	2	21,21	4022	1095
Alizadeh-V2 [31]	cDNA	62	3	42,9,11	4022	2093
Armstrong-V1 [32]	Affy	72	2	24,48	12582	1081
Armstrong-V2 [32]	Affy	72	3	24,20,28	12582	2194
Bhattacharjee [33]	Affy	203	5	139,17,6,21,20	12600	1543
Bittner [34]	cDNA	38	2	19, 9	8067	2201
Bredel [35]	cDNA	50	3	31,14,5	41472	1739
Chen [36]	cDNA	180	2	104,76	22699	85
Chowdary [37]	Affy	104	2	62,42	22283	182
Dyrskjot [38]	Affy	40	3	9,20,11	7129	1203
Garber [39]	cDNA	66	4	17,40,4,5	24192	4553
Golub-V1 [40]	Affy	72	2	47,25	7129	1877
Gordon [41]	Affy	181	2	31,150	12533	1626
Khan [42]	cDNA	83	4	29,11,18,25	6567	1069
Laiho [43]	Affy	37	2	8,29	22883	2202
Lapoint-V1 [44]	cDNA	69	3	11,39,19	42640	1625
Lapoint-V2 [44]	cDNA	110	4	11,39,19,41	42640	2496
Liang [45]	cDNA	37	3	28,6,3	24192	1411
Nutt-V1 [46]	Affy	50	4	14,7,14,15	12625	1377
Nutt-V2 [46]	Affy	28	2	14,14	12625	1070
Nutt-V3 [46]	Affy	22	2	7,15	12625	1152
Pomeroy-V1 [47]	Affy	34	2	25,9	7129	857
Pomeroy-V2 [47]	Affy	42	5	10,10,10,4,8	7129	1379
Ramaswamy [29]	Affy	190	14	11,10,11,11,22,10,11,10,30,11,11,11,11,20	16063	1363
Risinger [48]	cDNA	42	4	13,3,19,7	8872	1771
Shipp [49]	Affy	77	2	58,19	7129	798
Singh [50]	Affy	102	2	58,19	12600	339
Su [51]	Affy	174	10	26,8,26,23,12,11,7,27,6,28	12533	1571
Tomlins-V1 [52]	cDNA	104	5	27,20,32,13,12	20000	2315
Tomlins-V2 [52]	cDNA	92	4	27,20,32,13	20000	1288
West [53]	Affy	49	2	25,24	7129	1198
Yeoh-V1 [54]	Affy	248	2	43,205	12625	2526

$$D_i^2 = \sum_j D_{ij}^2 \tag{1}$$

Finally, the Spearman coefficient is calculated using the equation:

$$SRC_i = 1 - \frac{6 * D_i^2}{P^3 - P}, \tag{2}$$

where P is the number of candidate algorithms. The value of this coefficient ranges from $[-1, 1]$. The larger is the value of SRC_i, the greater is the similarity between the suggested and the ideal rankings for the dataset i.

In order to evaluate the rankings generated for datasets in the test set, we calculated the average of the Spearman's correlation for all these datasets (Equation 3).

$$SRC = \frac{1}{32} * \sum_{i \in test} SRC_i \tag{3}$$

The result of our approach was compared to a default ranking method, where the average ranking is suggested for all datasets. In our case, the default ranking is: SL=6.41, AL=4.60, CL=3.84, KM=2.31, M=3.40, SP=3.07, SNN=4.36. In Table II, we show the mean and standard deviation for the Spearman coefficient for the rankings generated by our approach and for the default ranking.

As it can be seen, the rankings generated by our method were more correlated to the ideal ranking. In fact, according to a hypothesis test, at a significance level of 0.05, the mean of the correlation value found with our method was significantly higher than that obtained with the default ranking.

VI. FINAL REMARKS

In this work, we proposed a new approach to providing knowledge for the selection of clustering algorithms. We can point out contributions of this work to two different fields: (1) we applied meta-leaning concepts to a problem which had

TABLE II
MEAN OF THE SPEARMAN COEFFICIENT

Method	SRC
Default	0.59 ± 0.37
Meta-Leaner	0.75 ± 0.21

not yet been tackled: clustering algorithm selection; and (2) in terms of cluster analysis, we provided a novel framework to support non-expert users in the algorithm selection task.

In order to evaluate our framework, we developed a case study in the context of cancer gene expression microarray datasets. The experiment performed revealed good results in that our method, compared to the default ranking, generated rankings that were more correlated to the ideal ranking. In fact, according to a hypothesis test, at a significance level of 0.05, the mean of the Spearman coefficient value found with our method was significantly higher than that obtained with the default ranking.

Finally, we would like to highlight that several works can be developed from our proposal by implementing other meta-learners for different categories of datasets, and by using other meta-learning approaches that have not yet been used in the algorithm selection problem.

REFERENCES

[1] A. Kalousis, J. Gama, and M. Hilario, "On data and algorithms - understanding inductive performance," *Machine Learning*, vol. 54, no. 3, pp. 275–312, 2004.

[2] C. Giraud-Carrier, R. Vilalta, and P. Brazdil, "Introduction to the special issue on meta-learning," *Machine Learning*, vol. 54, no. 3, pp. 187–193, 2004.

[3] R. Vilalta and Y. Drissi, "A perspective view and survey of meta-learning," *Journal of Artificial Intelligence Review*, vol. 18, no. 2, pp. 77–95, 2002.

[4] R. B. C. Prudêncio and T. B. Ludermir, "Meta-learning approaches to selecting time series models," *Neurocomputing*, vol. 61, pp. 121–137, 2004.

[5] S. Monti, P. Tamayo, J. Mesirov, and T. Golub, "Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data," *Machine Learning*, vol. 52, pp. 91–118, 2003.

[6] J. Quackenbush, "Computational analysis of cDNA microarray data," *Nature Reviews*, vol. 6, no. 2, pp. 418–428, 2001.

[7] D. Slonim, "From patterns to pathways: gene expression data analysis comes of age," *Nature Genetics*, vol. 32, pp. 502–508, 2002.

[8] P. D'haeseleer, "How does gene expression clustering work?" *Nat Biotech*, vol. 23, no. 12, pp. 1499–1501, Dec. 2005. [Online]. Available: http://dx.doi.org/10.1038/nbt1205-1499

[9] R. Engels and C. Theusinger, "Using a data metric for preprocessing advice for data mining applications," in *Proceedings of the 13th European Conference on Artificial Intelligence (ECAI-98)*, H. Prade, Ed. John Wiley & Sons, 1998, pp. 430–434.

[10] P. Brazdil, C. Soares, and J. da Costa, "Ranking learning algorithms: Using IBL and meta-learning on accuracy and time results," *Machine Learning*, vol. 50, no. 3, pp. 251–277, 2003.

[11] D. Aha, "Generalizing from case studies: A case study," in *Proceedings of the 9th International Workshop on Machine Learning*. Morgan Kaufmann, 1992, pp. 1–10.

[12] A. Kalousis and M. Hilario, "Representational issues in meta-learning," in *Proceedings of the 20th International Conferente on Machine Learning*, 2003, pp. 313–320.

[13] R. B. C. Prudêncio, T. B. Ludermir, and F. A. T. de Carvalho, "A modal symbolic classifier to select time series models," *Pattern Recognition Letters*, vol. 25, no. 8, pp. 911–921, 2004.

[14] R. Leite and P. Brazdil, "Predicting relative performance of classifiers from samples," in *Proceedings of the 22nd International Conference on Machine Learning*, 2005.

[15] A. Kalousis and T. Theoharis, "Noemon: Design, implementation and performance results of an intelligent assistant for classifier selection," *Intelligent Data Analysis*, vol. 3, no. 5, pp. 319–337, 1999.

[16] A. Kalousis and M. Hilario, "Feature selection for meta-learning," *Lecture Notes in Computer Science*, vol. 2035, pp. 222–233, 2001.

[17] C. Koepf, C. C. Taylor, and J. Keller, "Meta-analysis: Data characterisation for classification and regression on a meta-level," in *Proceedings of the International Symposium on Data Mining and Statistics*, 2000.

[18] H. Bensusan and K. Alexandros, "Estimating the predictive accuracy of a classifier," in *Proceedings of the 12th European Conference on Machine Learning*, 2001, pp. 25–36.

[19] B. Pfahringer, H. Bensusan, and C. Giraud-Carrier, "Meta-learning by landmarking various learning algorithms," in *Proceedings of the 17th International Conference on Machine Learning, ICML'2000*. San Francisco, California: Morgan Kaufmann, 2000, pp. 743–750.

[20] G. Tsoumakas, D. Vrakas, N. Bassiliades, and I. Vlahavas, "Lazy adaptive multicriteria planning," in *Proceedings of the 16th European Conference on Artificial Intelligence, ECAI04*, 2004, pp. 693–697.

[21] A. K. Jain and R. C. Dubes, *Algorithms for clustering data*. Prentice Hall, 1988.

[22] R. Xu and D. Wunsch, "Survey of clustering algorithms," *IEEE Transactions on Neural Networks*, vol. 16, no. 3, pp. 645–678, 2005.

[23] L. Ertoz, M. Steinbach, and V. Kumar, "A new shared nearest neighbor clustering algorithm and its applications," in *Workshop on Clustering High Dimensional Data and its Applications*, 2002, pp. 105–115.

[24] J. A. M. Adya, F. Collopy and M. Kennedy, "Automatic identification of time series features for rule-based forecasting." *International Journal of Forecasting*, vol. 17, no. 2, pp. 143–157, 2001.

[25] D. Michie, D. J. Spiegelhalter, and C. C. Taylor, Eds., *Machine Learning, Neural and Statistical Classification*. Ellis Horwood, 1994.

[26] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, fifth edition ed. Prentice Hall, 2002.

[27] C.-C. Chang and C.-J. Lin, *LIBSVM: a library for support vector machines*, 2001, software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[28] G. W. Milligan and M. C. Cooper, "A study of standardization of variables in cluster analysis," *Journal of Classification*, vol. 5, pp. 181–204, 1988.

[29] S. Ramaswamy, P. Tamayo, R. Rifkin, S. Mukherjee, C. H. Yeang, M. Angelo, C. Ladd, M. Reich, E. Latulippe, J. P. Mesirov, T. Poggio, W. Gerald, M. Loda, E. S. Lander, and T. R. Golub, "Multiclass cancer diagnosis using tumor gene expression signatures." *Proc Natl Acad Sci U S A*, vol. 98, no. 26, pp. 15 149–15 154, Dec 2001. [Online]. Available: http://dx.doi.org/10.1073/pnas.211566398

[30] K. Stegmaier, K. N. Ross, S. A. Colavito, S. OMalley, B. R. Stockwell, and T. R. Golub, "Gene expression-based high-throughput screening(ge-hts) and application to leukemia differentiation," *Nature Genetics*, vol. 36, no. 3, pp. 257–263, 2004.

[31] A. A. Alizadeh, M. B. Eisen, R. E. Davis, C. Ma, I. S. Lossos, A. Rosenwald, J. C. Boldrick, H. Sabet, T. Tran, X. Yu, J. I. Powell, L. Yang, G. E. Marti, T. Moore, J. Hudson, L. Lu, D. B. Lewis, R. Tibshirani, G. Sherlock, W. C. Chan, T. C. Greiner, D. D. Weisenburger, J. O. Armitage, R. Warnke, R. Levy, W. Wilson, M. R. Grever, J. C. Byrd, D. Botstein, P. O. Brown, and L. M. Staudt, "Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling," *Nature*, vol. 403, pp. 503–511, 2000.

[32] S. A. Armstrong, J. E. Staunton, L. B. Silverman, R. Pieters, M. L. den Boer, M. D. Minden, S. E. Sallan, E. S. Lander, T. R. Golub, and S. J. Korsmeyer, "Mll translocations specify a distinct gene expression profile that distinguishes a unique leukemia." *Nat Genet*, vol. 30, no. 1, pp. 41–47, Jan 2002. [Online]. Available: http://dx.doi.org/10.1038/ng765

[33] A. Bhattacharjee, W. G. Richards, J. Staunton, C. Li, S. Monti, P. Vasa, C. Ladd, J. Beheshti, R. Bueno, M. Gillette, M. Loda, G. Weber, E. J. Mark, E. S. Lander, W. Wong, B. E. Johnson, T. R. Golub, D. J. Sugarbaker, and M. Meyerson, "Classification of human lung carcinomas by mrna expression profiling reveals distinct adenocarcinoma subclasses." *Proc Natl Acad Sci U S A*, vol. 98, no. 24, pp. 13 790–13 795, Nov 2001. [Online]. Available: http://dx.doi.org/10.1073/pnas.191502998

[34] M. Bittner, P. Meltzer, Y. Chen, Y. Jiang, E. Seftor, M. Hendrix, M. Radmacher, R. Simon, Z. Yakhini, A. Ben-Dor, N. Sampas, E. Dougherty, E. Wang, F. Marincola, C. Gooden, J. Lueders, A. Glatfelter, P. Pollock, J. Carpten, E. Gillanders, D. Leja, K. Dietrich, C. Beaudry, M. Berens, D. Alberts, and V. Sondak, "Molecular classification of cutaneous malignant melanoma by gene expression profiling." *Nature*, vol. 406, no. 6795, pp. 536–540, Aug 2000. [Online]. Available: http://dx.doi.org/10.1038/35020115

[35] M. Bredel, C. Bredel, D. Juric, G. R. Harsh, H. Vogel, L. D. Recht, and B. I. Sikic, "Functional network analysis reveals extended gliomagenesis pathway maps and three novel myc-interacting genes in human gliomas." *Cancer Res*, vol. 65, no. 19, pp. 8679–8689, Oct 2005. [Online]. Available: http://dx.doi.org/10.1158/0008-5472.CAN-05-1204

[36] X. Chen, S. T. Cheung, S. So, S. T. Fan, C. Barry, J. Higgins, K.-M. Lai, J. Ji, S. Dudoit, I. O. Ng, M. van de Rijn, D. Botstein, and P. O. Brown, "Gene expression patterns in human liver cancers," *Mol. Biol. Cell*, vol. 13, no. 6, pp. 1929–1939, 2002. [Online]. Available: http://www.molbiolcell.org/cgi/content/abstract/13/6/1929

[37] D. Chowdary, J. Lathrop, J. Skelton, K. Curtin, T. Briggs, Y. Zhang, J. Yu, Y. Wang, and A. Mazumder, "Prognostic gene expression signatures can be measured in tissues collected in rnalater preservative." *J Mol Diagn*, vol. 8, no. 1, pp. 31–39, Feb 2006.

[38] L. Dyrskjt, T. Thykjaer, M. Kruhffer, J. L. Jensen, N. Marcussen, S. Hamilton-Dutoit, H. Wolf, and T. F. Orntoft, "Identifying distinct classes of bladder carcinoma using microarrays." *Nat Genet*, vol. 33, no. 1, pp. 90–96, Jan 2003. [Online]. Available: http://dx.doi.org/10.1038/ng1061

[39] M. E. Garber, O. G. Troyanskaya, K. Schluens, S. Petersen, Z. Thaesler, M. Pacyna-Gengelbach, M. van de Rijn, G. D. Rosen, C. M. Perou, R. I. Whyte, R. B. Altman, P. O. Brown, D. Botstein, and I. Petersen, "Diversity of gene expression in adenocarcinoma of the lung." *Proc Natl Acad Sci U S A*, vol. 98, no. 24, pp. 13 784–13 789, Nov 2001. [Online]. Available: http://dx.doi.org/10.1073/pnas.241500798

[40] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, and E. S. Lander, "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring." *Science*, vol. 286, no. 5439, pp. 531–537, Oct 1999.

[41] G. J. Gordon, R. V. Jensen, L.-L. Hsiao, S. R. Gullans, J. E. Blumenstock, S. Ramaswamy, W. G. Richards, D. J. Sugarbaker, and R. Bueno, "Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma." *Cancer Res*, vol. 62, no. 17, pp. 4963–4967, Sep 2002.

[42] J. Khan, J. S. Wei, M. Ringnr, L. H. Saal, M. Ladanyi, F. Westermann, F. Berthold, M. Schwab, C. R. Antonescu, C. Peterson, and P. S. Meltzer, "Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks." *Nat Med*, vol. 7, no. 6, pp. 673–679, Jun 2001. [Online]. Available: http://dx.doi.org/10.1038/89044

[43] P. Laiho, A. Kokko, S. Vanharanta, R. Salovaara, H. Sammalkorpi, H. Jrvinen, J.-P. Mecklin, T. J. Karttunen, K. Tuppurainen, V. Davalos, S. Schwartz, D. Arango, M. J. Mkinen, and L. A. Aaltonen, "Serrated carcinomas form a subclass of colorectal cancer with distinct molecular basis." *Oncogene*, vol. 26, no. 2, pp. 312–320, Jan 2007. [Online]. Available: http://dx.doi.org/10.1038/sj.onc.1209778

[44] J. Lapointe, C. Li, J. P. Higgins, M. van de Rijn, E. Bair, K. Montgomery, M. Ferrari, L. Egevad, W. Rayford, U. Bergerheim, P. Ekman, A. M. DeMarzo, R. Tibshirani, D. Botstein, P. O. Brown, J. D. Brooks, and J. R. Pollack, "Gene expression profiling identifies clinically relevant subtypes of prostate cancer." *Proc Natl Acad Sci U S A*, vol. 101, no. 3, pp. 811–816, Jan 2004. [Online]. Available: http://dx.doi.org/10.1073/pnas.0304146101

[45] Y. Liang, M. Diehn, N. Watson, A. W. Bollen, K. D. Aldape, M. K. Nicholas, K. R. Lamborn, M. S. Berger, D. Botstein, P. O. Brown, and M. A. Israel, "Gene expression profiling reveals molecularly and clinically distinct subtypes of glioblastoma multiforme." *Proc Natl Acad Sci U S A*, vol. 102, no. 16, pp. 5814–5819, Apr 2005. [Online]. Available: http://dx.doi.org/10.1073/pnas.0402870102

[46] C. L. Nutt, D. R. Mani, R. A. Betensky, P. Tamayo, J. G. Cairncross, C. Ladd, U. Pohl, C. Hartmann, M. E. McLaughlin, T. T. Batchelor, P. M. Black, A. von Deimling, S. L. Pomeroy, T. R. Golub, and D. N. Louis, "Gene expression-based classification of malignant gliomas correlates better with survival than histological classification." *Cancer Res*, vol. 63, no. 7, pp. 1602–1607, Apr 2003.

[47] S. L. Pomeroy, P. Tamayo, M. Gaasenbeek, L. M. Sturla, M. Angelo, M. E. McLaughlin, J. Y. H. Kim, L. C. Goumnerova, P. M. Black, C. Lau, J. C. Allen, D. Zagzag, J. M. Olson, T. Curran, C. Wetmore, J. A. Biegel, T. Poggio, S. Mukherjee, R. Rifkin, A. Califano, G. Stolovitzky, D. N. Louis, J. P. Mesirov, E. S. Lander, and T. R. Golub, "Prediction of central nervous system embryonal tumour outcome based on gene expression." *Nature*, vol. 415, no. 6870, pp. 436–442, Jan 2002. [Online]. Available: http://dx.doi.org/10.1038/415436a

[48] J. I. Risinger, G. L. Maxwell, G. V. R. Chandramouli, A. Jazaeri, O. Aprelikova, T. Patterson, A. Berchuck, and J. C. Barrett, "Microarray analysis reveals distinct gene expression profiles among different histologic types of endometrial cancer." *Cancer Res*, vol. 63, no. 1, pp. 6–11, Jan 2003.

[49] M. A. Shipp, K. N. Ross, P. Tamayo, A. P. Weng, J. L. Kutok, R. C. T. Aguiar, M. Gaasenbeek, M. Angelo, M. Reich, G. S. Pinkus, T. S. Ray, M. A. Koval, K. W. Last, A. Norton, T. A. Lister, J. Mesirov, D. S. Neuberg, E. S. Lander, J. C. Aster, and T. R. Golub, "Diffuse large b-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning." *Nat Med*, vol. 8, no. 1, pp. 68–74, Jan 2002. [Online]. Available: http://dx.doi.org/10.1038/nm0102-68

[50] D. Singh, P. G. Febbo, K. Ross, D. G. Jackson, J. Manola, C. Ladd, P. Tamayo, A. A. Renshaw, A. V. D'Amico, J. P. Richie, E. S. Lander, M. Loda, P. W. Kantoff, T. R. Golub, and W. R. Sellers, "Gene expression correlates of clinical prostate cancer behavior." *Cancer Cell*, vol. 1, no. 2, pp. 203–209, Mar 2002.

[51] A. I. Su, J. B. Welsh, L. M. Sapinoso, S. G. Kern, P. Dimitrov, H. Lapp, P. G. Schultz, S. M. Powell, C. A. Moskaluk, H. F. Frierson, and G. M. Hampton, "Molecular classification of human carcinomas by use of gene expression signatures." *Cancer Res*, vol. 61, no. 20, pp. 7388–7393, Oct 2001.

[52] S. A. Tomlins, R. Mehra, D. R. Rhodes, X. Cao, L. Wang, S. M. Dhanasekaran, S. Kalyana-Sundaram, J. T. Wei, M. A. Rubin, K. J. Pienta, R. B. Shah, and A. M. Chinnaiyan, "Integrative molecular concept modeling of prostate cancer progression." *Nat Genet*, vol. 39, no. 1, pp. 41–51, Jan 2007. [Online]. Available: http://dx.doi.org/10.1038/ng1935

[53] M. West, C. Blanchette, H. Dressman, E. Huang, S. Ishida, R. Spang, H. Zuzan, J. A. Olson, J. R. Marks, and J. R. Nevins, "Predicting the clinical status of human breast cancer by using gene expression profiles." *Proc Natl Acad Sci U S A*, vol. 98, no. 20, pp. 11 462–11 467, Sep 2001. [Online]. Available: http://dx.doi.org/10.1073/pnas.201162998

[54] E.-J. Yeoh, M. E. Ross, S. A. Shurtleff, W. K. Williams, D. Patel, R. Mahfouz, F. G. Behm, S. C. Raimondi, M. V. Relling, A. Patel, C. Cheng, D. Campana, D. Wilkins, X. Zhou, J. Li, H. Liu, C.-H. Pui, W. E. Evans, C. Naeve, L. Wong, and J. R. Downing, "Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling." *Cancer Cell*, vol. 1, no. 2, pp. 133–143, Mar 2002.