

Three Visual Cluster Validity Methods for Object and Relational Data

Jim Bezdek
Computer Science Dept.
University of West Florida
jbezdek@uwf.edu

Abstract: This talk is about three visual cluster validity methods developed by the authors that can be used for both object and relational data sets. The original method VAT (visual assessment of clustering tendency) works nicely for dissimilarity data up to about $n = 5000$ objects, but VAT quickly bumps up against storage and resolution limits, and is of limited utility for large data sets. The second method in the family is sVAT (scalable VAT). This algorithm generates a sample taken from the rows (or columns) of very large, square, relational data, and builds a VAT image on the sample. sVAT is shown to be exact when the data contain compact, separated clusters in the sense of Dunn, and sVAT can be used with arbitrarily large data sets. The third method is called coVAT. This method can be used to estimate the number of clusters in objects represented by the rows and columns of a rectangular dissimilarity matrix, as well as the number of mixed and unmixed clusters in the union of the row and column objects. We show examples of using coVAT to detect the number of clusters in the four canonical clustering problems that are present in rectangular relational data.

About the speaker: Jim Bezdek received the BSCE from the U. of Nevada (Reno) in 1969, and the Ph.D. in Applied Math from Cornell in 1973. He is currently the Nystul Professor of Computer Science at the University of West Florida. His previous experience includes the directorship of Boeing's HTC Inf. Proc. Lab, and a term as head of Computer Science at the University of South Carolina.

Jim's interests include woodworking, optimization, motorcycles, pattern recognition, fishing, vision and image processing, skiing, computational neural networks, blues music, medical applications and cigars. Jim is the founding editor of the Int'l. Jo. Approximate Reasoning and the IEEE Transactions on Fuzzy Systems. He has been a distinguished lecturer for the IEEE and ACM, is an IEEE fellow, and was president of the IEEE Neural Networks Council in 1997-1998.

Current research topics: multiple prototype classifier designs, mixed fuzzy-possibilistic c-means clustering models, rule extraction with clustering, generalized nearest prototype classifier networks, fusion of heterogeneous fuzzy data, target recognition with LADAR data, mammographic image analysis, topics in cluster validity, robotic control models, fuzzy learning vector quantization, clustering with genetic algorithms, and acceleration of image processing algorithms.