mlr: A new Package to conduct Machine Learning Experiments in R

Bischl, Bernd [1]

[1]Fakultät Statistik, TU Dortmund `bischl@statistik.tu-dortmund.de`

The *mlr* package [1] provides a generic, object-oriented interface to many machine learning methods in R for classification and regression and can easily be extended with further ones. It enables the researcher to rapidly conduct complex experiments or implement his own meta-methods using the package's building blocks.

Resampling like cross-validation, bootstrapping and subsampling are used to assess the generalization performance. Hyperparameters of learners can be tuned by grid search or more sophisticated deterministic or stochastic search methods like e.g. Nelder-Mead or CMA-ES [3]. The same holds true for variable selection. Here, mainly the wrapper approach [4] is currently implemented, but the package will be extended with filters and a combination of the two.

Benchmark experiments with two levels of resampling, e.g. nested cross-validation, can be specified with few lines of code to compare different classes of learning algorithms. An interface to the benchmark package by Eugster and Leisch [2] is provided, which enables exploratory and inferential analysis of the results.

Parallel high-performance computing is supported through the *snowfall*, *nws* and *multicore* packages [5]. Experiments can be converted to parallelized versions with a simple configuration command, without touching any further code. The job granularity of scheduled tasks can be changed, so jobs don't complete too early, providing a better scale-up for problems of different sizes.

The talk will include short use cases to explain the package and its optimization algorithms, some general remarks about its programming methodology and a live R demo.

References

[1] B. Bischl. The mlr package: Machine Learning in R. `http://mlr.r-forge.r-project.org`.

[2] M. J. A. Eugster and F. Leisch. Bench plot and mixed effects models: First steps toward a comprehensive benchmark analysis toolbox. In P. Brito, editor, *Compstat 2008— Proceedings in Computational Statistics*, pages 299–306. Physica Verlag, Heidelberg, Germany, 2008.

[3] T. Glasmachers and C. Igel. Uncertainty handling in model selection for support vector machines. In *PPSN*, pages 185–194, 2008.

[4] R. Kohavi and G. H. John. Wrappers for feature subset selection. *Artificial Intelligence*, Vol. 97, pages 273–324, 1997.

[5] M. Schmidtberger, M. Morgan, D. Eddelbüttel, H. L. Yu, L. Tierney and U. Mansmann. State of the Art in Parallel Computing with R. *Journal of Statistical Software*, 31, Issue 1, 1–27, 2009.