

UNIVERSITY OF CENTRAL FLORIDA & THE SCHOOL OF EECS

present the Spring 2009

EECS Seminar Series



Dr. Lawrence O. Hall

*Chair and Professor, Department of Computer Science and Engineering,
University of South Florida*

“Scalable Fuzzy Clustering Algorithms”

Friday, February 20, 2009 • 2:00 p.m. • Harris Center (HEC) 101

Clustering is the most typical way to group unlabeled data. Today, there are very large unlabeled data sets available. Many of these data sets are too large to fit in the memory of a typical computer. Some of these data sets are so large that they can only be treated as data streams because not all of the data can be stored in a cost-effective manner. Fuzzy clustering algorithms are known to be very useful on small to medium-size data sets. This talk focuses on how to make some well understood classic fuzzy clustering algorithms scale to very large data sets and streaming data sets.

The goal is to be able to create a data partition that reflects the whole data set, but requires practical computation times. In particular, we show that the fuzzy *c*-means families of algorithms can be scaled to provide data partitions that are very close and potentially identical to what you would get if you were able to cluster all the data. The general idea is to cluster subsets of the data and create weighted examples from the subsets. The weighted examples from a previous partition(s) are used with new data to create a new partition which reflects the examples currently loaded in memory and those partitioned previously. This process can be repeated until all the data has been clustered. Several variations on the theme of summarizing previous partitions with a set of weighted examples are given. Some history can be ignored, for example, in time changing data streams. One could choose also to cluster summarizations. Experimental data sets include several which contain tens of millions of examples, as well as streaming data sets. Results from real-world data sets show excellent partitions are obtained. For tractable size data sets it is shown they are comparable to those from fuzzy *c*-means when it clusters all the data.

DR. LAWRENCE O. HALL

Lawrence O. Hall is a Professor and the Chair of the Department of Computer Science and Engineering at University of South Florida. He received his Ph.D. in Computer Science from the Florida State University in 1986 and a B.S. in Applied Mathematics from the Florida Institute of Technology in 1980. He is a fellow of the IEEE. His research interests lie in distributed machine learning, extreme data mining, bioinformatics, pattern recognition and integrating AI into image processing. The exploitation of imprecision with the use of fuzzy logic in pattern recognition, AI and learning is a research theme. He has authored or co-authored over 65 publications in journals, as well as many conference papers and book chapters. Some recent publications appear in the IEEE Transactions on Pattern Analysis and Machine Intelligence, Neural Computation, Information Fusion, Journal of Machine Learning research, IEEE Transactions on Systems, Man, and Cybernetics, Pattern Recognition, the International Conference on Pattern Recognition, the Multiple Classifier Systems Workshop, and the FUZZ-IEEE conference (<http://isl.csee.usf.edu/ailab/hall.html>). He received the IEEE SMC Society Outstanding contribution award in 2008. He received an Outstanding Research achievement award from the Univ. of South Florida in 2004. A past president of NAFIPS. The former vice president for membership of the SMC society. He was the President of the IEEE Systems, Man and Cybernetics society for 2006-7. He was the Editor-In-Chief of the IEEE Transactions on Systems, Man and Cybernetics, Part B, 2002-05. He is the Vice President for Publications of the IEEE Biometrics Council. Also, associate editor for IEEE Transactions on Fuzzy Systems, International Journal of Intelligent Data Analysis, the International Journal of Pattern Recognition and Artificial Intelligence and International Journal of Approximate Reasoning.

SCHOOL OF ELECTRICAL ENGINEERING & COMPUTER SCIENCE

www.eecs.ucf.edu • click on “Colloquiums & Seminars”