
SUELY OLIVEIRA, The University of Iowa
Parallel Computing Large-scale Data Problems

Large data problems appear in different applications. For example, LiDAR measurements for terrain in geoinformatics and data warehouses for weather modeling. Knowledge of diverse areas of computer science plays a significant role in data science. My research has shifted from scientific computing to data sciences, including the creation of a certificate in Large Data Analysis involving Computer Science, Statistics and Mathematics Departments at the University of Iowa. In the capstone course for this new certificate, we work on various projects involving different types of data sets, parallel algorithms with MapReduce and other parallel paradigms such message passing and GPUs. One type of application I will cover in this talk is classification with a large number of data points and a moderate number of features. It is an improvement on standard soft-margin SVM algorithms. The new algorithm is very fast in terms of numbers of iterations, and relatively easy to parallelize in distributed memory clusters.