

Gradual Pattern Mining Tool on Cloud

Dickson Owuor^{*,**}, Anne Laurent^{*}
Joseph Orero^{**}, Olivier Lobry^{***}

^{*}LIRMM Univ Montpellier, CNRS, Montpellier, France
doowuor@lirimm.fr
anne.laurent@umontpellier.fr
^{**}FIT Strathmore University, Nairobi, Kenya
{dowuor, jorero}@strathmore.edu
^{***}OSU OREME Univ Montpellier, Montpellier, France
olivier.lobry@umontpellier.fr

Résumé. This paper describes an approach that illustrates how gradual pattern mining algorithms are integrated into a containerized *Docker* Cloud platform that implements *OGC* (Open Geospatial Consortium) SensorThings API (application interface). We present a practical application of the SensorThings API as a source of real-time data streams and propose an architecture that allows for extraction of gradual patterns among these data streams.

1 Introduction

Scientific researchers are constantly collecting, crossing and analyzing data in order to help them understand various phenomena. For example environmental data is important for helping to understand phenomena like global warming, rainfall patterns etc. Most of such data is collected using sensors that are enabled to upload data in Cloud frameworks (Liang et al., 2016; Hajj-Hassan et al., 2018; Joshi et Simon, 2018). These researchers obviously wish to spend more time studying phenomena than configuring research tools. Fortunately, integrating data analysis tools into Cloud frameworks alleviates such configuration hustles.

It is important to highlight that understanding certain phenomena may require knowledge to be extracted from numerous unrelated data sets. The technique of combining unrelated data sets in order to identify interesting correlations, also known as *data crossing*, is increasingly gaining more interest in numerous research domains (Hajj-Hassan et al., 2018).

Temporal gradual pattern (GP) mining is an instance of a data analysis technique that allows for extraction of gradual correlations among attributes of a data set. For instance given a set of an ordered time-series data set with attributes {Temperature, Mosquitoes}, a temporal GP may take the form : “*the higher the temperature, the more mosquitoes almost 2 hours later*” (Owuor et al., 2019). Time-series data sets that are collected in real-time are also known as *data streams*. The OGC SensorThings API is an example of a framework that generates and stores data streams from numerous environmental sensors.

The idea of a software implementation that enables the access of GP mining algorithms through a SaaS (Software-as-a-service) model through a Web interface is interesting. In a SaaS