



Evaluation Methods for Machine Learning

Papers from the AAAI Workshop

Technical Report WS-06-06



AAAI Press

American Association for Artificial Intelligence

AAAI Press

445 Burgess Drive
Menlo Park, California 94025

ISBN 978-1-57735-288-4 WS-06-06



978157735-2884

Evaluation Methods for Machine Learning

Papers from the AAAI Workshop

*Chris Drummond, William Elazmeh,
and Nathalie Japkowicz, Cochairs*

Technical Report WS-06-06

AAAI Press
Menlo Park, California

Copyright © 2006, AAAI Press

The American Association for Artificial Intelligence
445 Burgess Drive
Menlo Park, California 94025 USA

AAAI maintains compilation copyright for this technical report and retains the right of first refusal to any publication (including electronic distribution) arising from this AAAI event. Please do not make any inquiries or arrangements for hardcopy or electronic publication of all or part of the papers contained in these working notes without first exploring the options available through AAAI Press and *AI Magazine* (concurrent submission to AAAI and another publisher is not acceptable). A signed release of this right by AAAI is required before publication by a third party.

Distribution of this technical report by any means including electronic (including, but not limited to the posting of the papers on any Website) without permission is prohibited.

ISBN 978-1-57735-288-4 WS-06-06

Manufactured in the United States of America

Organizing Committee

Chris Drummond
NRC Institute for Information Technology, Canada
Christopher.Drummond@nrc-cnrc.gc.ca

William Elazmeh
University of Ottawa, Canada
welazmeh@site.uottawa.ca

Nathalie Japkowicz
University of Ottawa, Canada
nat@site.uottawa.ca

Program Committee

Rich Caruana, Cornell University, United States.
Nitesh Chawla, University of Notre Dame, United States.
Ira Cohen, HP Labs, United States.
Charles Elkan, University of California, United States.
Peter Flach, University of Bristol, United Kingdom.
George Forman, Hewlett-Packard Labs, United States.
Robert C. Holte, University of Alberta, Canada.
Aleksander Kolcz, University of Colorado, United States.
Sylvain Letourneau, NRC Institute for Information Technology, Canada.
Dragos Margineantu, The Boeing Company, United States.
Stan Matwin, University of Ottawa, Canada.
Bianca Zadrozny, Universidade Federal Fluminense, Brazil.

Invited Speakers

James D. Malley, National Institutes of Health.
Richard Caruana, Cornell University.
Dragos D. Margineantu, The Boeing Company.

This AAAI-06 Workshop was held July 17, 2006,
in Boston, Massachusetts USA

Contents

Preface / vii

Machine Learning as an Experimental Science (Revisited) / 1
Chris Drummond

Why Question Machine Learning Evaluation Methods?
An Illustrative Review of the Shortcomings of Current Methods / 6
Nathalie Japkowicz

Evaluating Model Selection Abilities of Performance Measures / 12
Jim Huang and Charles X. Ling

Evaluating Probability Estimates from Decision Trees / 18
Nitesh V. Chawla and David A. Cieslak

Beyond Accuracy, F-score, and ROC: A Family of Discriminant
Measures for Performance Evaluation / 24
Marina Sokolova, Nathalie Japkowicz, and Stan Szpakowicz

Evaluation Classifiers: Practical Considerations for Security Applications / 30
Alvaro A. Cárdenas and John S. Baras

Confidence Interval for the Difference in Classification Error / 36
William Elazmeh, Nathalie Japkowicz, and Stan Matwin

Evaluating the Explanatory Value of Bayesian Network
Structure Learning Algorithms / 42
Patrick Shaughnessy and Gary Livingston

