

ABSTRACTS

15th Seminar D² Seminar Series

Florence Center for Data Science 'Double' Seminar Series

Fabio Schoen - Department of Information Engineering, University of Florence

Title: *Clustering for Optimization, Optimization for Clustering*

Abstract: In this talk I will present two fundamental problems in data science: the global optimization problem (i.e., how to find globally optimal solutions to a mathematical programming problem) and the problem of clustering multi-dimensional data (i.e. how to efficiently group data according to similarity). The aim of this talk is to present the connections between these fundamental problems and to show how each of them can be used to improve the performance of the other one. For Global Optimization problems, the idea of clustering dates back to the 80's, when researchers used clustering techniques to recognize the regions of attraction of local optima, in the search for the global one. Due to reasons that I will be explaining during the seminar, those approaches were abandoned, however we have shown that, provided some modifications are introduced, they might prove very interesting for modern global optimization.

On the other side, clustering high dimensional data is clearly an optimization problem, as we would like to group points so that a measure of similarity within groups is maximized. Recent computational approaches have been developed in which classical clustering techniques, like, e.g., K-means, are used as local optimization tools which, when embedded in a higher level global optimization strategy, can produce significantly better clusters.

This talk is partly based on research done in collaboration with dr. Luca Tigli, PhD, and dr. Pierluigi Mansueto

Alessandro Panunzi & Lorenzo Gregori - Department of Humanities, University of Florence

Title: *Towards action concepts identification through unsupervised and semi-supervised clustering on a multimodal cross-linguistic ontology*

Abstract: This work presents the steps performed on IMAGACT ontology of action to identify cognitively consistent action concepts through machine learning methods. IMAGACT contains a set of 1,010 actions, represented by video scenes, and enriched with linguistic data in 14 languages. Each scene is linked to the full set of verbs that can be used to refer the depicted action, in every language. Starting from these data, an automatic clustering of scenes has been performed, using the linked lexical items as a feature set, following the idea that similar actions can be referred by a similar group of verbs. In order to obtain an evaluation of the clusters, a wide set of surveys have been set up, and action similarity judgements from human raters have been collected. These data have been analyzed together with automatic clustering metrics to evaluate the clustering and to tune the algorithm. The presentation will also focus on similarity evaluation issues emerging from a task that involves human perception and cognitive processing.