BIG DATA LEARNING WITH EVOLUTIONARY ALGORITHMS

Isaac Triguer, Mikel Galar

In the era of big data, the leverage of recent advances achieved in distributed technologies enables data mining techniques to discover unknown patterns or hidden relations from voluminous data in a faster way. Extracting knowledge from big data becomes a very interesting and challenging task where we must consider new paradigms to develop scalable algorithms. However, evolutionary models for machine learning and data mining cannot be straightforwardly adapted to the new space and time requirements. Hence, existing algorithms should be redesigned or new ones developed in order to take advantage of their capabilities in the big data context. Moreover, several issues are posed by real-world complex big data problems besides from computational complexity, and big data mining techniques should be able to deal with challenges such as dimensionality, class-imbalance, and lack of annotated samples among others.

In the first part of this tutorial, we will provide a brief introduction to the big data problem, including MapReduce, as the most representative programing paradigm, as well as an overview of recent technologies (Hadoop ecosystem, Spark). Then, we will dive into the field of big data analytics, explaining the challenges that come to the evolutionary models and introducing machine learning libraries such as Mahout, MLlib and FlinkML.

Afterwards, we will go across the main topic of the CEC 2017: evolutionary models in the big data context. Some cases of study will be presented for evolutionary instance selection/generation, feature selection/weighting and imbalanced data classification. Finally, we will carry out a live demonstration with the MLlib and the evolutionary models we have developed for imbalanced big data classification.

Tutorial Chair
Qingfu Zhang · qingfu.zhang@cityu.edu.hk