Lecture Notes in Artificial Intelligence 3944

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Joaquin Quiñonero-Candela   Ido Dagan
Bernardo Magnini   Florence d'Alché-Buc (Eds.)

Machine Learning Challenges

Evaluating Predictive Uncertainty
Visual Object Classification
and Recognizing Textual Entailment

First PASCAL Machine Learning Challenges Workshop, MLCW 2005
Southampton, UK, April 11-13, 2005
Revised Selected Papers

Springer

Volume Editors

Joaquin Quiñonero-Candela
Max Planck Institute for Biological Cybernetics
Spemannstr. 38, 72076 Tübingen, Germany
E-mail: joaquin@first.fraunhofer.de

Ido Dagan
Bar Ilan University
Ramat Gan 52900, Israel
E-mail: dagan@macs.biu.ac.il

Bernardo Magnini
ITC-irst, Centro per la Ricerca Scientifica e Tecnologica
Via Sommarive 14, 38050 Povo (Trento), Italy
E-mail: magnini@itc.it

Florence d'Alché-Buc
Université d'Evry-Val d'Essonne
IBISC CNRS FRE 2873 and GENOPOLE
523, Place des terrasses, 91000 Evry, France
E-mail: florence.dalche@ibisc.univ-evry.fr

Library of Congress Control Number: 2006924677

CR Subject Classification (1998): I.2.6-8, I.2.3, I.4-7, F.1, F.2, F.4

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743
ISBN-10 3-540-33427-0 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-33427-9 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springer.com

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11736790 06/3142 5 4 3 2 1 0

Preface

The first **PASCAL** Machine Learning Challenges Workshop (MLCW 2005) (see, `www.pascal-network.org/Workshops/PC04/`) was held in Southampton, UK, during April 11-13, 2005. This conference was organized by the Challenges programme of the European Network of Excellence PASCAL (Pattern Analysis, Statistical modelling and ComputationAl Learning) in the framework of the IST Programme of the European Community. First annually and now quarterly, the PASCAL Challenges Programme plays the role of selecting and sponsoring challenging tasks, either practical or theoretical. The aim is to raise difficult machine learning questions and to motivate innovative research and development of new approaches. Financial support covers all the work concerning the cleaning and labelling of the data as well as the preparation of evaluation tools for ranking the results. For the first round of the programme, four challenges were selected according to their impact in the machine learning community, supported from summer 2004 to early spring 2005 by PASCAL and finally invited to participate in MLCW 2005 :

- The first challenge, called "Evaluating Predictive Uncertainty", dealt with the fundamental question of assigning a degree of confidence to the outputs of a classifier or a regressor.
- The goal of the second challenge, called "Visual Object Classes", was to recognise objects from a number of visual objects classes in realistic scenes.
- The third challenge task, called "Recognizing Textual Entailment", consisted in recognizing, given two texts fragments, whether the meaning of one text can be inferred (entailed) from the other.
- The fourth challenge was concerned with the assessment of "Machine Learning Methodologies to Extract Implicit Relations from Documents".

Each of these challenges raised noticeable attention in the research community, attracting numerous participants. The idea behind having a unique workshop was to make participants in different challenges exchange and benefit from the research experienced in other challenges. For the workshop, the session chairs made a first selection among submissions leading to 34 oral contributions. This book is concerned with selected proceedings of the first three challenges, providing a large panel of machine learning issues and solutions. A second round of selection was made to extract the 25 contributed chapters that make up this book, resulting in a selection rate of one half for the three considered challenges whose description follows.

Evaluating Predictive Uncertainty Challenge
When making decisions based on predictions, it is essential to have a measure of the uncertainty associated to them, or predictive uncertainty. Decisions are of course most often based on a loss function that is to be minimized in expectation. One common approach in machine learning is to assume knowledge of the loss

function, and then train an algorithm that outputs decisions that directly minimize the expected loss. In a realistic setting, however, the loss function might be unknown, or depend on additional factors only determined at a later stage. A system that predicts the presence of calcification from a mammography should also provide information about its uncertainty. Whether to operate or not will depend on the particular patient, as well as on the context in general. If the loss function is unknown, expressing uncertainties becomes crucial. Failing to do so implies throwing information away.

There does not seem to be a universal way of producing good estimates of predictive uncertainty in the machine learning community, nor a consensus on the ways of evaluating them. In part this is caused by deep fundamental differences in methodology (classical statistics, Bayesian inference, statistical learning theory). We decided to organize the Evaluating Predictive Uncertainty Challenge (`http://predict.kyb.tuebingen.mpg.de/`) to allow the different philosophies to compete directly on the empirical battleground. This required us to define losses for probabilistic predictions. Twenty groups of participants competed on two classification and three regression datasets before the submission deadline of December 11, 2004, and a few more after the deadline. We present six contributed chapters to this volume, by all the winners plus authors of other outstanding entries.

Visual Objects Classes
The PASCAL Visual Object Classes Challenge ran from February to March 2005 (`http://www.pascal-network.org/challenges/VOC/`). The goal of the challenge was to recognize objects from a number of visual object classes in realistic scenes (i.e., not pre-segmented objects). Although there already exist benchmarks such as the so-called ‘Caltech 5’ (faces, airplanes, motorbikes, cars rear, spotted cats) and UIUC car side images, largely used by the community of image recognition, it appears now that the developed methods are achieving such good performance that they have effectively saturated on these datasets, and thus the datasets are failing to challenge the next generation of algorithms. Such saturation can arise because the images used do not explore the full range of variability of the imaged visual class. Some dimensions of variability include: clean vs. cluttered background; stereotypical views vs. multiple views (e.g., side views of cars vs. cars from all angles); degree of scale change, amount of occlusion; the presence of multiple objects (of one or multiple classes) in the images.

Given this problem of saturation of performance, the Visual Object Classes Challenge was designed to be more demanding by enhancing some of the dimensions of variability listed above compared to the databases that had been available previously, so as to explore the failure modes of different algorithms. Four object classes were selected: motorbikes, bicycles, cars and people. Twelve teams entered the challenge. This book includes a contributed review chapter about the methods and the results achieved by the participants.

Recognizing Textual Entailment
Semantic analysis of language has been addressed traditionally through interpretation into explicitly stipulated meaning representations. Such semantic interpretation turned out to be a very difficult problem, which led researchers to approximate semantic processing at shallow lexical and lexical-syntactic levels. Usually, such approaches were developed in application-specific settings, without having an encompassing application-independent framework for developing and evaluating generic semantic approaches.

The Recognizing Textual Entailment (RTE) challenge was an attempt to form such a generic framework for applied semantic inference in text understanding. The task takes as input a pair of text snippets, called *text* (T) and *hypothesis* (H), and requires determining whether the meaning of T (most likely) entails that of H or not. The view underlying the RTE task is that different natural language processing applications, including question answering, information extraction, (multi-document) summarization, and machine translation, have to address the language variability problem and recognize that a particular target meaning can be inferred from different text variants. The RTE task abstracts this primary inference need, suggesting that many applications would benefit from generic models for textual entailment.

It is worth emphasizing some relevant features of the task, which contributed to its success:

– RTE is interdisciplinary: the task has been addressed with both machine learning and resource-based NLP techniques. It also succeeded to bridge, as a common benchmark, over different application-oriented communities.
– RTE was a really challenging task: RTE-1, in several respects, was a simplification of the complete task (e.g., we did not consider temporal entailment), but it proved to be at the state of the art of text understanding.
– The challenge attracted 17 participatants and made a strong impact in the research community, followed by a related ACL 2005 workshop and a dozen more conference publications later in 2005, which used the publicly available RTE-1 dataset as a standard benchmark.

February 2006

Joaquin Quiñonero-Candela
Ido Dagan
Bernardo Magnini
Florence d'Alché-Buc
MLCW 2005

Organization

The first (PASCAL) Machine Learning Challenges Workshop (MLCW 2005) was organized by the Challenges programme of the Network of Excellence PASCAL in Southampton, UK, April 11-13, 2005.

PASCAL Challenges Programme Committee

Programme Chairs	Florence d'Alché–Buc (Université d'Evry) Steve Gunn (University of Southampton) Michèle Sebag (Université de Paris XI)
Programme committee	Samy Bengio (IDIAP-Martigny) Alex Clark (Royal Holloway, University of London) Walter Daelemans (University of Antwerp) Cyril Goutte (Xerox Research Centre Europe) Steve Gunn (University of Southampton) Klaus-Robert Mueller (Fraunhofer FIRST) John Shawe-Taylor (University of Southampton) Bill Triggs (INRIA) Chris Watkins (Royal Holloway, University of London)

Programme Committee of (PASCAL) MLCW 2005

Conference Chair	Florence d'Alché–Buc (Université d'Evry) Steve Gunn (University of Southampton)
Local Organization	Eileen Simon (University of Southampton)

Session Chairs

Evaluating Predictive Uncertainty	Joaquin Quiñonero-Candela (Max Planck Institute for Biological Cybernetics, Fraunhofer FIRST and TU Berlin)
Visual Object Classes	Christopher Williams (University of Edinburgh) Andrew Zisserman (University of Oxford)

Recognizing Textual Entailment	Ido Dagan (Bar-Ilan University) Oren Glickman (Bar-Ilan University) Bernardo Magnini (ITC-irst, Istituto per la Ricerca Scientifica e Tecnologica)
Evaluating Machine Learning for Information Extraction	Mary Elaine Califf (Illinois State University) Fabio Ciravegna (University of Sheffield) Dayne Freitag (Fair Isaac Corporation) Nicholas Kushmerick (University College Dublin) Neil Ireson (University of Sheffield) Alberto Lavelli (ITC-irst, Istituto per la Ricerca Scientifica e Tecnologica, Povo)

Sponsoring Institution

PASCAL (Pattern Analysis, Statistical modelling and ComputationAl Learning): European Network of Excellence, IST Programme of the European Community, grant number IST-2002-506778.

Acknowledgements

FAB would like to thank cheerfully Steve Gunn and Michèle Sebag for co-managing the Challenges Programme, Eileen Simon for the nice local organization of the workshop at Southampton and John Shawe-Taylor for his support and his coordination of the European Network of Excellence PASCAL.

Table of Contents