DATA SCIENCE SEMINAR
JAN. 27TH @ 1:30PM – 5345 HERZBERG BLDG

DATA ANALYTICS FOR HETEROGENEOUS DATA

By Julio J. Valdes, National Research Council

ABSTRACT

Heterogeneous data refers to objects described by features of different nature (e.g. mixtures of numeric, qualitative (nominal), ordinal, interval, images, documents, signals, graphs, etc.). In addition to the complexity introduced by the heterogeneity of the attributes, the information usually is incomplete (missing values) and is obtained with different degrees and types of uncertainty. An example is the case of a patient, described by non-numeric variables (e.g. gender), ordinal variables (pain intensity), numeric variables (e.g. temperature, blood pressure), image variables (e.g. X-ray), document variables (e.g. a medical laboratory report), signal variables (e.g. ECG), etc. All of these variables provide information about an object as a single whole entity. A given dataset may contain hundreds, thousands or even millions of such objects.

Modern developments in sensor, communication and computer technologies have revolutionized data acquisition by increasing the amount of information obtained from a targeted problem (the 'big data' buzzword), which has received a lot of attention. However, another degree of heterogeneity of the information obtained.

Most data analytic procedures in general are oriented to homogenous data (mostly numeric data). Those among them that have capabilities for handling missing information do so usually via imputation and fewer accept plain data absence. When dealing with problems involving heterogeneous data the usual approaches are i) to work with a (homogeneous) subset of the information and/or ii) to redefine the data attributes so that the resulting information is acceptable by the data processing procedure.

This presentation illustrates an approach to processing heterogeneous information sensu-strictu accepting data incompleteness and uncertainties. Real world examples are presented for important operations (overlooked) consequence has been the increasing in data analytics like classification, regression and data visualization.

Hosted by the Carleton University Institute for Data Science
www.carleton.ca/cuids

FREE
TO ATTEND

NO REGISTRATION REQUIRED

COMPLIMENTARY LIGHT REFRESHMENTS

ABOUT THE SPEAKER

Dr. Julio J. Valdes has a PhD in mathematics. He joined the National Research Council (NRC) as a Research Officer and is an Adjunct Professor at the Department of Engineering and Computer Science, Université du Québec en Outaouais (2006-present). He is also an Adjunct Professor at the School of Information Technology and Engineering, University of Ottawa (2008-present).

CARLETON UNIVERSITY
1125 Colonel By Drive
Ottawa, ON
K1S 5B6

QUESTIONS
Contact Kathryn Elliott
kathryn.elliott@carleton.ca
613-520-2600 ext. 3244