

James C. Bezdek



Wahoo, Giant Trevally, Maui, Hawaii

Jim received the PhD in Applied Mathematics from Cornell University in 1973. Jim is past president of NAFIPS (North American Fuzzy Information Processing Society), IFSA (International Fuzzy Systems Association) and the IEEE CIS (Computational Intelligence Society): founding editor the *Int'l. Jo. Approximate Reasoning* and the *IEEE Transactions on Fuzzy Systems*: Life fellow of the IEEE and IFSA; and a recipient of the IEEE 3rd Millennium, IEEE CIS Fuzzy Systems Pioneer, and IEEE technical field award Rosenblatt medals. Jim's interests: woodworking, optimization, motorcycles, pattern recognition, cigars, clustering in very large data, fishing, visual methods for clustering, blues music, wireless sensor networks, poker and co-clustering in rectangular relational data. Jim retired in 2007, and will be coming to a university near you soon (especially if there is fishing nearby).

Talk 1 : Nominal Time is 90 minutes

Introduction to Pattern Recognition, Fuzzy Models and Cluster Analysis, with selected applications to medical image processing

A. Pattern Recognition 101. This first part of Talk A defines and exemplifies the four basic problems of pattern recognition: process description, feature analysis, clustering and classifier design. We discuss data types and measures of similarity for numerical pattern recognition systems.

B. Fuzzy Models 101. This section begins with a discussion of uncertainty in the real world, and how it is modeled in science. I discuss Zadeh's 1965 paper using the set-theoretic description of fuzzy models. The mathematical and philosophical differences between fuzzy and probabilistic models are discussed. Membership functions are on display, and are seen to be the atomic units of fuzzy models. They are easy to get - in pattern recognition.

C. Cluster Analysis 101. This section opens with material on label vectors and how they are used to build matrix representations of partitions. I characterize the three canonical problems of clustering: tendency assessment (does the data have cluster substructure?); clustering (how do we find partitions of the data?); and validation (are the partitions we find accurate and/or useful?). Then I identify the four types of models used in clustering: partition only (single linkage); prototype only (self-organizing maps); partition and prototypes (hard, fuzzy and possibilistic c-means; EM algorithm for Gaussian mixture decomposition).

D. The hard and fuzzy c-Means clustering models and algorithms. I will give a short history and the development of the basic models and algorithms for hard and fuzzy c-means. Then I will discuss the use of these models for elementary segmentation of medical images.

E. Single Linkage Clustering. This last section of Talk A describes the single linkage (SL) clustering algorithm. An understanding of this method will enable attendees to understand and appreciate the results I discuss in Talk 2.

Talk 2 : Nominal Time is 90 minutes

Anomaly Detection in (Wireless) Sensor Networks: Visual Assessment and Clustering in Environmental Monitoring Networks

A. Wireless sensor networks (WSNs). There are four types of network anomalies: isolated and epoch anomalies are aberrant behavior internal to a single node; second order anomalies are atypical behavior of an entire node; and higher order anomalies are one or more subtrees of nodes in the network that exhibit anomalous behavior. We will discuss two types of models to detect anomalies; DCAD models that use data capture by level sets of elliptical summaries; and ESAD models that rely on visual assessment of elliptical summaries, with detection based on single linkage clustering.

B. The DCAD model. This section describes three models that use data capture by level sets of ellipsoids having effective radii chosen with differing assumptions (viz., % of points captured, % of points within k standard deviations from the mean, and % of points captured based on the chi-squared distribution. Examples are given using real WSN data from the Intel Berkeley Research Lab (IBRL).

C. The ESAD model. This method uses visual assessment of elliptical summaries for anomaly detection. These models begin with four measures of similarity on sets of ellipsoids, namely compound normalized, transformation energy, Bhattacharya distance and focal dissimilarity. We define the four measures and compare them with some simple two-dimensional examples that reveal surprising differences between human and mathematical assessment of elliptical similarities.

D. The ESAD model. The ellipsoidal similarities computed above easily become dissimilarities, so we can apply visual assessment techniques, in particular, the recursive iVAT method, to images of the (dis) similarity data. These images enable us to assess cluster tendency amongst the set of ellipsoids, and estimate the number of clusters (of elliptical summaries) in the data.

E. Anomaly detection in real WSNs. We show that these images are capable of detecting anomalous behavior with numerical examples using both real WSN and artificial data. The real data include the IBRL network, the Great Barrier Reef Ocean Observation System, and the Grand St. Bernard network for wind monitoring in a mountain pass on the border between France and Switzerland. Our model reliably detects first and second order anomalies in each of the three real data sets. An aberrant node, Cyclone Hamish and node drift cause the anomalies. These examples illustrate that the ESAD model for detecting unusual events in environmental monitoring networks is quite effective. Although not exemplified here, we believe that the ESAD model will be useful in many other data collection situations in which unusual events (anomalies) are of interest.