

Mining Big Data Streams with Apache SAMOA

Albert Bifet

Huawei Noah's Ark Lab, Hong Kong

Abstract. In this talk, we present Apache SAMOA, an open-source platform for mining big data streams with Apache Flink, Storm and Samza. Real time analytics is becoming the fastest and most efficient way to obtain useful knowledge from what is happening now, allowing organizations to react quickly when problems appear or to detect new trends helping to improve their performance. Apache SAMOA includes algorithms for the most common machine learning tasks such as classification and clustering. It provides a pluggable architecture that allows it to run on Apache Flink, but also with other several distributed stream processing engines such as Storm and Samza.