

Department of Computer Science

St. Francis Xavier University

Presents

Novel Metrics for Evaluation of Unsupervised Learning Consistency with Applications in Ensemble Clustering

by

Yunfeng Shangguan

St. Francis Xavier University

M.Sc. Thesis Proposal Presentation

July 28, 2021 at 9am AST via Zoom

Unsupervised learning (UL) algorithms or clustering algorithms have a strong tendency to be inconsistent with themselves, when repeatedly run on the same problem, due to the large amount of randomness commonly relied upon in the methods available. In this work, we hypothesize that a novel method for evaluating the consistency of disagreement (CoD) between the segmentations of UL algorithms may be helpful in assessing UL algorithm stability and towards guiding the machine learning community towards the development of more consistent UL algorithms. Consistency is a measure of error variability among multiple results of clustering algorithm. We investigate algorithmic properties that influence UL stability and further hypothesize that ensemble approaches to UL algorithms can assist in improving the CoD. Ensemble approaches for clustering algorithms are described, developed and evaluated in the context of the proposed consistency of disagreement UL algorithm evaluation metric that is also developed as part of this thesis.