LKCA

In the real world, the categorical data is widespread. With the deepening of research, clustering analysis of categorical data has attracted wide attention, which has become a hot spot in the research of cluster analysis. The package of k-type clustering algorithm provides a complementary algorithm methodology to enrich clustering algorithm, especially clustering categorical data. And they are designed to facilitate research efforts on this novel direction.

LKCA is a software toolbox for clustering algorithm. It provides the open-source package for use in R that implements the clustering algorithms to clustering categorical data. The package is designed to facilitate the development of new algorithm in this research. The versions of the K-type package is available from

<http://github>

The K-type package comes with detailed documentation. The documentation is available from

<http://github>

This documentation describes the setup and usage of K-type package. All functions and related data structures are explained in detail.

Copyright

License: GUN GENERAL PUBLIC LICENSE Version 3

The architecture of LKCA is based on two modules. These modules in the architecture are related to each other, and all codes follow the R standards. This guarantees that the package is simple, easy to implement, and easily extendable.

In the module 2, the packages provide three operation modes to implement the clustering algorithm, single thread, multi thread, and distribute computation. Meanwhile, multi-thread operation is used to take advantage of multi-core machines to cluster categorical data. In the parallel operation, it is provided with multiple CPU to execute multiple threads at the same time which equivalently create a set of copies of functions running in parallel .Through the multi thread operation, it will improve the overall processing performance. In addition, by using distributed computing technology, the task will be decomposed into a number of small parts, and assigned to multiple computers for processing, which can save the overall computing time,and greatly improve the computational efficiency.

Based on the above operation mode, the module 1 includes k-modes algorithm, fuzzy k-modes algorithm, SV-k-modes algorithm, and fuzzy SV-k-modes algorithm to achieve the clustering for categorical data.

The implementation of the algorithms in module 1 depends on the sub functions, including the distance function, find mode function, and initial class center selection function.