A Brief Overview of Unsupervised Clustering Methods

Olfa Nasraoui
Department of Computer Engineering & Computer Science
University of Louisville,
olfa.nasraoui_AT_louisville.edu

One major limitation of many classical clustering algorithms is that they assume that the number of clusters is known. However, in practice, the number of clusters may not be known. This problem is sometimes called *unsupervised clustering*. Unsupervised prototype-based clustering aims at determining the correct number of clusters, C, without any prior knowledge about it, using one of four approaches. The first approach is to proceed by repeating the clustering for several C values at a high computational cost, and using a validity measure to choose the best partiton. The second approach is to perform several passes through the data set, seeking one cluster at a time and then removing from the data set of the next pass the points belonging to a found cluster if it passes a validity test as in the GMVE [JMB91]. The problem with these approaches lies in the difficulty in designing validity measures that can truly evaluate the goodness of fit of a given cluster or partition because they usually require setting thresholds that can widely vary in practice. Also, most validity measures either assume a known underlying inlier or noise distribution or are very sensitive to noise, and hence are not appropriate for general robust clustering. The third approach consists of starting the clustering process with an overspecified number of clusters, and then merging similar clusters and eliminating spurious clusters until the correct number of clusters is left as in Compatible Cluster Merging [KF92]. The fourth and most recent approach is based on Competitive Agglomeration [FK97], which starts by partitioning the data set into an overspecified number of clusters. Then, as the clustering progresses, adjacent clusters compete against each other for data points, and clusters that lose in the competition gradually become depleted and vanish.

Acknowledgment

This work is supported by the National Science Foundation (CAREER Award IIS-0133948 to O. Nasraoui). Partial support of earlier stages of this work by the Office of Naval Research grant (N000014-96-1-0439) and by the National Science Foundation Grant IIS 9800899 is also gratefully acknowledged.

Bibliography

FK97

H. Frigui and R. Krishnapuram.
Clustering by competitive agglomeration.
Pattern Recognition, 30(7):1109-1119, 1997.

JMB91

J. M. Jolion, P. Meer, and S. Bataouche.
Robust clustering with applications in computer vision.
IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(8):791-802, Aug. 1991.

KF92
R. Krishnapuram and C.-P. Freg.
Fitting an unknown number of lines and planes to image data through compatible cluster merging.
Pattern Recognition, 25(4), 1992.

About this document ...

A Brief Overview of Unsupervised Clustering Methods

This document was generated using the **LaTeX2HTML** translator Version 2002 (1.62)

The command line arguments were:
latex2html `-split 0 -image_type gif UnsupervisedClustering.tex`