**LKCA: Library k-modes-type Clustering Algorithms**

For the numerical data, methods of cluster analysis have been well-explored. As the categorical data lack of the inherent geometric properties, the distance between objects cannot be defined naturally. Therefore, the corresponding clustering model and its algorithm design differentiate from that of the numerical data. In recent years, the problem of clustering categorical data has attracted more attention.

LKCA is a software toolbox for k-modes-type clustering algorithm. It provides the comprehensive open-source library for use in R that implements the clustering algorithms for categorical data. The library is designed to facilitate the development of the new algorithms in this direction and make comparisons between the new methods and existing ones available. LKCA is available from

[https://github.com/FuyuanCao/LKCA](https://github.com/cfy/LKCA)

The LKCA library comes with detailed documentation. The documentation is available from

<https://github.com/FuyuanCao/LKCA/tree/master/manual>

This documentation describes the setup and usage of LKCA. All functions and related data structures are explained in detail.

The LKCA architecture is composed of four modules, that is, k-modes algorithm, fuzzy k-modes algorithm, SV-k-modes algorithm, fuzzy SV-k-modes algorithm. The four modules in the LKCA architecture are designed independently, and all codes follow the R standards.

In each module, LKCA provides three patterns to cluster categorical data, including single-threaded, multi-threaded and distributed computation. In the multi-threaded operation, it is provided with multiple CPU to execute multiple threads at the same time, which equivalently creates a set 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.

The implementation of the clustering algorithms depends on these sub functions, including the distance function, finding cluster centers function, and initial cluster center selection function.