

“A change is gonna come”

André C. P. L. F. de Carvalho

Machine Learning for Data Analysis Center

ICMC-Universidade de Sao Paulo, Brasil



Conventional data mining tasks assume that the whole dataset is available. However, most practical problems continuously generate new data, in a data stream. Data stream mining is a research area that investigates the extraction of knowledge from large volumes of continuously generated data. In this talk, I will discuss the main approaches and show some applications of data stream mining.

“A Grammatical Evolution based Hyper-Heuristic for the Automatic Design of Split Criteria”

Marcio Basgalupp

Machine Learning for Data Analysis Center

ICT– Universidade Federal de São Paulo, Brasil



Top-down induction of decision trees (TDIDT) is a powerful method for data classification. A major issue in TDIDT is the decision on which attribute should be selected for dividing the nodes in subsets, creating the tree. For performing such a task, decision trees make use of a split criterion, which is usually an information-theory based measure. Apparently, there is no free-lunch regarding decision-tree split criteria, as is the case of most things in machine learning. Each application may benefit from a distinct split criterion, and the problem we pose here is how to identify the suitable split criterion for each possible application that may emerge. In this presentation we expose a grammatical evolution algorithm for automatically generating split criteria through a context-free grammar. We name our new approach ESC-GE (Evolutionary Split Criteria with Grammatical Evolution). It is empirically evaluated on public gene expression datasets, and we compare its performance with state-of-the-art split criteria, namely the information gain and gain ratio. Results show that ESC-GE outperforms the baseline criteria in the domain of gene expression data, indicating its effectiveness for automatically designing tailor-made split criteria.

Date: Wednesday, 10.09.2014, 4 p.m.

Place: H9