

Experiment databases

A new way to share, organize and learn from experiments

Joaquin Vanschoren · Hendrik Blockeel ·
Bernhard Pfahringer · Geoffrey Holmes

Received: 10 March 2009 / Accepted: 16 December 2011 / Published online: 5 January 2012
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract Thousands of machine learning research papers contain extensive experimental comparisons. However, the details of those experiments are often lost after publication, making it impossible to reuse these experiments in further research, or reproduce them to verify the claims made. In this paper, we present a collaboration framework designed to easily share machine learning experiments with the community, and automatically organize them in public databases. This enables immediate reuse of experiments for subsequent, possibly much broader investigation and offers faster and more thorough analysis based on a large set of varied results. We describe how we designed such an experiment database, currently holding over 650,000 classification experiments, and demonstrate its use by answering a wide range of interesting research questions and by verifying a number of recent studies.

Keywords Experimental methodology · Machine learning · Databases · Meta-learning

Editor: Carla Brodley.

J. Vanschoren (✉) · H. Blockeel
LIACS, Universiteit Leiden, Niels Bohrweg 1, 2333CA Leiden, The Netherlands
e-mail: joaquin@liacs.nl

H. Blockeel
e-mail: hendrik.blockeel@cs.kuleuven.be

J. Vanschoren · H. Blockeel
Dept. of Computer Science, Katholieke Universiteit Leuven, Celestijnenlaan 200A, 3001 Leuven, Belgium

B. Pfahringer · G. Holmes
Dept. of Computer Science, The University of Waikato, Private Bag 3105, Hamilton 3240, New Zealand

B. Pfahringer
e-mail: bernhard@cs.waikato.ac.nz

G. Holmes
e-mail: geoff@cs.waikato.ac.nz