Editorial of Special Issue on Software Architectures and Systems for Real Time Data Stream Analytics

Theodora Varvarigou , Dimitrios Zissis , Konstantinos Tserpes

PII: S0164-1212(17)30037-7
DOI: 10.1016/j.jss.2017.02.022
Reference: JSS 9925

To appear in: *The Journal of Systems & Software*

Please cite this article as: Theodora Varvarigou , Dimitrios Zissis , Konstantinos Tserpes , Editorial of Special Issue on Software Architectures and Systems for Real Time Data Stream Analytics, *The Journal of Systems & Software* (2017), doi: 10.1016/j.jss.2017.02.022

Editorial of Special Issue on Software Architectures and Systems for Real Time Data Stream Analytics

Introduction

Our surroundings are generating massive and potentially infinite sequences of data mainly sourced from non-stationary distributions in highly dynamic environments. Such streams of big data are produced by an increasing number of widely adopted systems including social media platforms, surveillance systems, sensor networks, telecommunication records, web logs, etc. Nowadays, real-time data processing at massive scale is becoming more and more of a requirement for businesses, but information technology is challenged when called upon to decipher the complexity of these environments, discover knowledge in these streams of data and produce actionable intelligence. Stream Big Data has high volume and complex data types, but the true challenge lies in its high velocity characteristic, especially when concerning applications that require real-time data mining and machine learning.

It is our pleasure to open the special issue on Software Architectures and Systems for Real Time Data Stream Analytics of the Journal of Software and Systems that includes 11 selected papers, out of which some are revised versions of those presented at the 1st International workshop on Real-time Data Stream Analytics (RTSrteams2015), which was held in conjunction with the 9th IEEE International Conference on Big Data Science and Engineering (IEEE BigDataSE-15) in Helsinki, Finland on August 20-22, 2015,. In this SI we attempt to provide an overview of the field, analysis of specific open scientific challenges, while present viable solutions to a number of these.

Papers

The first paper deals with the issue of the high memory footprint of large-scale time series representations. It is entitled: “Repeating patterns as symbols for long time series representation” by J. Sevcech and M. Bielikova. The authors propose a novel symbolic time series representation using repeating shapes as symbols. The representation is based on k-means clustering that can work on streaming time series. The proposed representation has proven to provide a lower bound on Euclidean distance and it has been shown to be successful in representing the time series on several examples from the UCR time series database and a long time series from energy domain. Furthermore, the authors suggest a design of a proper measure to estimate the distance between time series which is a topic that has gained friction lately in the research community.

The second article, “Data Stream Classification using Random Feature Functions and Novel Method Combinations” by D. Marrón et al, presents some methods for