

Tutorial:

" Genetic Programming as Advanced Machine Learning Technique for Big Data Mining"

Sana Ben Hamida

*Université Paris Nanterre
LAMSADE, Université Paris Dauphine*

Genetic Programming (GP) algorithms belong to the family of Evolutionary algorithms. They use special encoding and variation operators, but they are subject to the same Darwin evolutionary aspects and use the same parental and environmental selection. Several works using GP have demonstrated how these algorithms are able to breed computer programs for solving, or approximately solving, a wide variety of machine learning problems such as Sequence Induction, Symbolic Regression, Clustering, Discovering Game-Playing strategies, Forecasting, Induction of Decision Trees, etc. In the era of Big Data, the machine learning techniques have a new challenge: dealing with the huge size of the available data. Genetic Programming techniques have shown to be a very promising solution to learn from very large data sets.

The first part of this tutorial reminds the basics of Darwin evolutionary aspects and details how computer programs are encoded and evolved with GP. Some examples of application are presented in Python. The second part of the tutorial presents some solutions de scale GP to Big Data bases. Otherwise, detailed guidelines are given for the use of GP to solve a symbolic regression problem and a data classification problem using large data sets.

Organization:

The tutorial is organized on two parts as follows:

- **Part1:** Genetic Programming for Machine Learning
- **Part2:** Genetic Programming for Big Data Mining (two examples of application, symbolic regression and data classification)

Case-studies:

The practical part of the tutorial includes two case-studies with the expanded version of GPlearn tool in Python: a symbolic regression application and a data classification application.

Sana Ben-Hamida. Sana Ben-Hamida, Ph.D in computer science 2001, is an Associate Professor at Paris Nanterre University and an Associate Researcher at the Computer Science Laboratory (LAMSADE) of Paris Dauphine University. Her main research interests include Evolutionary Computation, Machine Learning and related applications. In recent years her interests have focused on problems related to scaling Evolutionary machine learning techniques to massive datasets.