

Grouping techniques for facing Volume and Velocity in the Big Data

Antonio Irpino

Università degli studi della Campania "Luigi Vanvitelli"

Facultad de Informática

Sala de Grados - Lunes 4 de Junio de 2018 - 17:30

Entrada libre hasta completar el aforo

Resumen:

The data deluge who we are living today is fostering the development of new techniques for effective and efficient methods for the analysis and the extraction of knowledge and insights from the data. The Big Data paradigm, in particular, the Volume and the Velocity features, is requiring to change our habits for treating data and for extracting information that is useful for discovering patterns and insights. Also, the exploratory data analysis must reformulate its aims. How many groups are in the data? How to deal with data that doesn't fit my PC memory? How to represent aggregated data or repeated measures on individuals? How data correlates? The Symbolic Data Analysis approach tried to reformulate the statistical thinking in this case. In this talk, we present some tools for working with aggregated data described by empirical distributions of values. Using some real-cases from different fields (data from sensors, official statistics or data stream), and the HistDAWass R package, we show some recent solutions for the unsupervised classification, and for feature selection in a subspace clustering context, and how to interpret the results.

Sobre Antonio Irpino:

Antonio Irpino is an Associate Professor of Statistics at the Dept. of Mathematics and Physics of the University of Campania "L.Vanvitelli" since 2015, where his research and teaching focus on Data Analysis. After receiving his Ph.D. in computational statistics at Federico II University of Naples (2001), he joined the University of Campania (formerly, Second University of Naples) in 2007 as a researcher. He has authored and co-authored more than 50 journal, chapters and conference papers on classification, unsupervised, supervised and semi-supervised, crisp and fuzzy techniques, data reduction techniques, for multivalued and aggregate data with applications in datastream and Symbolic data analysis especially when data are described by distributions. He is also the author of an R package for their treatment. He is a member of the SIS (Italian society of Statistics, since 2004) and of IASC-ISI (International Association for Statistical Computing, since 2008) of ISI (International Statistical Institute, since 2012).



UNIVERSIDAD
COMPLUTENSE
MADRID

Facultad de Informática

ANUNCIO DE CONFERENCIA

POSGRADO
