A COMPREHENSIVE DATASET FOR EVALUATING APPROACHES OF VARIOUS META-LEARNING TASKS

Matthias Reif
German Research Center for Artificial Intelligence, Trippstadter Strasse 122, 67663 Kaiserslautern, Germany

Keywords: Meta-learning, Ranking, Algorithm selection, Dataset, Pattern recognition, Classification.

Abstract: New approaches in pattern recognition are typically evaluated against standard datasets, e.g. from UCI or StatLib. Using the same and publicly available datasets increases the comparability and reproducibility of evaluations. In the field of meta-learning, the actual dataset for evaluation is created based on multiple other datasets. Unfortunately, no comprehensive dataset for meta-learning is currently publicly available. In this paper, we present a novel and publicly available dataset for meta-learning based on 83 datasets, six classification algorithms, and 49 meta-features. Different target variables like accuracy and training time of the classifiers as well as parameter dependent measures are included as ground-truth information. Therefore, the meta-dataset can be used for various meta-learning tasks, e.g. predicting the accuracy and training time of classifiers or predicting the optimal parameter values. Using the presented meta-dataset, a convincing and comparable evaluation of new meta-learning approaches is possible.

1 INTRODUCTION

For a convincing evaluation of new pattern recognition methods, appropriate datasets are essential and a sound and fair comparison of competitive methods requires that each method should be evaluated on exactly the same data. Therefore, many scientific papers use for their evaluations the same datasets from common sources like the UCI machine learning repository (Asuncion and Newman, 2007) or StatLib (Vlachos, 1998).

In meta-learning, a dataset is based on multiple other datasets and contains experience knowledge about how learning algorithms, so called target algorithms, performed on these datasets. Therefore, it is required that multiple target algorithms are applied on multiple datasets. Depending on the number of considered algorithms and datasets, the creation of a meta-dataset can be very computational expensive. For the meta-learning step, datasets are represented by characteristics of them, so called meta-features.

Unfortunately, previous publications in the domain of meta-learning typically use their own data for evaluation that is not publicly available. The reproduction of such a meta-dataset is theoretically possible, but very hard in practice due to missing information about used datasets, parameter values, and implementations. Moreover, meta-learning methods are usually evaluated only on a small number of underlying datasets using a set of unoptimized target classifiers that is not diverse. In this paper, we present a novel dataset that overcomes this limitations. The dataset was created using 83 datasets from different domains and sources, six target classifiers with different theoretical foundations including a parameter optimization, and 49 meta-features, calculated by an R-script that we made publicly available as well. Additionally, the presented dataset includes multiple target measures such as accuracy and run-time that are also available for each parameter combination considered during optimization.

The rest of the paper is structured as follows. First, we give a more detailed introduction to meta-learning in Section 2. In Section 3, we describe the creation of the dataset. The final section comprises the conclusion.

2 META-LEARNING

Meta-learning uses knowledge about algorithms and known datasets in order to make a prediction for a new dataset. Datasets are represented by their properties

Reif M. (2012).
A COMPREHENSIVE DATASET FOR EVALUATING APPROACHES OF VARIOUS META-LEARNING TASKS.
In *Proceedings of the 1st International Conference on Pattern Recognition Applications and Methods*, pages 273-276
DOI: 10.5220/0003736302730276

using different measures, so called meta-features.

The meta-features the desired target variable is computed for all known datasets. These data construct the training data for the meta-learning step. The resulting model is used for predicting the target variable for a new, unknown dataset by applying it on the meta-features of the new dataset. This approach is illustrated in Figure 1.

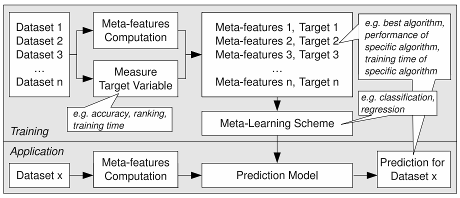


Figure 1: Meta-learning uses meta-features and the desired target value of known datasets for creating a meta-model (top). This model is later used to predict the target value for a new dataset (bottom).

The target variable depends on the goal of the meta-learning approach. In the following, we will present several meta-learning tasks that can be directly applied on the presented meta-dataset.

Best Classifier. In this task, the target variable is the best classifier for each single dataset according to some performance measure, e.g. the classification accuracy. Since this is a classification problem, any classifier can be used. The outcome of the prediction model is the best classifier for the new dataset. This approach was investigated in (Bensusan and Giraud-Carrier, 2000a; Ali and Smith, 2006).

Ranking. The goal is to predict a ranked list of all considered target algorithms, sorted according to some performance measure, e.g. accuracy or time. The target variable consists of the sorted list and a nearest neighbor approach and scores are typically used to predict the ranking. (Brazdil and Soares, 2000; Brazdil et al., 2003; Vilalta et al., 2004).

Quantitative Prediction. This approach directly predicts the performance or run-time of the target algorithm in an appropriate unit. Since the prediction is independently performed for each considered target algorithm, separate regression model has to be trained. The quantitative prediction of error values was evaluated by (Gama and Brazdil, 1995; Sohn, 1999; Köpf et al., 2000; Bensusan and Kalousis, 2001) and the prediction of training-time was evaluated by (Reif et al., 2011).

Predicting Parameters. Besides algorithm selection, meta-learning can also be used for parameter prediction. In this context, the target variable is one parameter value or a set of parameter values. Soares et al. already investigated the parameter selection using meta-learning for the Support Vector Machine classifier (Soares et al., 2004; Soares and Brazdil, 2006).

3 META-DATASET

In this section, the components of the dataset and its creation will be described in more detail.

Meta-features. Meta-features can be grouped according to their underlying analysis concepts. The presented meta-dataset includes 49 meta-features from the following six groups.

Simple Features are directly and easily accessible properties of the dataset which need almost no computations such as number of classes or number of attributes. We included 17 simple meta-features.

Statistical Features use statistical analysis methods and tests (Engels and Theusinger, 1998; Sohn, 1999). Seven measures have been included, e.g. skewness and kurtosis.

Information-theoretic Features typically use entropy measures of the attributes and the class label (Segrera et al., 2008). We used seven features of this group.

Model-based Features create a model of the data, e.g. a decision tree, and use properties of it, e.g. the width and height of the tree, as features. We followed (Peng et al., 2002) and used 17 properties of a decision tree.

Landmarking Features apply fast computable classifiers, e.g. Naive Bayes or 1-Nearest Neighbor, on the dataset (Pfahringer et al., 2000; Bensusan and Giraud-Carrier, 2000b) and use the resulting performance as meta-features. The meta-dataset contains 14 landmarking features.

Time-based Features are specialized for time predictions. They contain time measures of several computations regarding the dataset, e.g. the time for computing the other meta-features. Meta-features of this group have the benefit that they are able to take the performance of the computer into account. Nine different time-measures have been included as presented in (Reif et al., 2011).

The complete list of meta-features can be found on the dataset website[1].

Datasets. We used 83 datasets from the UCI machine learning repository (Asuncion and Newman, 2007), from StatLib (Vlachos, 1998), and from the book "Analyzing Categorical Data" (Simonoff, 2003). All datasets contain 10 to 435 samples with 1 to 69

[1] http://www.dfki.uni-kl.de/ reif/datasets/

Table 1: The classifiers, the number of optimized parameters, and the number of evaluated parameter combinations used for creating the meta-dataset.

Classifier	Parameters	Combinations
Decision Tree	5	161051
k-NN	2	152
MLP	3	242
Naive Bayes	1	2
Ripper	4	2662
SVM	2	225

nominal and numeric attributes and 2 to 24 classes. The complete list can also be found on the website[1].

Classifiers. We selected classifiers that use different learning foundations like tree or rule based learners but also statistical and instance-based learners as well as neural networks. The selected classification algorithms as well as the number of parameters optimized during evaluation are listed in Table 1. Complete details are given on the website[1].

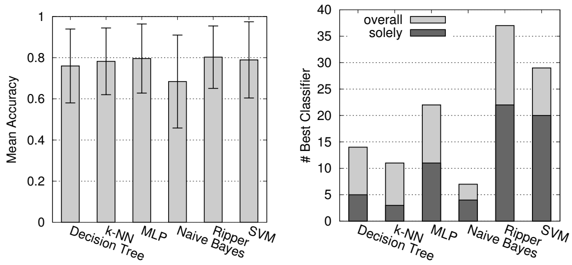
3.1 Generation

After all features were normalized to the range $[0,1]$ and nominal features have been converted to numeric features for the SVM and MLP classifiers, every classifier was evaluated on each dataset using a grid search and 10-fold cross-validation. The accuracy of a classifier is the highest accuracy achieved during the search. The total training time of a classifier is the run-time of the search. Accuracy and training time were also recorded for every considered parameter combination.

The ranking of classifiers for a single dataset was determined by ordering the classifiers according to their accuracy or total training time, respectively. The best classifier for a dataset is the top-ranked classifier. However, several classifiers may achieve the same accuracy for a dataset. In such cases, classifiers with equal accuracy were ordered according to their prior probability of being the best classifier. A different ordering, if necessary, can be easily achieved by using the provided accuracy values.

The ground-truth data was created using RapidMiner (Mierswa et al., 2006). Target times were gathered by measuring the thread CPU time. For the calculation of the meta-features, we wrote an R script that is freely available on the website[1] and can be used to easily extend the meta-dataset by more datasets.

Based on the generated data, we created several variants of the meta-dataset that are directly applicable to one of the tasks described in Section 2. All of these variants share most of the meta-features and principally differ by the target variable. Variants with an accuracy related target value contain all meta-features but the time-based measures whereas the variants for time-based predictions contain all meta-features but the landmarking features. Datasets for parameter prediction contain all parameter combinations. All variants are available as separate plain CSV files and in the XRFF format[2] on our website[1].



(a) The mean accuracy achieved by the classifiers over all 83 datasets including standard deviation.

(b) The number of datasets on which the classifier achieved the highest accuracy overall and solely.

Figure 2: Statistics of the classifiers.

3.2 Statistics

Finally, we present some statistics of the meta-dataset. Figure 2(a) shows the classification accuracy achieved by the target classifiers averaged over all datasets including standard deviation. It is visible that the more sophisticated algorithms achieve almost the same average accuracy, but the simple *k*-Nearest Neighbor algorithm achieved comparable results, as well.

However, if we look at the frequency of a classifier being the best choice for a dataset, the differences are more significant. Figure 2(b) shows how often a classifier achieved the highest accuracy solely (dark gray) and how often it achieved the highest accuracy where another classifier achieved this value as well (light gray). It is visible that SVM and Ripper seem to be superior for many cases, but also the simple approaches of *k*-Nearest Neighbor and Naive Bayes are the best classifiers for several datasets.

4 CONCLUSIONS

In this paper, we presented a novel and publicly available dataset that allows rapid experiments and evaluations of various meta-learning approaches.

[2]http://weka.wikispaces.com/XRFF

The dataset is based on six classifiers with different theoretical foundations, 83 datasets from different domains, and 49 meta-features from six different groups. The R-script for computing the meta-features is also publicly available to make extensions of the meta-dataset easier.

A brief analysis of the gathered data showed that the accuracy of a specific classifier has a large deviation and that also very simple classifiers like Naive Bayes are still the best choice for some datasets. Both aspects make the presented meta-dataset and meta-learning in general a challenging task.

REFERENCES

Ali, S. and Smith, K. A. (2006). On learning algorithm selection for classification. *Applied Soft Computing*, 6:119–138.

Asuncion, A. and Newman, D. (2007). UCI machine learning repository. http://www.ics.uci.edu/∼mlearn/MLRepository.html University of California, Irvine, School of Information and Computer Sciences.

Bensusan, H. and Giraud-Carrier, C. (2000a). Casa batl is in passeig de grcia or how landmark performances can describe tasks. In *Proceedings of the ECML-00 Workshop on Meta-Learning: Building Automatic Advice Strategies for Model Selection and Method Combination*, pages 29–46.

Bensusan, H. and Giraud-Carrier, C. G. (2000b). Discovering task neighbourhoods through landmark learning performances. In *PKDD '00: Proceedings of the 4th European Conference on Principles of Data Mining and Knowledge Discovery*, pages 325–330, London, UK. Springer Berlin / Heidelberg.

Bensusan, H. and Kalousis, A. (2001). Estimating the predictive accuracy of a classifier. In De Raedt, L. and Flach, P., editors, *Machine Learning: ECML 2001*, volume 2167 of *Lecture Notes in Computer Science*, pages 25–36. Springer Berlin / Heidelberg.

Brazdil, P., Soares, C., and da Costa, J. P. (2003). Ranking learning algorithms: Using IBL and meta-learning on accuracy and time results. *Machine Learning*, 50(3):251–277.

Brazdil, P. B. and Soares, C. (2000). Zoomed ranking: Selection of classification algorithms based on relevant performance information. In *In Proceedings of Principles of Data Mining and Knowledge Discovery, 4th European Conference*, pages 126–135.

Engels, R. and Theusinger, C. (1998). Using a data metric for preprocessing advice for data mining applications. In *Proceedings of the European Conference on Artificial Intelligence (ECAI-98*, pages 430–434. John Wiley & Sons.

Gama, J. and Brazdil, P. (1995). Characterization of classification algorithms. In Pinto-Ferreira, C. and Mamede, N., editors, *Progress in Artificial Intelligence*, volume 990 of *Lecture Notes in Computer Science*, pages 189–200. Springer Berlin / Heidelberg.

Köpf, C., Taylor, C., and Keller, J. (2000). Meta-analysis: From data characterisation for meta-learning to meta-regression. In *Proceedings of the PKDD-00 Workshop on Data Mining, Decision Support,Meta-Learning and ILP*.

Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., and Euler, T. (2006). Yale: Rapid prototyping for complex data mining tasks. In Ungar, L., Craven, M., Gunopulos, D., and Eliassi-Rad, T., editors, *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 935–940, New York, NY, USA. ACM.

Peng, Y., Flach, P., Soares, C., and Brazdil, P. (2002). Improved dataset characterisation for meta-learning. In Lange, S., Satoh, K., and Smith, C., editors, *Discovery Science*, volume 2534 of *Lecture Notes in Computer Science*, pages 193–208. Springer Berlin / Heidelberg.

Pfahringer, B., Bensusan, H., and Giraud-Carrier, C. (2000). Meta-learning by landmarking various learning algorithms. In *In Proceedings of the Seventeenth International Conference on Machine Learning*, pages 743–750. Morgan Kaufmann.

Reif, M., Shafait, F., and Dengel, A. (2011). Prediction of classifier training time including parameter optimization. In *34th Annual German Conference on Artificial Intelligence KI11, Berlin, Germany*.

Segrera, S., Pinho, J., and Moreno, M. (2008). Information-theoretic measures for meta-learning. In Corchado, E., Abraham, A., and Pedrycz, W., editors, *Hybrid Artificial Intelligence Systems*, volume 5271 of *Lecture Notes in Computer Science*, pages 458–465. Springer Berlin / Heidelberg.

Simonoff, J. S. (2003). *Analyzing Categorical Data*. Springer Texts in Statistics. Springer Berlin / Heidelberg.

Soares, C. and Brazdil, P. B. (2006). Selecting parameters of SVM using meta-learning and kernel matrix-based meta-features. In *SAC '06: Proceedings of the 2006 ACM symposium on Applied computing*, pages 564–568, New York, NY, USA. ACM.

Soares, C., Brazdil, P. B., and Kuba, P. (2004). A meta-learning method to select the kernel width in support vector regression. *Machine Learning*, 54(3):195–209.

Sohn, S. Y. (1999). Meta analysis of classification algorithms for pattern recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(11):1137 –1144.

Vilalta, R., Giraud-carrier, C., Brazdil, P. B., and Soares, C. (2004). Using meta-learning to support data mining. *International Journal of Computer Science and Applications*, 1(1):31–45.

Vlachos, P. (1998). StatLib datasets archive. http://lib.stat.cmu.edu Department of Statistics, Carnegie Mellon University.