

Introduction

Ubiquitous Knowledge Discovery

João Gama^{a,*} and Michael May^b

^a*LIAAD, University of Porto, Porto, Portugal*

^b*Fraunhofer IAIS, Sankt Augustin, Germany*

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 peer2peer systems. These large scale distributed systems have in many cases 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, P2P, 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 the current 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 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.

The goal of this special issue is to promote an interdisciplinary forum for researchers who deal with sequential learning, anytime learning, real-time learning, online learning, etc. from ubiquitous and distributed data. Distributed Learning from Data Streams is a recent and increasing research area with challenging applications and contributions from fields like Data Bases, Data Mining, Machine Learning, and Statistics. The selected papers cover a large spectrum in the research of Ubiquitous Knowledge Discovery that goes from frequent pattern mining algorithms, distributed clustering, outlier detection to multi-relational learning. The common concept in all the papers is that learning occurs while data continuously flows eventually produced from distributed sources.

The first paper, Gama and Pereira presents a new distributed clustering algorithm which reduces both the dimensionality and the communication burden between sensors. Bifet and Gavaldà propose new algorithms for adaptively mining closed rooted trees, both labeled and unlabeled, from data streams that change over time. Closed patterns are powerful representatives of frequent patterns, since they eliminate

*Corresponding author: João Gama, LIAAD – INESC Porto L.A. Rua de Ceuta, 118-6 andar, 4050-190, Porto, Portugal.
E-mail: jgama@fep.up.pt.

redundant information. Depaire, Falcon, Vanhoof, and Wets introduce a collaborative clustering approach suitable in peer-to-peer networks where different data sites want to cluster their local data as if they consolidated their data sets, but which is prevented by privacy restrictions. Appice, Ceci, Turi, and Malerba propose an extension of a relational algorithm for multi-level frequent pattern discovery, which resorts to data sampling and distributed computation in Grid environments. Marascu and Masegla propose a cluster-based method for outlier detection that automatically adjusts to any clustering result.

In summary, the five selected papers represent some of the latest research in an emerging field with rapid and exciting growth and report the current state of research in ubiquitous knowledge discovery. The reader will notice that Ubiquitous Knowledge Discovery poses many new challenges to the data mining community.

The editors would like to thank Intelligent Data Analysis journal, the Editor-in-Chief, Professor Fazel Famili, and the authors who submitted their work. A special word goes to the anonymous reviewers for their excellent collaboration. This work was developed under the auspices of coordination action KDubiq-WG3 (funded by the European Commission under IST-6FP-021321) and the project Knowledge Discovery from Ubiquitous Data Streams (PTDC/EIA-EIA/098355/2008).