

# **A Comparison of Software Implementations of SOM Clustering Procedures**

by

Haiyan Li, University of Maryland  
Bruce Golden, University of Maryland  
Edward Wasil, American University  
Paul Zantek, University of Maryland

---

Presented at ANNIE 2002, November 2002

## Focus of the Paper

- Self-organizing maps (SOMs) are routinely used in clustering
- New software for SOM-based clustering continues to emerge
- How well do these software packages perform?
- We construct 96 data sets and evaluate the performance of 4 SOM-based clustering procedures as well as the K-means algorithm
- Classification accuracy is measured using the cluster recovery rate and the Rand statistic.

# Introduction

- Clustering is a common activity in data mining
- The goal is to partition the observations of a data set into clusters
- The observations within a cluster should be similar
- Observations in different clusters should be dissimilar
- Numerous applications in biology, business, and engineering

## Self-Organizing Maps (SOMs)

- Developed by Teuvo Kohonen in early 1980s
- Observations are mapped onto a two-dimensional hexagonal grid
- Related to MDS and Sammon maps, but ensures better spacing
- Colors are used to indicate clusters
- Software: SOM\_PAK (Public domain, WWW), Viscovery (Eudaptics, Austria)

