3rd Workshop on IoT Large Scale Machine Learning from Data Streams

Preface

Workshop Description

The volume of data is rapidly increasing due to the development of the technology of information and communication. This data comes mostly in the form of streams. Learning from this ever-growing amount of data requires flexible learning models that self-adapt over time. Traditional one shot memory based learning methods trained offline from a static historic data cannot cope with evolving data streams. This is because firstly, it is not feasible to store all incoming data over time and secondly the generated models become quickly obsolete due to data distribution changes, also known as 'concept drift'. The basic assumption of offline learning is that data is generated by a stationary process and the learning models are consistent with future data. However, in multiple applications like web mining, social networks, network monitoring, sensor networks, telecommunications, financial forecasting etc., data samples arrive continuously as unlimited streams often at high-speed. Moreover, the phenomena generating these data streams may evolve over time. In this case, the environment in which the system or the phenomenon generated the data is considered to be dynamic, evolving or non-stationary.

Learning methods used to learn from data generated by dynamically evolving and potentially non-stationary processes must take into account many constraints: (pseudo) real-time processing, high-velocity, and dynamic multi-form change such as concept drift and novelty. In addition in data streams scenarios, the number of classes is often unknown in advance. Therefore, new classes can appear any time and they must be detected and the predictor structure must be updated.

Therefore, data generated by phenomena in dynamical environments are characterized by: (1) potentially unlimited size, (2) sequential access to data samples in the sense that once an observation has been processed, it cannot be retrieved easily unless it is explicitly stored in memory and (3) unpredictable, dependent and not identical distributed observations.

Consequently, learning from streams of evolving and unbounded data requires developing new algorithms and methods able to learn under the following constraints: (1) random access to observations is not feasible or it has high costs, (2) memory is small with respect to the size of data, (3) data distribution or phenomena generating the data may evolve over time, which is known as concept drift and (4) the number of classes may evolve overtime.

It is worthwhile to emphasize that streams are very often generated by distributed sources, especially with the advent of Internet of Things and therefore processing them centrally may not be efficient especially if the infrastructure is large and complex. Scalable and decentralized learning algorithms are potentially more suitable and efficient.

This combined tutorial and workshop aimed at discussing the problem of learning from data streams generated by evolving non-stationary processes. It overviewed the advances of techniques, methods and tools that are dedicated to manage, exploit and interpret data streams in non-stationary environments. In particular, the event examined the problems of modeling, prediction, and classification based on learning from data streams.

The workshop aimed at presenting new research advances related to data streams processing. The complementarity between these methods and tools, together with methods and techniques discussed during the tutorial, was investigated in order to define new suggestions to develop and improve these methods as well as defining new application domains.

December 2018 Joao Gama

Organization

Workshop Chairs

Moamar Sayed-Mouchaweh	Computer Science and Automatic Control Labs, High Engineering School of Mines, Douai, France
Albert Bifet	LTCI, Telecom ParisTech; Paris, France
João Gama	Laboratory of Artificial Intelligence and Decision Support, University of Porto, Porto, Portugal
Rita Ribeiro	Laboratory of Artificial Intelligence and Decision Support, University of Porto, Porto, Portugal

Program Committee Members

Edwin Lughofer	Johannes Kepler University of Linz, Austria
Sylvie Charbonnier	Université Joseph Fourier-Grenoble, France
Bruno Sielly Jales Costa	IFRN, Natal, Brazil
Fernando Gomide	University of Campinas, Brazil
José A. Iglesias	Universidad Carlos III de Madrid, Spain
Anthony Fleury	Mines-Douai, Institut Mines-Télécom, France
Teng Teck Hou	Nanyang Technological University, Singapore
Plamen Angelov	Lancaster University, UK
Igor Skrjanc	University of Ljubljana, Slovenia
Indre Zliobaite	Aalto University, Finland
Elaine Faria	Univ. Uberlandia, Brazil
Mykola Pechenizkiy	TU Eindonvhen, Netherlands
Raquel Sebastião	Univ. Aveiro, Portugal

Keynote Speaker

Bernhard Pfahringer	University of Waikato

Program Committee

Thiago Andrade	INESC TEC
Albert Bifet	LTCI, Telecom ParisTech
Brais Cancela	Universidade da Coruña
Douglas Castilho	Federal University of Minas Gerais
Jonathan De Andrade Silva	University of Sao Paulo at Sao Carlos

Andre de Carvalho	University of São Paulo
Elaine Faria	Federal University of Uberlandia
Carlos Ferreira	LIAAD INESC Porto LA
Joao Gama	University of Porto
Ana Nogueira	INESC TEC
Rita P. Ribeiro	University of Porto
Maria Pedroto	University of Porto
Fabiola Pereira	Federal University of Uberlandia
Moamar Sayed-Mouchaweh	Ecole des Mines de Douai
Ricardo Sousa	LIAAD\INESC
Shazia Tabassum	University of Porto
Bruno Veloso	INESC TEC
Michal Wozniak	Wroclaw University of Technology
Indre Zliobaite	University of Helsinki