Parallelization of Mixtures of Multivariate t-Factor Analyzers Software

Mathieu Chalifour, Dr. Jeffrey Andrews*

MacEwan University

POSTER

Mixtures of modified t-factor analyzers (MMtFA) are a family of statistical models that are used to find inherent groups in data. The significance of this family of models is that, unlike many other methods currently used, they are robust with respect to outliers and can be applied to high dimensional data. Although the field of clustering has been around for some time, it has only been in the last decade that the methodology has expanded due to modern computer capabilities, however, the MMtFA models take a significant amount of time to fit even on modern computers. This is a problem in the sense that applied scientists may be unwilling to adopt this approach to cluster analysis purely due to the time costs. The main goal of the research was to optimize the software for the MMtFA family, which is being developed for the R computing platform (free, open-source, compatible with nearly all operating systems). This was done in two ways: parallelization and general efficiency fixes. Parallelization allows each model in the (MMtFA) family software to be estimated independently using different processors within a computer (most desktops have at least four processor cores), instead of a single processor. This greatly reduces the time expense for running the algorithm. Since time is not the only factor of great importance we will also provide comparisons with other clustering algorithms across several data sets to measure performance.