

(1.2) Every picture tells a story: Visual Cluster Assessment in Relational Data

Abstract. The VAT/iVAT, algorithms are the parents of a large family of visual assessment models. This talk is divided into three pieces. Part 1 is a prerequisite to Parts 2 and 3, which are independent of each other. I can cover only those topics in Parts 2 and 3 that match the interests of the audience.

Part 1. Definitions of the three canonical problems of cluster analysis: tendency assessment, clustering, and cluster validity. History of Visual Clustering. Applications: role-based compliance assessment, eldercare time series data, and anomaly detection in wireless sensor networks.

Part 2. Extension to siVAT, scalable iVAT for big data. This is the basis of clusiVAT and clusiVAT+ for clustering in big data (Topic 4 below). Application: image segmentation. Extension to coiVAT for assessment of co-clustering tendency in the four clustering problems associated with rectangular relational data. Application: response of 18 Fetal Bovine Serum Treatments to the treatment of fibroblasts in gene expression data.

Part 3. Five Easy Pieces:

asiVAT: non-symmetric data. Application: Social Networks (Monastery data)

impVAT: missing data. Application: Social Networks (Karate club data)

clusiVAT: big data. Applications: clustering in big (synthetic) data, MIT video trajectories

inciVAT/inc-siVAT: streaming data. Application: anomaly detection in Heron Island data.

LOFiVAT: immunization of iVAT and Single linkage to inlier contamination: Application: Grand St. Bernard weather station.