

Lecture Notes in Artificial Intelligence

6202

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

Subseries of Lecture Notes in Computer Science

Michael May Lorenza Saitta (Eds.)

Ubiquitous Knowledge Discovery

Challenges, Techniques, Applications

Series Editors

Randy Goebel, University of Alberta, Edmonton, Canada

Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Wolfgang Wahlster, DFKI and University of Saarland, Saarbrücken, Germany

Volume Editors

Michael May

Fraunhofer IAIS

Schloss Birlinghoven, 53754 Sankt Augustin, Germany

E-mail: michael.may@iais.fraunhofer.de

Lorenza Saitta

Università del Piemonte Orientale Amedeo Avogadro

Dipartimento di Informatica

Viale Teresa Michel 11, 13100, Alessandria, Italy

E-mail: saitta@di.unipmn.it

Library of Congress Control Number: 2010936640

CR Subject Classification (1998): I.2.4, I.2, H.3-4, C.2, F.1, H.2.8

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743

ISBN-10 3-642-16391-2 Springer Berlin Heidelberg New York

ISBN-13 978-3-642-16391-3 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

springer.com

© Springer-Verlag Berlin Heidelberg 2010

Printed in Germany

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

Printed on acid-free paper 06/3180

Preface

Over the last years, ubiquitous computing has started to create a new world of small, heterogeneous, and distributed devices that have the ability to sense, to communicate and interact in ad hoc or sensor networks and peer-to-peer systems. These large-scale distributed systems, in many cases, have to interact in real-time with their users. Knowledge discovery in ubiquitous environments (KDubiq) is an emerging area of research at the intersection of the two major challenges of highly distributed and mobile systems and advanced knowledge discovery systems. It aims to provide a unifying framework for systematically investigating the mutual dependencies of otherwise quite unrelated technologies employed in building next-generation intelligent systems: machine learning, data mining, sensor networks, grids, peer-to-peer networks, data stream mining, activity recognition, Web 2.0, privacy, user modeling and others.

In a fully ubiquitous setting, the learning typically takes place *in situ*, inside the small devices. Its characteristics are quite different from currently mainstream data mining and machine learning. Instead of offline-learning in a batch setting, sequential learning, anytime learning, real-time learning, online learning, etc.—under real-time constraints from ubiquitous and distributed data—is needed. Instead of learning from stationary distributions, concept drift (the change of a distribution over time) is the rule rather than the exception. Instead of large stand-alone workstations, learning takes place in unreliable, highly resource constrained environments in terms of battery power and bandwidth.

To explore this emerging field of research, a networking project has been funded since 2006 by the European Commission under grant IST-FP6-021321¹: KDubiq (knowledge discovery in ubiquitous environments) is a coordination action at the intersection of the two major challenges of highly distributed and mobile systems and advanced knowledge discovery systems. A basic assumption of the project is that what seems to be a bewildering array of different methodologies and approaches for building “smart,” “adaptive,” “intelligent” ubiquitous knowledge discovery systems can be cast into a coherent, integrated set of key areas centered on the notion of learning from experience. The objective of KDubiq is to provide this common perspective, and to shape a new area of research. For doing so, the KDubiq coordination action has coordinated relevant research done on learning in many subfields, including:

- machine learning and statistics
- knowledge discovery in databases or data mining
- distributed and embedded computing
- mobile computing

¹ See the website www.kdubiq.org for details about the project.

- human computer interaction (HCI)
- cognitive science

A major goal was to create for the first time a forum to bring these individual research lines together, to consolidate the results that have already been achieved, and to pave the way for future research and innovative applications. For doing so, KDubiq has organized a large number of workshops, summer schools, tutorials and dissemination events to bring together this new community.² One important means to focus the activities and discussions was a collaborative effort to provide a *blueprint for the design of ubiquitous knowledge discovery systems*. A number of working groups on relevant topics have been established. Their goal was to create a conceptual framework for this new line of research, to survey the state of the art, and to identify future challenges, both on the theoretical and the applications side.

The result of this collaborative effort is Part I of this book. This blueprint manifests the vision and serves as a practical guide for further, integrated advances in this field, towards, in the long-term, building truly autonomous intelligent systems.

Overview of the Book. Part I of the book aims to provide a conceptual foundation for the new field of ubiquitous knowledge discovery, discussing the state of the art, highlighting challenges and problems, and proposing future directions. Although at some points technical examples are given for illustration, the aim of this chapter is rather on the non-technical, conceptual side.

While Part I is divided into individually authored chapters, it should be seen as a collaborative effort by the working groups of the KDubiq coordination action. Each chapter was read and commented by the other working group members, and influenced by the discussions and findings of the individual working groups.³ Hence, the chapters should be seen as an integrated whole.

Part I of the book is structured as follows. Chapter 1 gives an introduction to the topic and the fundamental issues. Chapter 2 provides an overview on three distributed infrastructures for ubiquitous computing. Chapter 3 discusses how the learning setting itself changes in a ubiquitous environment, when compared to a traditional learning set-up. Chapter 4 defines general characteristics of data in ubiquitous environments. Chapter 5 takes up the issues of privacy and security, arguing that they are critical for the deployment and user acceptance of KDubiq systems. Chapter 6 is devoted to the human-centric view of ubiquitous knowledge discovery systems. Finally, Chapter 7 contains a collection of potential application areas for KDubiq, providing pointers to the state of the art, to existing applications (if available) and to challenges for future research.

² See <http://www.kdubiq.org/kdubiq/images/KDubiq.newsletter.pdf> for details.

³ In some cases, project partners provided input for some sections, but not for the whole chapter. Where this is the case, it is stated in the footnotes of individual sections.

Part II contains selected approaches to ubiquitous knowledge discovery and treats specific aspects in detail. The contributions have been carefully selected to provide illustrations and in-depth discussions for some of the major findings of Part I.

The contribution by Antoine Cornuéjols takes up in greater detail two fundamental challenges for learning in ubiquitous environments: incrementality and non-stationarity. The chapter by Severo and Gama investigates change detection in temporal data, when the assumption of stationarity is not met. Sharfman, Schuster, and Keren have the topic of monitoring changes in highly distributed data stream systems using a geometric approach. The contribution by Inan and Saygin addresses another highly important dimension of ubiquitous knowledge discovery: privacy. It discusses the problem of clustering horizontally partitioned spatio-temporal data (the data sets at the nodes have the same attributes, but different instances) in a privacy-preserving manner. The chapter by Katharina Morik describes a peer-to-peer Web 2.0 application for collaborative structuring of multimedia collections. The chapter by Rasmus Pedersen broadens the discussion and provides an overview on the topic of learning in micro-information systems. The final chapter by Hillol Kargupta and co-workers describes the Mine-Fleet system, one of the few commercially available ubiquitous knowledge discovery systems. Coordinating a network with more than 50 partner institutions and several hundred individual members is a complex, and sometimes daunting, task. We thank the many researchers and practitioners that contributed to the discussions that led to this book in various ways⁴; the invited speakers that helped us to sharpen our understanding of the research issues involved; the numerous workshop and summer school attendees; the project reviewers for constructive criticism; and the EC project officers Fabrizio Sestini and Paul Hearn for their support. We also thank Tino Sanchez for maintaining the project website. By far the greatest thanks, however, go to Ina Lauth. She has coordinated the network activities for three years and did a superb job in making it both a vibrant and pleasant experience for everyone, thereby having a great share in the successful outcome of the project.

The preparation of this book has been supported by the European Commission under grant KDubiq, IST-FP6-021321, which we gratefully acknowledge. We hope that the reader will agree with us that ubiquitous knowledge discovery in many ways holds potential for radically changing the way machine learning and data mining is done today, and share our excitement about this new field of research.

June 2010

Michael May
Lorenza Saitta

⁴ They are too numerous to list individually here, but see the KDubiq newsletter, the members page on KDubiq.org, and the acknowledgments for individual chapters of Part I.

Table of Contents

I A Blueprint for Ubiquitous Knowledge Discovery

Introduction: The Challenge of Ubiquitous Knowledge Discovery	3
<i>Michael May and Lorenza Saitta</i>	
Ubiquitous Technologies	19
<i>Assaf Schuster and Ran Wolff</i>	
Resource Aware Distributed Knowledge Discovery	40
<i>João Gama and Antoine Cornuéjols</i>	
Ubiquitous Data	61
<i>Andreas Hotho, Rasmus Ulslev Pedersen, and Michael Wurst</i>	
Privacy and Security in Ubiquitous Knowledge Discovery	75
<i>Fosca Giannotti and Yücel Saygin</i>	
A Human-Centric Perspective on Ubiquitous Knowledge Discovery	90
<i>Bettina Berendt and Ernestina Menasalvas</i>	
Application Challenges for Ubiquitous Knowledge Discovery	108
<i>Koen Vanhoof and Ina Lauth</i>	

II Case Studies

On-Line Learning: Where Are We So Far?	129
<i>Antoine Cornuéjols</i>	
Change Detection with Kalman Filter and CUSUM	148
<i>Milton Severo and João Gama</i>	
A Geometric Approach to Monitoring Threshold Functions over Distributed Data Streams	163
<i>Izchak Sharfman, Assaf Schuster, and Daniel Keren</i>	
Privacy Preserving Spatio-temporal Clustering on Horizontally Partitioned Data	187
<i>Ali Inan and Yücel Saygin</i>	
Nemoz — A Distributed Framework for Collaborative Media Organization	199
<i>Katharina Morik</i>	

Micro Information Systems and Ubiquitous Knowledge Discovery	216
<i>Rasmus Ulslev Pedersen</i>	
MineFleet®: The Vehicle Data Stream Mining System for Ubiquitous Environments	235
<i>Hillol Kargupta, Michael Gilligan, Vasundhara Puttagunta, Kakali Sarkar, Martin Klein, Nick Lenzi, and Derek Johnson</i>	
Author Index	255