Algunos paquetes de Análisis Cluster de R 2.13

1. **amap**: package includes standard hierarchical clustering and k-means. We optimize implementation (with a parallelized hierarchical clustering) and allow the possibility of using diferent distances like Eulidean or Spearman (rank-based metric). We implement a principal component analysis (with robusts methods).
2. **cba**: Implements clustering techniques such as Proximus and Rock, utility functions for efficient computation of cross distances and data manipulation.
3. **cclust**: Convex Clustering methods, including Kmeans algorithm, On-line Update algorithm (Hard Competitive Learning) and Neural Gas algorithm (Soft Competitive Learning) and calculation of several indexes for finding the number of clusters in a data set.
4. **clvalid:** The R package clValid contains functions for validating the results of a clustering analysis. There are three main types of cluster validation measures available, "internal", "stability", and "biological". The user can choose from nine clustering algorithms in existing R packages,including hierarchical, K-means, self-organizing maps (SOM), and model based clustering. In addition, we provide a function to perform the self-organizing tree algorithm (SOTA) method of clustering. Any combination of validation measures and clustering methods can be requested in a single function call. This allows the user to simultaneously evaluate several clustering algorithms while varying the number of clusters, to help determine the most appropriate method and number of clusters for the dataset of interest.
5. **clue:** Cluster ensembles are collections of individual solutions to a given clustering problem which are useful or necessary to consider in a wide range of applications. The R package~clue provides an extensible computational environment for creating and analyzing cluster ensembles, with basic data structures for representing partitions and hierarchies, and facilities for computing on these, including methods for measuring proximity and obtaining consensus and \secondary" clusterings.
6. **clues:** Determining the optimal number of clusters appears to be a persistent and controver-sial issue in cluster analysis. Most existing R packages targeting clustering require the user to specify the number of clusters in advance. However, if this subjectively chosen number is far from optimal, clustering may produce seriously misleading results. In order to address this vexing problem, we develop the R package clues to automate and evaluate the selection of an optimal number of clusters, which is widely applicable in the _eld of clustering analysis. Package clues uses two main procedures, shrinking and partitioning, to estimate an optimal number of clusters by maximizing an index function, either the CH index or the Silhouette index, rather than relying on guessing a pre-speci_ed number. Five agreement indices (Rand index, Hubert and Arabie's adjusted Rand index, Morey and Agresti's adjusted Rand index, Fowlkes and Mallows index and Jaccard index), which measure the degree of agreement between any two partitions, are also provided in clues. In addition to numerical evidence, clues also supplies a deeper insight into the partitioning process with trajectory plots.
7. **clustofVar:** Cluster analysis of a set of variables. Variables can be quantitative, qualitative or a mixture of both.
8. **clusttool:** Cluster results can change dramatically depending on the choice of the clustering method, the distance measure, and the number of clusters. Moreover, depending on the selected validity measure, there may be different results for the optimal number of clusters. Despite of the changing cluster results, each partition could still be informative and valuable. The results can give an interesting insight into the multivariate data structure even if the validity measure does not suggest the optimum for the chosen cluster number. It is thus desirable to perform cluster analysis in an exploratory context, by changing the cluster parameters and inspecting the results visually. For this purpose, this statistical tool has been developed. Data, subsets of the data, coordinates and maps can be selected. Furhtermore, different parameters like the distance measure, the clustering method, the number of clusters and the validity measure can be selected. Depending on the selection the clusters can be presented on maps. Additionally, plots of the cluster centres are provided.
9. **cluster:** Realiza cluster jerárquico.
10. **clustersim:** GDM Distance, Sokal-Michener Distance, Bray-Curtis Distance, Calinski-Harabasz Index, G2 Index, G3 Index, Silhouette Index, Krzanowski-Lai Index, Hartigan Index, Gap Index, DB Index, Data Normalization, HINoV method, Replication analysis for cluster validation, Clustering with several algorithms, distances, normalizations and icq indices, Symbolic interval distances, Plot functions, Random cluster generation.

11. **clusterfly:** Visualise clustering algorithms with GGobi. Contains both general code for visualising clustering results and specific visualisations for model-based, hierarchical and SOM clustering. See http://had.co.nz/clusterfly for more information.
12. **clv:** Package contains most of the popular internal and external cluster validation methods ready to use for the most of the outputs produced by functions coming from package "cluster". Package contains also functions and examples of usage for cluster stability approach that might be applied to algorithms implemented in "cluster" package as well as user defined clustering algorithms.
13. **dynamictreecut:** Contains methods for detection of clusters in hierarchical clustering dendrograms.
14. **flashclust:** See the description of hclust for details on available clustering methods. If members!=NULL, then d is taken to be a dissimilarity matrix between clusters instead of dissimilarities between singletons and members gives the number of observations per cluster. This way the hierarchical cluster algorithm can be 'started in the middle of the dendrogram', e.g., in order to reconstruct the part of the tree above a cut (see examples). Dissimilarities between clusters can be efficiently computed (i.e., without hclust itself) only for a limited number of distance/linkage combinations, the simplest one being squared Euclidean distance and centroid linkage. In this case the dissimilarities between the clusters are the squared Euclidean distances between cluster means. flashClust is a wrapper for compatibility with older code.
15. **flexclust:** The main function kcca implements a general framework for k-centroids cluster analysis supporting arbitrary distance measures and centroid computation. Further cluster methods include hard competitive learning, neural gas, and QT clustering. There are numerous visualization methods for cluster results (neighborhood graphs, convex cluster hulls, barcharts of centroids, ...), and bootstrap methods for the analysis of cluster stability.
16. **fpc:** Various methods for clustering and cluster validation. Fixed point clustering. Linear regression clustering. Clustering by merging Gaussian mixture components. Symmetric and asymmetric discriminant projections for visualisation of the separation of groupings. Cluster validation statistics for distance based clustering including corrected Rand index. Clusterwise cluster stability assessment. Methods for estimation of the number of clusters: Calinski-Harabasz, Tibshirani and Walther's prediction strength. Gaussian/multinomial mixture fitting for mixed continuous/categorical variables. Veriablewise statistics for cluster interpretation. DBSCAN clustering. Interface functions for many clustering methods implemented in R, including estimating the number of clusters with kmeans, pam and clara. Modality diagnosis for Gaussian mixtures. Note that the use of the package mclust (called by function prabclust) is protected by a special license, see
    http://www.stat.washington.edu/mclust/license.txt. For an overview see package? fpc.
17. **gclus:** Orders panels in scatterplot matrices and parallel coordinate displays by some merit index. Package contains various indices of merit, ordering functions, and enhanced versions of pairs and parcoord which color panels according to their merit level.
18. **MOCCA**: This package provides methods to analyze cluster alternatives based on multi-objective optimization of cluster validation indices.
19. **Modalclust**: Perfroms Modal Clustering (MAC) including Hierarchical Modal Clustering (HMAC) along with their parallel implementation (PHMAC) over several processors. These model-based non-parametric clustering techniques can extract clusters in very high dimensions with arbitrary density shapes. By default clustering is performed over several resolutions and the results are summarized as a hierarchical tree. Associated plot functions are also provided. There is a package vignette that provides many examples
20. **SGCS:** Graph based clustering summaries for spatial point patterns. Includes Connectivity function, Cumulative connectivity function and clustering function, plus the triplet intensity function T.
21. **sigclust:** SigClust is a statistical method for testing the significance of clustering results. SigClust can be applied to assess the statistical significance of splitting a data set into two clusters. For more than two clusters, SigClust can be used iteratively.