FAKULTÄT FÜR !NFORMATIK

Magisterstudium:
Computational Intelligence

Diplomarbeitspräsentationen der Fakultät für Informatik

Meta-Learning: Machine Learning with No Idea?

Manuel J. A. Eugster

Technische Universität Wien
Institut für Statistik und Wahrscheinlichkeitstheorie
Arbeitsbereich: Computational Intelligence
Betreuer: Univ.-Prof. Dr. Friedrich Leisch

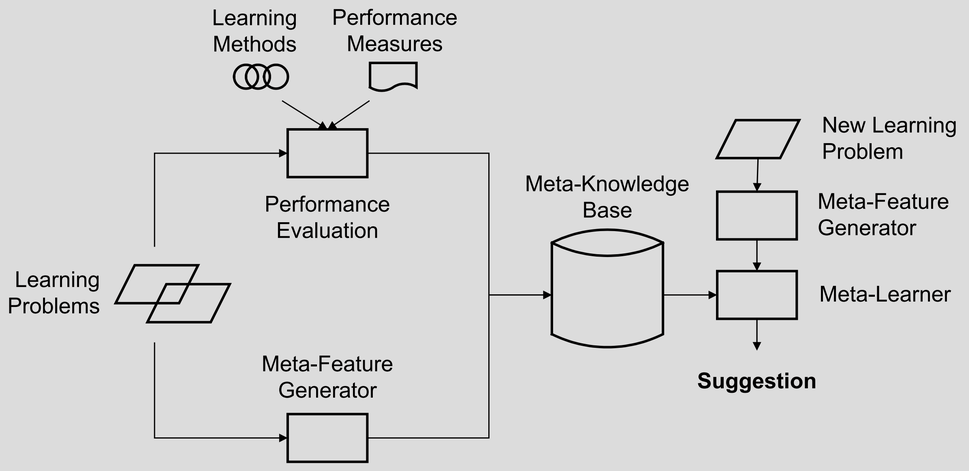
Problem

If an analyst has to solve a problem with statistical and machine learning methods he has the possibility to choose from a lot of methods. Each of these methods has its advantages and disadvantages and yields in a different solution, but in most cases the selection of a method is based on the strength of the analyst‘s preferences, experiences and background knowledge about the data. This thesis tries to give a suggestion concerning the applicability of different methods for a new classification learning problem based on the experience gained on previous learning problems.

Basic Concept

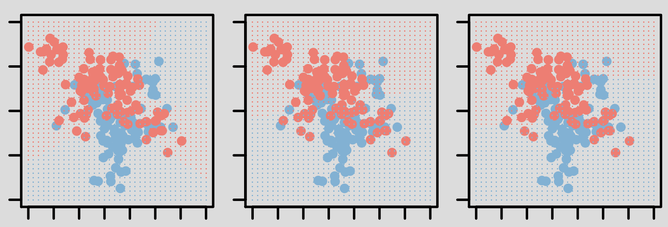
This thesis uses an approach called *Meta-Learning for Machine Learning*. Thereby the idea is to collect the information from the benchmark process and relate them to the learning problems using a characterisation with various statistical and information-theoretical measures. This data is collected in the meta-knowledge base and used for the formulation of a new learning problem to predict a suggestion. Based on the favored kind of suggestion different formulations are possible.

Given a new unknown learning problem, the characterisation of this problem is computed and the meta-learner uses the meta-knowledge base and the characterisation to create a suggestion concerning the best method or a ranking of all available methods.



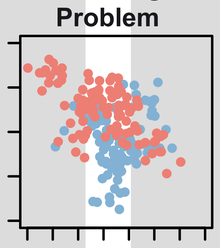
Performance Evaluation

... and other methods, all with different solutions.

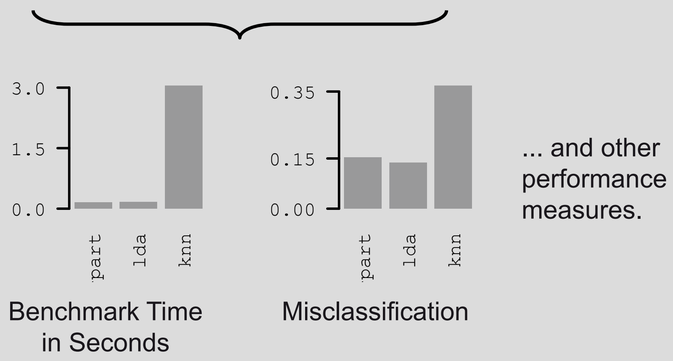


Learning Methods

Learning Problem



Characterisation



Benchmark Time in Seconds

Misclassification

... and other performance measures.

The performance evaluation is done with benchmark experiments. The learning problems are seen as real-world problems where the true data generating process is unknown but a single dataset is available. *B* (e.g. 250) independent training samples are drawn and the corresponding test samples are defined in terms of the out-of-bootstrap observations. This results in a block-design procedure and standard statistical tests can be used to obtain a ranking of the methods on the datasets.

Meta-Feature Generator

n	attr	...	skew	frac1	sd.ratio
690	15	...	*NA*	0.48	1.62

The characterisation is done with 32 statistical and information-theoretical measures. It describes nominal and continuous attributes and their relation to each other. If a measure is not applicable to a dataset, the return value is „not available“.

To put the characterised datasets in relation a similarity measure is defined. Based on the ordered and normalized characterisation of two datasets $\langle m_1, ..., m_c \rangle$ and $\langle m'_1, ..., m'_c \rangle$ the distance measure is:

$$d = \sum_{i=1}^{c} d_{ii}^2 \qquad \text{with} \qquad d_{ii} = \begin{cases} m_i - m'_i, \text{ if } m_i, m'_i \in \Re \\ 1, \text{ if one of } m_i \text{ or } m'_i \text{ is } NA \\ 0, \text{ if both } m_i \text{ and } m'_i \text{ are } NA \end{cases}$$

The idea behind this definition is that if one specific character measure is not applicable on both datasets, the two are equal at this measure and the distance between them should be zero. On the other hand, if a measure is applicable on one dataset but not on the other dataset, the distance should be as large as possible.

Meta-Knowledge Base

	n	attr	...	skew	frac1	sd.ratio	**Response**
Problem 1	690	15	...	*NA*	0.48	1.62	lda / 0.34 / ...
...			...				
Problem *N*	1200	2	...	0.02	*NA*	*NA*	rpart / 0.01 / ...

The meta-knowledge base combines the information from the characterisation and the benchmark experiments, thus the problem of giving a suggestion is reduced to a new learning problem, the so-called meta-learning problem. Based on the kind of information used for the response variable from the benchmark process the meta-learning problem can be formulated differently. If one wants to predict a peformance measure it is formulated with regression problems and if one wants to predict methods it is formulated with classification problems. A third approach predicts a relative ranking of the methods using the performance information of neighbour problems. In this thesis a regression-based approach and the zoomed ranking are implemented and analysed.

Meta-Learner and their Suggestions

Nadaraya-Watson-Epanechnikov Ranking:

Characterisation	Response
Problem 1	0.34
...	
Problem *N*	0.01

Meta-Learning Problem knn

Characterisation	Response
Problem 1	0.40
...	
Problem *N*	0.39

Meta-Learning Problem lda

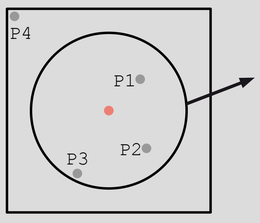
... and for all other methods.

In this regression-based approach the Nadaraya-Watson estimator with the Epanechnikov quadratic kernel is used. A regression problem is established for each learning method, consisting of the dataset characterisation as attributes and the performance measure (e.g. misclassification or benchmark time) as response. The meta-learner estimates the performance measure for each method on the new problem and generates a ranking based on these estimations.

Suggestion for a new problem:

knn	lda	...
0.23	0.53	...

Zoomed Ranking:



	Method (Misclassification / Benchmark Time)
Problem 1	knn (0.34 / 0.2), lda (0.4 / 0.2), ...
Problem 2	knn (0.3 / 2.3), lda (0.5 / 0.4), ...
Problem 3	knn (0.01 / 1.5), lda (0.3 / 0.9), ...

In this approach the *k*-nearest problems to the new problem based on their characterisation and the above defined distance measure are located and the corresponding performance measures are used to establish a relative ranking using different schemata. Basic idea of all schemata is to calculate a pairwise ratio between all algorithms and problems and aggregate these measures to a ranking rate per algorithm.

Suggestion for a new problem: knn < lda < ...

Implementation

The implementation takes place with the R system. I created three packages covering an abstraction of a learning problem, the benchmarking and the meta-learning framework.

Results

The practical usage is shown with a case study consisting of 21 classification learning problems and 6 classification methods. I predicted for each of the problems their corresponding method rankings with the rest of the meta-knowledge base. The Nadaraya-Watson-Epanechnikov meta-learner provided the best suggestions, with its usage the truly best method is benchmarked as the second one on average.

Kontakt: Manuel J. A. Eugster, manuel@mjae.net