Zoomed Ranking: Selection of Classification Algorithms Based on Relevant Performance Information*

Pavel B. Brazdil and Carlos Soares
LIACC/FEP, University of Porto
R. Campo Alegre 823, 4150 Porto, Portugal
{pbrazdil,csoares}@ncc.up.pt

June 16, 2000

The need for methods which would assist the user in selecting classification algorithms for a new problem has frequently been recognized as an important issue.

Previous meta-learning approaches to algorithm selection consist of suggesting one algorithm or a small group of algorithms that are expected to perform well on the given problem [1, 3]. We believe that a more informative and flexible solution is to provide rankings of the candidate algorithms [5, 2]. A ranking can be used to select just one algorithm, or more, if enough resources are available. The problem of constructing rankings can be seen as an alternative to other Machine Learning methods, such as classification and regression.

One reason why selecting appropriate classification algorithms is difficult is related to the issue of how to assess their performance. The evaluation measure that is commonly used in classification problems is error rate. However, it is often important to consider a *combination of several criteria* [5]. We may consider, for instance, time to learn and time to apply a learned model. With the increasing computational power available today, it may be argued that time is not so important. Still some classification algorithms may take weeks or months to run on the volumes of data that organizations commonly store nowadays, which may not be acceptable. We may also consider interpretability, since in many applications the decision model must be understood. Not much work has been done to incorporate these criteria into the KDD process.

Recently, several methods that generate rankings of algorithms based on their past performance have been developed with promising results. Some are based only on accuracy [2], others on accuracy and time [6]. These methods were tested using all available information, i.e. concerning all datasets used in

*A complete version of this paper can be found in `http://www.ncc.up.pt/~csoares/inv/papers/brazdil_00b.ps.gz`.

the past. However, considering the NFL theorem [7] we cannot expect that all that information is relevant for the problem at hand.

We, therefore, address the problem of algorithm selection by dividing it into two distinct phases. In the first one we identify a subset of relevant datasets. For that purpose we present a technique called *zooming*. It employs the k-Nearest Neighbor algorithm with a distance function based on a set of statistical, information theoretic and other dataset characterization measures to identify datasets that are similar to the one at hand. We have selected dataset measures that were previously used for the same purpose [4], although we do not expect that all have them are appropriate. Work is under way to select the most predictive subset of those measures.

In the second phase we proceed to construct a ranking on the basis of the performance information of the candidate algorithms on the selected datasets. We present the *adjusted ratio of ratios* ranking method. This method processes performance information on accuracy and total execution time. It is based on the ratio of success rate ratio and an adjusted time ratio:

$$ARR^{d_i}_{a_p,a_q} = \frac{\frac{SR^{d_i}_{a_p}}{SR^{d_i}_{a_q}}}{1 + \frac{\log\left(\frac{T^{d_i}_{a_p}}{T^{d_i}_{a_q}}\right)}{K_T}} \tag{1}$$

where $SR^{d_i}_{a_p}$ and $T^{d_i}_{a_p}$ are the success rate and time of algorithm a_p on dataset d_i, respectively, and K_T[1] is a user-defined value that determines the relative importance of time. We have devised a user-friendly way to set this parameter. It is approximated by $K_T = 1/X\%$, where X is the accuracy one is willing to trade for a 10 times speedup or slowdown. This is represented as 10x≅X%.

The formula may seem ad-hoc at first glance, but its form can be related to the ones used in other areas of science. We can look at the ratio of success rates, $SR^{d_i}_{a_p}/SR^{d_i}_{a_q}$, as a measure of the advantage, and the ratio of times, $T^{d_i}_{a_p}/T^{d_i}_{a_q}$, as a measure of the disadvantage of algorithm a_p relative to algorithm a_q on dataset d_i. The former can be considered a *benefit* while the latter a *cost*. Thus, by dividing a measure of the benefit by a measure of the cost, we assess the overall quality of an algorithm. A similar philosophy underlies the efficiency measure of Data Envelopment Analysis (DEA) that has been proposed for multicriteria evaluation of data mining algorithms [5].

We go beyond the previous approaches in that we provide a methodology not only for constructing rankings but also for their empirical evaluation. It is based on the correlation between the ranking generated by the method for a given dataset with an *ideal ranking* constructed from the performance of the algorithms on that dataset. The same evaluation methodology can be used by others in the search for the best possible ranking method(s).

[1]Here, to avoid confusion with the number of nearest-neighbors (k), we refer to the compromise between time and accuracy as K_T, rather than K, as in [6].

We report experiments varying the number of neighbors and the relative importance of accuracy and time. The results obtained are compared to results obtained by ARR without zooming to assess whether zooming brings any significant improvement. It appears that zooming improves the quality of the generated rankings, although the results obtained are not significantly different according to the Friedman's test.

In summary, our contributions are (1) exploiting rankings rather then classification or regression, (2) providing an evaluation methodology for ranking, (3) providing a way of combining success rate and time and (4) exploiting dataset characteristics to select relevant performance information prior to ranking.

References

[1] P. Brazdil, J. Gama, and B. Henery. Characterizing the applicability of classification algorithms using meta-level learning. In F. Bergadano and L. de Raedt, editors, *Proceedings of the European Conference on Machine Learning (ECML-94)*, pages 83–102. Springer-Verlag, 1994.

[2] P. Brazdil and C. Soares. A comparison of ranking methods for classification algorithm selection. In *Proceedings of the European Conference on Machine Learning ECML2000 (to Be Published)*, 2000.

[3] A. Kalousis and T. Theoharis. NOEMON: Design, implementation and performance results of an intelligent assistant for classifier selection. *Intelligent Data Analysis*, 3(5):319–337, November 1999.

[4] G. Lindner and R. Studer. AST: Support for algorithm selection with a CBR approach. In C. Giraud-Carrier and B. Pfahringer, editors, *Recent Advances in Meta-Learning and Future Work*, pages 38–47. J. Stefan Institute, 1999. http://ftp.cs.bris.ac.uk/cgc/ICML99/lindner.ps.Z.

[5] G. Nakhaeizadeh and A. Schnabl. Development of multi-criteria metrics for evaluation of data mining algorithms. In D. Heckerman, H. Mannila, D. Pregibon, and R. Uthurusamy, editors, *Proceedings of the Fourth International Conference on Knowledge Discovery in Databases & Data Mining*, pages 37–42. AAAI Press, 1998.

[6] C. Soares. Ranking classification algorithms on past performance. Master's thesis, Faculty of Economics, University of Porto, 1999. http://www.ncc.up.pt/~csoares/miac/thesis_revised.zip.

[7] D.H. Wolpert and W.G. Macready. No free lunch theorems for search. Technical Report SFI-TR-95-02-010, The Santa Fe Institute, 1996. http://lucy.ipk.fhg.de:80/~stephan/nfl/nfl.ps.