

ECML 2007 PRDD
WARSAW POLAND

THE 18TH EUROPEAN CONFERENCE ON MACHINE LEARNING
AND
THE 11TH EUROPEAN CONFERENCE ON PRINCIPLES AND PRACTICE
OF KNOWLEDGE DISCOVERY IN DATABASES

STATE-OF-THE-ART IN
DATA STREAM MINING

TUTORIAL NOTES

presented by
Mohamed Gaber and Joao Gama

September 17, 2007
Warsaw, Poland

Prepared and presented by:

Mohamed Gaber

Tasmanian ICT Centre, CSIRO ICT Centre, Australia

João Gama

Laboratory of Artificial Intelligence and Decision Support,
INESC-Porto, University of Porto, Portugal

Tutorial Summary

Data streams became ubiquitous as many sources produce data continuously and rapidly. Examples of streaming data include customer click streams, telephone records, web logs, multimedia data, and sets of retail chain transactions. Data streams have brought new challenges to the data mining research community. In consequence, new techniques are needed to process streaming data in reasonable time and space. The goal of this tutorial is to present and discuss the research problems, issues and challenges in learning from data streams. We will present the state-of-the-art techniques in change detection, clustering, classification, frequent patterns, and time series analysis from data streams. Applications of mining data streams in different domains are highlighted. Open issues and future directions will conclude this tutorial. The tutorial also points to data stream mining resources.

Specific goals and objectives

- Introducing the area of data stream mining
- Giving a detailed explanation of the major techniques in the area
- Emphasizing the research issues and challenges

Expected background of the audience

Basic knowledge of data mining concepts and techniques is required.

Outline

1. Introduction
2. Data Streams
3. Change Detection
4. Learning Descriptive Models from Data Streams
5. Learning Predictive Models from Data Streams
6. Frequent pattern mining
7. Time series analysis in data streams
8. Applications of mining data streams
9. Future Directions

Prepared and presented by:

Mohamed Gaber

Tasmanian ICT Centre, CSIRO ICT Centre, Australia

Mohamed Medhat Gaber is a research scientist at Commonwealth Scientific and Industrial Research Organization (CSIRO), Australia. He has published more than 40 articles. Mohamed has served in the program committees of several international and local conferences and workshops in the area of data mining. He has also been serving as a reviewer for the special issues of international journals in the area of data stream mining. He was the co-chair of the International Workshop on Mining Evolving and Streaming Data held in conjunction with ICDM 2006. He is the co-chair of the International Workshop on Knowledge Discovery from Ubiquitous Data Streams to be held in conjunction with ECML/PKDD 2007 and the ACM Workshop on Knowledge Discovery from Sensor Data to be held in conjunction with ACM SIGKDD 2007.

João Gama

Laboratory of Artificial Intelligence and Decision Support, INESC-Porto, University of Porto.

Joao Gama is a researcher at LIACC, the Laboratory of Artificial Intelligence and Computer Science of the University of Porto, working at the Machine Learning group. His main research interest is in Learning from Data Streams. He has published several articles in change detection, learning decision trees from data streams, hierarchical Clustering from streams, etc. Editor of special issues on Data Streams in Intelligent Data Analysis, J. Universal Computer Science, and New Generation Computing Co-chair of a series of Workshops on Knowledge Discovery in Data Streams, ECML 2004, Pisa, Italy, ECML 2005, Porto, Portugal, ICML 2006, Pittsburg, US, ECML 2006 Berlin, Germany, SAC2007, Korea, and the ACM Workshop on Knowledge Discovery from Sensor Data to be held in conjunction with ACM SIGKDD 2007.

Joao and Mohamed are editing a book titled: *Learning from Data Streams-Processing Techniques in Sensor Networks* to be published by Springer.

