

A tour on big data classification: Selected Computational Intelligence approaches

Francisco Herrera

University of Granada, Spain

Abstract

In this age, big data applications are increasingly becoming the main focus of attention because of the enormous increment of data generation and storage that has taken place in the last years, in science, business, ... This situation becomes a challenge when huge amounts of data are processed to extract knowledge because the data mining techniques are not adapted to the new space and time requirements. We must consider the new paradigms to develop scalable algorithms.

The paradigm MapReduce was introduced by Google for processing large amounts of information allowing to deal with large datasets. The open source implementation called Hadoop for this paradigm created by Yahoo led the development of a popular platform, allowing a wide use. Recently other frameworks as Apache Spark are emerging. Different machine learning libraries are developed for these frameworks, such as Mahout (Hadoop) and MLlib (Spark).

In this talk we will pay attention to the big data classification problem, we will analyze the available learning algorithms, and we will focus the attention on the soft computing based approaches for feature selection and fuzzy rule learning. We will discuss the reduce phase for models fusion, showing some cases of study.