Lecture Notes in Artificial Intelligence 6549

Edited by R. Goebel, J. Siekmann, and W. Wahlster

Subseries of Lecture Notes in Computer Science

Christos Dimitrakakis Aris Gkoulalas-Divanis
Aikaterini Mitrokotsa Vassilios S. Verykios
Yücel Saygin (Eds.)

Privacy and Security Issues in Data Mining and Machine Learning

International ECML/PKDD Workshop, PSDML 2010
Barcelona, Spain, September 24, 2010
Revised Selected Papers

Springer

Volume Editors

Christos Dimitrakakis
Johann Wolfgang Goethe University, Frankfurt, Germany
E-mail: dimitrakakis@fias.uni-frankfurt.de

Aris Gkoulalas-Divanis
IBM Research – Zurich, Rüschlikon, Switzerland
E-mail: AGD@zurich.ibm.com

Aikaterini Mitrokotsa
Ecole Polytechnice Fédérale de Lausanne, Switzerland
E-mail: katerina.mitrokotsa@epfl.ch

Vassilios S. Verykios
University of Thessaly, Volos, Greece
E-mail: verykios@inf.uth.gr

Yücel Saygin
Sabanci University, Tuzla, Istanbul, Turkey
E-mail: ysaygin@sabanciuniv.edu

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-19895-3 e-ISBN 978-3-642-19896-0
DOI 10.1007/978-3-642-19896-0
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011923363

CR Subject Classification (1998): I.2, H.2.8, K.6.5, K.4.1

LNCS Sublibrary: SL 7 – Artificial Intelligence

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This volume contains the papers presented at PSDML 2010: ECML/PKDD Workshop on Privacy and Security issues in Data Mining and Machine Learning held on September 24, 2010 in Barcelona, Spain.

The purpose of the workshop was to bring together researchers from different areas of data mining and machine learning, with an interest in privacy and security, to discuss recent results and open problems and to enable future collaborations. We received 21 submissions, each of which received at least 2, and on average 3.6, reviews. We wish to thank the reviewers for their excellent feedback to the authors, which directly contributed to the workshop's success. The committee decided to accept 11 papers for an engaging full-day program touching upon multiple aspects of the workshop's theme.

One theme was data privacy, i.e., how to perform computations on data without revealing the data itself or any sensitive knowledge that can be mined from the data. This was explored for general computations (such as the eigenvector computation paper by Pathak and Raj and the work by Grosskreutz et al. on group discovery), for anonymous data publication (such as the work by Cano and Torra, who studied the suitability of additive noise to protect sensitive microdata while taking data edits into account), and for supervised learning (such as the work by Pathak and Raj on Gaussian classification and the Gavin and Velcin paper on quadratic error minimization).

Security applications, focusing on detecting malicious behavior in computer systems, formed another major part of the workshop schedule. Kruger et al. contribute a method, employing n-grams and matrix factorization, for automatically mapping network payloads onto a low-dimensional space, enabling visualization and anomaly detection. In a similar vein, Mao et al. use generalized n-grams to represent and detect attacks in network traffic, while the problem of filtering in recommender systems is tackled using a soft two-tier classifier employing bag-of-words and other message statistics as features.

Finally, the open problems and position papers session resulted in interesting and fruitful discussions. Charles Elkan presented the idea of using importance weights to preserve privacy in data mining, and Blaine Nelson gave a detailed overview of an adversarial setting where the opponent actively tries to evade detection by the classifier, which raises many interesting theoretical questions.

We would once more like to thank the workshop participants for their interesting contributions, the reviewers for their diligent work and the ECML/PKDD Workshops Chairs for making this workshop possible.

November 2010

Christos Dimitrakakis
Aris Gkoulalas-Divanis
Aikaterini Mitrokotsa
Yücel Saygin
Vassilios Verykios

Conference Organization

Program Chairs

- Christos Dimitrakakis, Goethe University of Frankfurt, Germany
- Aris Gkoulalas-Divanis, IBM Research – Zurich, Switzerland
- Aikaterini Mitrokotsa, EPFL University, Switzerland
- Yücel Saygin, Sabanci University, Turkey
- Vassilios S. Verykios, University of Thessaly, Greece

Program Committee

- Ulf Brefeld, Yahoo Research, Spain
- Michael Bruckner, University of Postdam, Germany
- Mike Burmester, Florida State University, USA
- Kamalika Chaudhuri, University of California at San Diego, USA
- Peter Christen, Australian National University, Australia
- Chris Clifton, Purdue University, USA
- Maria Luisa Damiani, University of Milan, Italy
- Juan M. Estevez-Tapiador, University of York, UK
- Elena Ferrari, University of Insubria, Italy
- Dimitrios Kalles, Hellenic Open University, Greece
- Murat Kantarcioglu, University of Texas at Dallas, USA
- Kun Liu, Yahoo! Labs, California, USA
- Daniel Lowd, University of Oregon, USA
- Grigorios Loukides, Vanderbilt University, USA
- Emmanuel Magkos, Ionian University, Greece
- Bradley Malin, Vanderbilt University, USA
- Mohamed Mokbel, University of Minnesota, USA
- Blaine Nelson, UC Berkeley, USA
- Ercan Nergiz, Sabanci University, Turkey
- Roberto Perdisci, Georgia Institute of Technology, USA
- Pedro Peris-Lopez, TU Delft, The Netherlands
- Aaron Roth, Carnegie Mellon University, USA
- Benjamin I.P. Rubinstein, University of California, USA
- Jianhua Shao, Cardiff University, UK
- Jessica Staddon, PARC, USA
- Angelos Stavrou, George Mason University, USA
- Grigorios Tsoumakas, Aristotle University of Thessaloniki, Greece
- Shobha Venkataraman, AT&T, USA
- Philip S. Yu, University of Illinois at Chicago, USA

External Reviewers

- Acar Tamersoy, Vanderbilt University, USA
- Kamalika Chaudhuri, University of California, USA

Table of Contents