LKCA User's Manual in R

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

CONTENTS

1 Overview of LKCA

1.1 Introduction

The k-means algorithm is well known for its efficiency in clustering large data set, however, working only on numeric data limits the use of these k-means-type algorithm. To solve this problem, Huang presented a k-modes algorithm, which extends the k-means algorithm by using a simple matching dissimilarity measure for categorical objects, modes instead of means for clusters, and a frequency-base method to update modes in the clustering process to minimize the clustering cost function.

But the limitation of k-modes algorithm is that each attribute of the objects only has a single categorical value. In real application, some objects maybe have more than one value for an attributes. For instance, many people may have several job titles and more than one hobby or email. Such data representation is very widespread in questionnaire, banking, insurance, retails, and medical databases. The capacity to deal with the categorical data with set-valued attributes is undoubtedly important for clustering algorithm. To cluster categorical data with set-values attributes, Cao proposed a SV-k-modes algorith.

As an emerging research direction, k-modes-type clustering algorithm provides a new, complementary algorithmic methodology to enrich clustering algorithm, especially clustering categorical data in big data analytics. But to the best of our knowledge, there is no comprehensive open-source packages existing for this problem. To facilitate research efforts on this research direction, we develop the open-source library called LKCA (Library k-modes-type Clustering Algorithm).

The main contribution of the LKCA library lies on two aspects. (1) It is the first comprehensive open-source library for clustering categorical data. (2) It is written in R and RStudio respectively, easy to use, and completely open source. We hope it will facilitate the development of clustering categorical data and encourage researchers to extend LKCA and share their algorithms through the LKCA framework.

1.2 Architecture of LKCA

The LKCA architecture is based on a separation of four modules, that is, k-modes algorithm, fuzzy k-modes algorithm, SV-k-modes algorithm, fuzzy SV-k-modes algorithm, as shown in Fig.1. The four modules in the LKCA architecture are designed independently, and all codes follow the R standards.

In each modules, LKCA provides three patterns to implement clustering algorithm, 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 create 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, find mode function, and initial-center selection function.

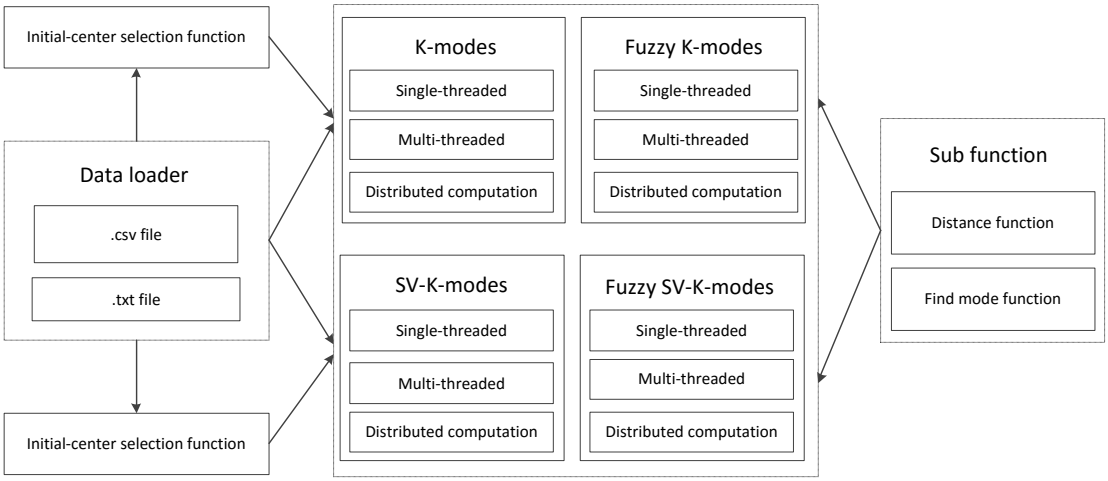

Figure 1 The architecture of LKCA

1.3 Core function overview

The core functions provided in the LKCA library are listed in Table 1.

Table 1 Core functions

R Function	Corresponding Algorithm
K-modes-S	K-modes algorithm(Single-threaded)
K-modes-M	K-modes algorithm(Multi-threaded)
K-modes-D	K-modes algorithm(Distributed Computation)
F-K-modes-S	Fuzzy K-modes algorithm(Single-threaded)
F-K-modes-M	Fuzzy K-modes algorithm(Multi-threaded)
F-K-modes-D	Fuzzy K-modes algorithm(Distributed Computation)
K-dis	Distance function of K-modes algorithm
K-center	Find mode function of K-modes algorithm
SV-K-modes-S	SV-K-modes algorithm(Single-threaded)
SV-K-modes-M	SV-K-modes algorithm(Multi-threaded)
SV-K-modes-D	SV-K-modes algorithm(Distributed Computation)
F-SV-K-modes-S	Fuzzy SV-K-modes algorithm(Single-threaded)
F-SV-K-modes-M	Fuzzy SV-K-modes algorithm(Multi -threaded)
F-SV-K-modes-D	Fuzzy SV-K-modes algorithm(Distributed Computation)
SV-K-dis	Distance function of SV-K-modes algorithm
SV-K-center	Find mode function of SV-K-modes algorithm
K-modes-k-initial-center	Initial cluster centers of k-modes
SV-K-modes-k-initial-center	Initial cluster centers of SV-k-modes

2 Setup in R

2.1 Getting and installing

To run the package, it is required that (1) Windows 7 operating system and above the version;(2) R 2.14.0, or above to be installed; (3) RStudio (4) The package should be loaded to the

folder of current path, as shown in Figure 2.

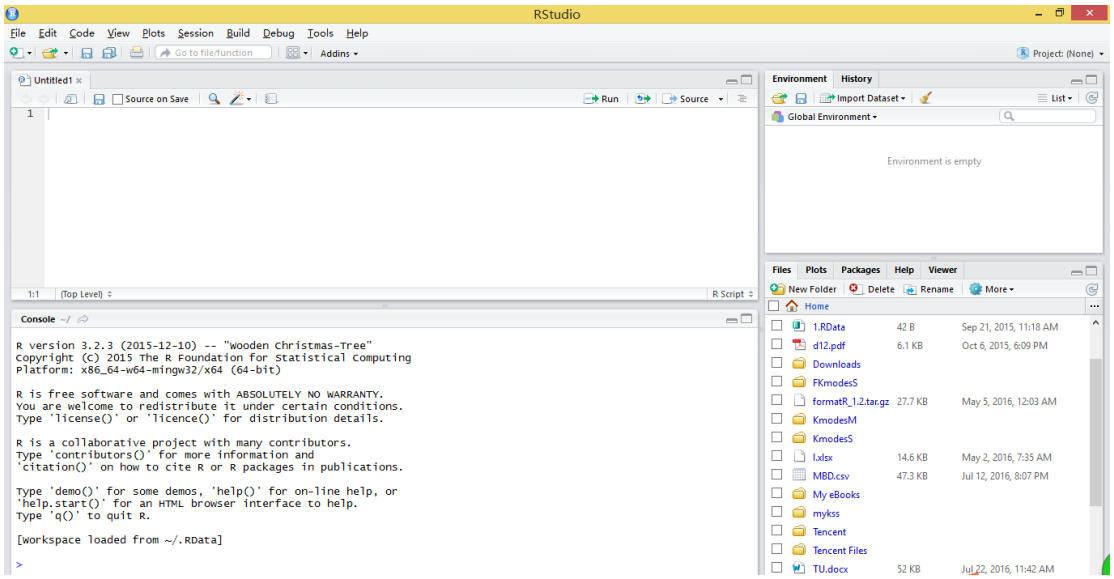

Figure 2 Loading the package

To use RStudio, you must install R in advance. Meanwhile, we can also use R to complete our work, as shown in Figure3. In this paper, we choose to use RStudio, which is a powerful and useful tool for R.

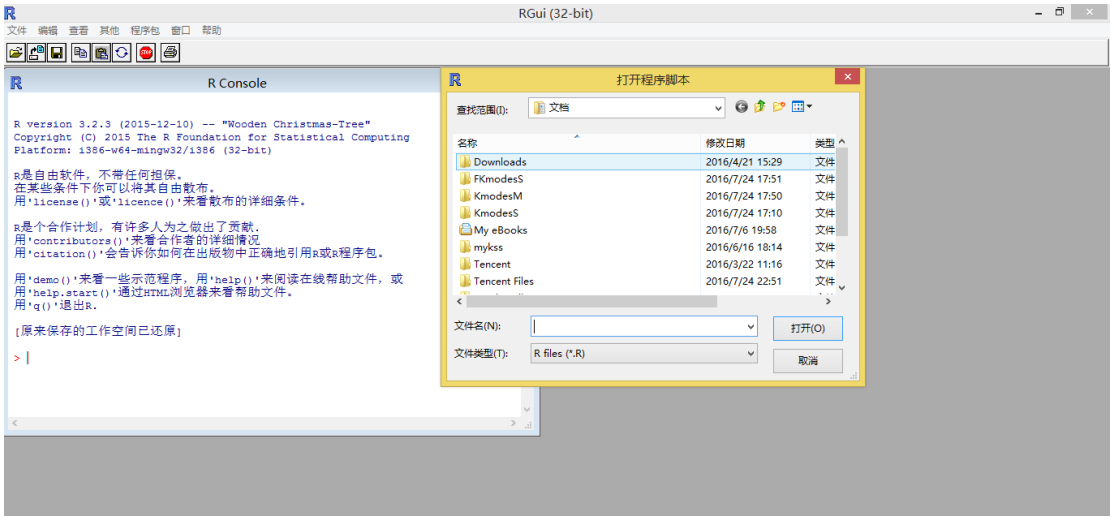

Figure 3 the Interface of R

2.2 Data format

It is required that a data set should be correctly imported to RStudio. Currently, R supports the csv file, txt file.

An example in RStudio is given in Figure 4. The lines represent the number of instances, and the columns represent the number of attributes. “header=FALSE or TRUE” represents that whether the first line is used as the column name. If the first line is used as the column name, we will input “ header=True ”. If not, input “ header=FALSE ”.

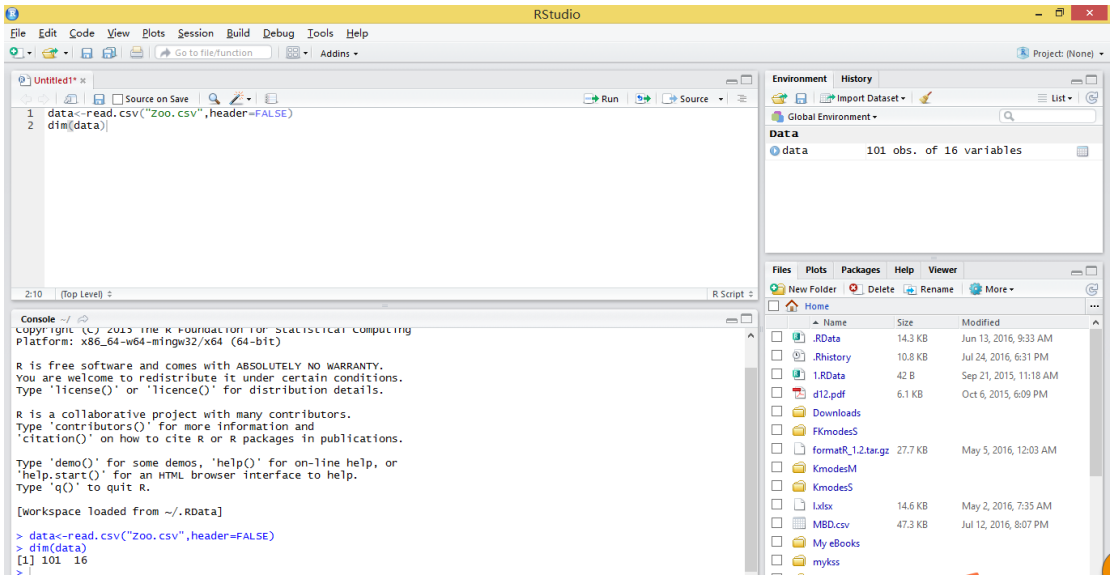

Figure 4 Example of Data format

This denotes that the data set contains 101 data instances, and the 16 columns represent the number of attributes. “header=FALSE ” represents that the first line is not used as the column name.

3 Core functions

3.1 Hard and Fuzzy k-modes clustering algorithm

3.1.1 K-modes-S

Description

Implement the K-Modes algorithm for single-threaded.

Usage

Hard-K-Mode (data, k, InitialCenters)

return (iter, cid, time)

Arguments

Inputs	data	Data set to be clustered
	K	Number of clusters
	InitialCenters	Initial cluster centers
Outputs	iter	Iteration number
	cid	Clustering result
	time	Computational cost

Details

The k-modes algorithm extends the k-means algorithm by using a simple matching dissimilarity measure for categorical objects, modes instead of means for clusters, and a frequency-base method to update modes in the clustering process to minimize the clustering cost function.

Reference

Z.Huang, ” Extensions to the k-means algorithm for clustering large data sets with categorical values,” Data Mining and Knowledge Discovery, vol.2, no. 3. Pp. 283 - 304, 1998

Example

>data=read.csv (“Zoo.csv”, header=FALSE)

>K=7

>InitialCenters=NULL

>library (KmodesS)

>Hard-K-Mode (data, K, InitialCenters)

Results

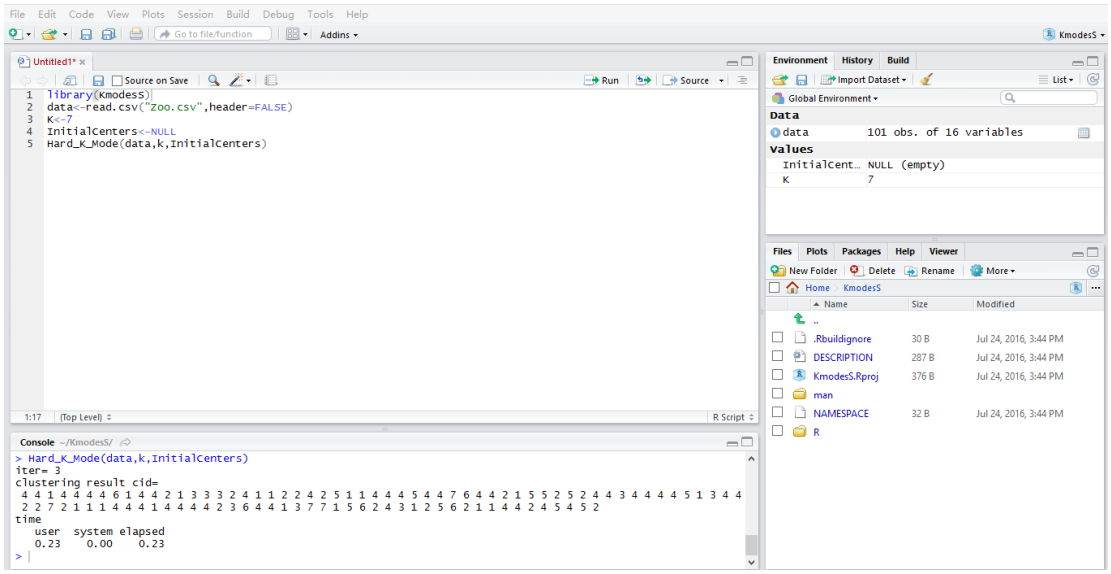

Figure 5 Example of the K-modes -S function

3.1.2 K-modes-M

Description

Implement the K-Modes algorithm for multi-thread operation.

Usage

Hard-K-Mode-M (data, k, InitialCenters)

return (iter, cid, time)

Arguments

Inputs	data	Data set to be clustered
	K	Number of clusters
	InitialCenters	Initial cluster centers
Outputs	iter	Iteration number
	cid	Clustering result
	time	Computational cost

Details

Multi thread operation is used to take advantage of multi-core machines and clusters of categorical data with reduce elapsed times. The operation o is provided with multiple CPU to execute multiple threads at the same time which equivalently create a set of copies of KmodesS package running in parallel .Through the multi thread operation, it will improve the overall processing performance.

Reference

Z.Huang, ” Extensions to the k-means algorithm for clustering large data sets with categorical values,” Data Mining and Knowledge Discovery, vol.2, no. 3. Pp. 283 - 304, 1998

Example

>library (KmodesM)

>data=read.csv (“ Zoo.csv ”, header=FALSE)

>K=7

>InitialCenters=NULL

>Hard-K-Mode-M (data, K, InitialCenters)

Results

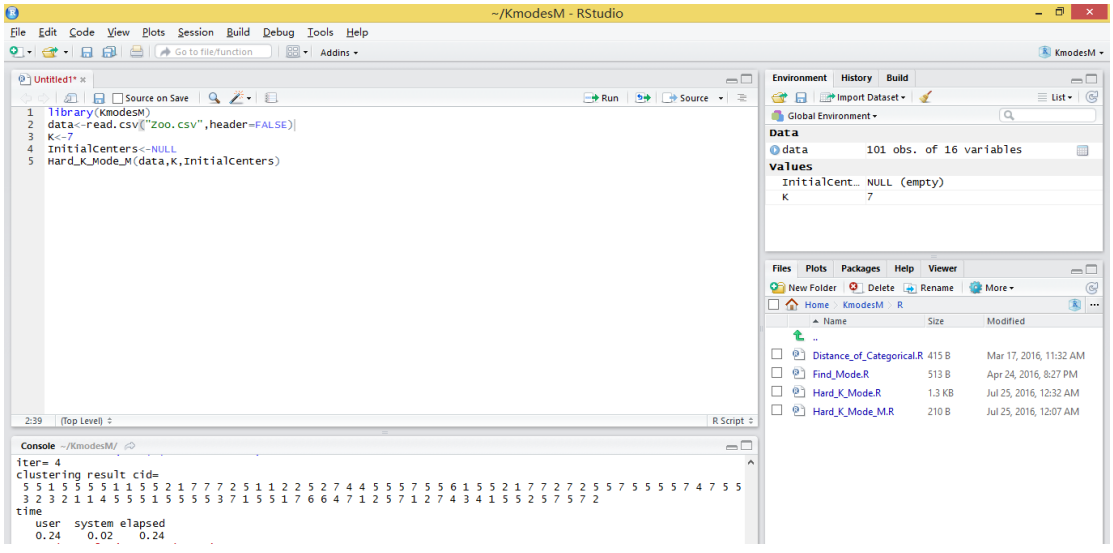

Figure 6 Example of the K-modes -M function

3.1.3 K-modes-D

Description

Implement the K-Modes algorithm for distributed computing.

Usage

Hard-K-Mode-D (data, k, InitialCenters)

return (nj)

Arguments

Inputs	data	Data set to be clustered
	K	Number of clusters
	InitialCenters	Initial cluster centers
Outputs	nj	Clustering result

Details

The distributed computing of the k-modes algorithm needs to rely on the RHadoop, which is a tool that connects R and Hadoop to realize the distributed computing of massive data.

Reference

Z.Huang, ” Extensions to the k-means algorithm for clustering large data sets with categorical values,”,Data Mining Knowledge Discovery, vol.2,no.3,Sept.1998

Example

```
>library ( KmodesD )
>data=read.csv ( “ Zoo.csv ”, header=FALSE )
>K=7
>InitialCenters=NULL
>Hard-K-Mode-D ( data, K, InitialCenters )
```

Results

```
$key
NULL

$val
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14]
[1,]    0    1    1    0    1    0    0    0    1     1     0     0     2     1
[2,]    1    0    0    1    0    0    1    1    1     1     0     0     4     1
[3,]    0    0    1    0    0    1    1    1    1     0     0     1     0     1
[4,]    0    0    0    1    0    1    1    1    1     1     0     1     0     1
[5,]    0    1    1    0    1    1    0    0    1     1     0     0     2     1
[6,]    0    0    1    0    0    1    1    0    1     1     0     0     0     1
[7,]    0    0    1    0    1    0    0    0    1     1     0     0     2     1
     [,15] [,16]
[1,]     0     0
[2,]     0     1
[3,]     0     0
[4,]     0     1
[5,]     0     1
[6,]     0     0
[7,]     0     0
```

Figure 7 Example of the K-modes -D function

3.1.4 F-K-modes-S

Description

Implement the Fuzzy- K-Modes algorithm for single-threaded.

Usage

Fuzzy-K-mode-S (data, K, InitialCenters, a)

return (iter, cid, time)

Arguments

Inputs	data	Data set to be clustered
	K	Number of clusters
	InitialCenters	Initial cluster centers
	a	Index
Outputs	cid	Clustering result
	time	Computational cost
	iter	Iteration number

Details

The fuzzy k-modes algorithm is achieved by the development of a new procedure to generate the fuzzy partition matrix from categorical data within the framework of the fuzzy k-means algorithm. The main result of the algorithm is to find the fuzzy cluster modes when the simple matching dissimilarity measure is used for categorical objects.

Reference

Z.Huang and Michael K.Ng, ”A fuzzy k-modes algorithm for clustering categorical data,” Fuzzy System, IEEE Transaction on, vol. 7. No. 4 pp. 446-452.1999

Example

>data=read.csv (“Zoo.csv”, header=FALSE)

>K=7

>InitialCenters=NULL

>a=1.1

>library (FKmodesS)

>Fuzzy-K-Mode-S (data, K, InitialCenters, a)

Results

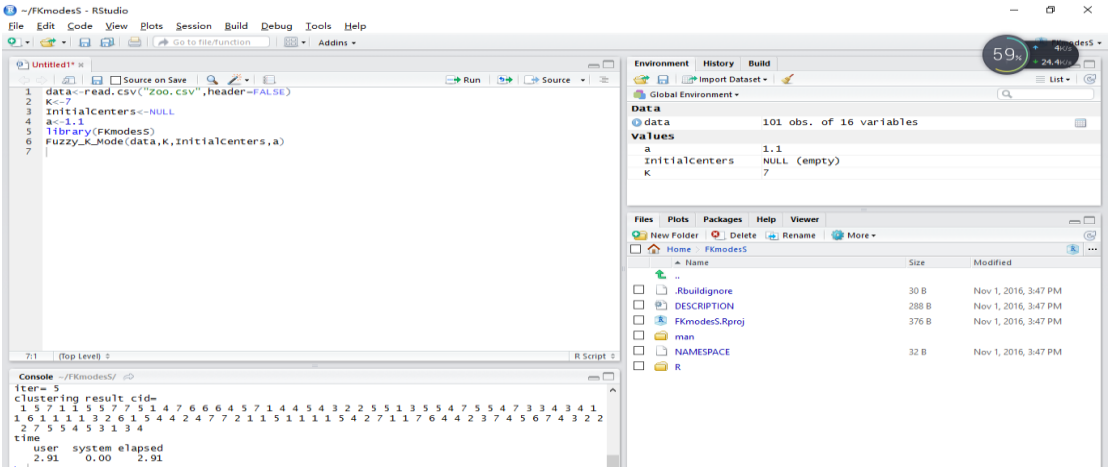

Figure 8 Example of the Fuzzy-K-Modes -S function

3.1.5 F-K-modes-M

Description

Implement the Fuzzy- K-Modes algorithm for multi-thread operation.

Usage

Fuzzy-K-Mode-M (data, K, InitialCenters, a)

return(iter, cid, time)

Arguments

Inputs	data	Data set to be clustered
	K	Number of clusters
	InitialCenters	Initial cluster centers
	a	Index
Outputs	cid	Clustering result
	time	Computational cost
	iter	Iteration number

Details

Please refer to the details of the F-K-modes-S mentioned above.

Reference

Z.Huang and Michael K.Ng,"A fuzzy k-modes algorithm for clustering categorical data," Fuzzy System, IEEE Transaction on, vol. 7. No. 4 pp. 446-452.1999

Example

>data=read.csv ("Zoo.csv", header=FALSE)

>K=7

>InitialCenters=NULL

>a=1.1

>library (FKmodesM)

>Fuzzy-K-Mode-M (data, K, InitialCenters, a)

Results

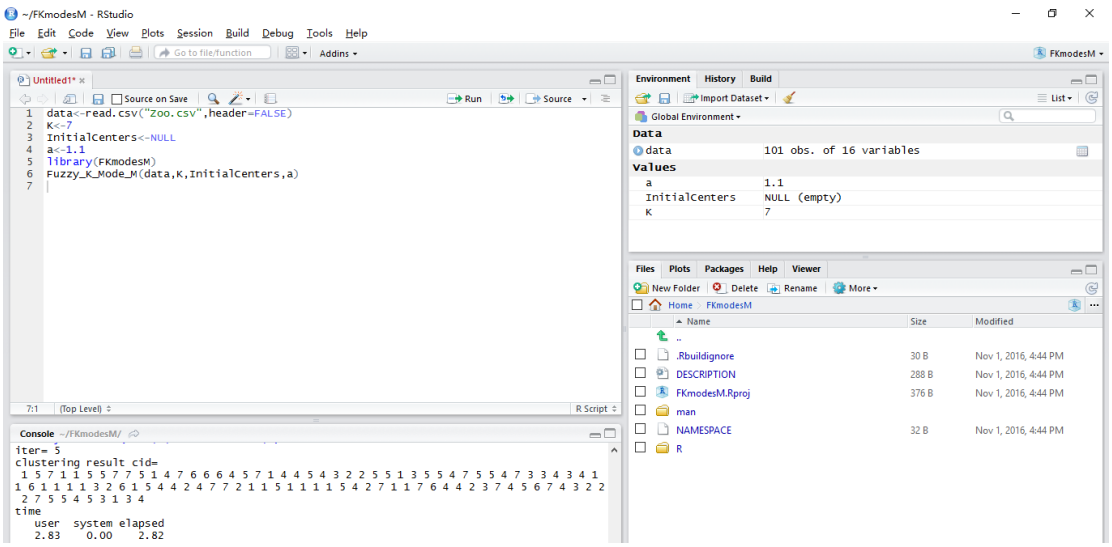

Figure 9 Example of the Fuzzy-K-modes -M function

3.1.6 F-K-modes-D

Description

Implement the Fuzzy- K-Modes algorithm for distributed computing.

Usage

Fuzzy-K-Mode-D (data, K, InitialCenters, a)

return (iter, cid, time)

Arguments

Inputs	data	Data set to be clustered
	K	Number of clusters
	InitialCenters	Initial cluster centers
	a	index
Outputs	nj	Clustering result

Details

Please refer to the details mentioned above.

Reference

Z.Huang and Michael K.Ng,"A fuzzy k-modes algorithm for clustering categorical data," Fuzzy System, IEEE Transaction on, vol. 7. No. 4 pp. 446-452.1999

Example

```
>data=read.csv ( " Zoo.csv ", header=FALSE)
>K=7
>InitialCenters=NULL
>a=1.1
>library ( FKmodesD )
>Fuzzy-K-Mode-D ( data, K, InitialCenters, a)
```

Results

```
$val
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14]
[1,]    0    0    1    0    0    1    1    1    1     1     0     0     4     0
[2,]    1    0    0    1    0    0    0    1    1     1     0     0     2     1
[3,]    0    0    1    0    0    1    1    1    1     0     0     1     0     1
[4,]    1    0    1    0    0    0    1    0    0     1     0     0     6     0
[5,]    1    0    0    1    0    0    1    1    1     1     0     0     4     1
[6,]    1    0    0    1    0    0    0    1    1     1     0     0     4     1
[7,]    1    0    0    1    0    0    0    1    1     1     0     0     4     1
     [,15] [,16]
[1,]     0     0
[2,]     0     0
[3,]     0     0
[4,]     0     0
[5,]     0     1
[6,]     0     1
[7,]     1     0

>
```

Figure 10 Example of the Fuzzy-K-modes -D function

3.2 Hard and Fuzzy SV-k modes clustering algorithm

3.2.1 SV-K-modes -S

Description

Implement the SV-K-modes algorithm for single-threaded.

Usage

New-ratio-k-multi-modes (Data, K, InitialCenters)

return (time)

Arguments

Inputs	Data	Data set to be clustered
	K	Number of clusters
	InitialCenters	Initial cluster centers
Outputs	Time	Computational cost

Details

The SV-k-modes clustering algorithm is designed to cluster the objects with set-valued attributes. The algorithm can partition data with mixed single-valued and set-valued attributes and the k-modes algorithm is its special case.

Reference

Fuyuan Cao, Joshua Zhexue Huang, Jiye Liang, Xingwang Zhao and Yinfeng Meng. “SV-k-modes: an Algorithm for Clustering Categorical Data with Set-valued Attributes,” Neural Networks and Learning System. IEEE Transaction, vol.13.No.9,April 2015

Example

>library (SVKS)

>Data=read.csv (“ MBD.csv ” , header=FALSE)

>K=10

>InitialCenters=NULL

>New-ratio-k-multi-modes (Data, K, InitialCenters)

Results

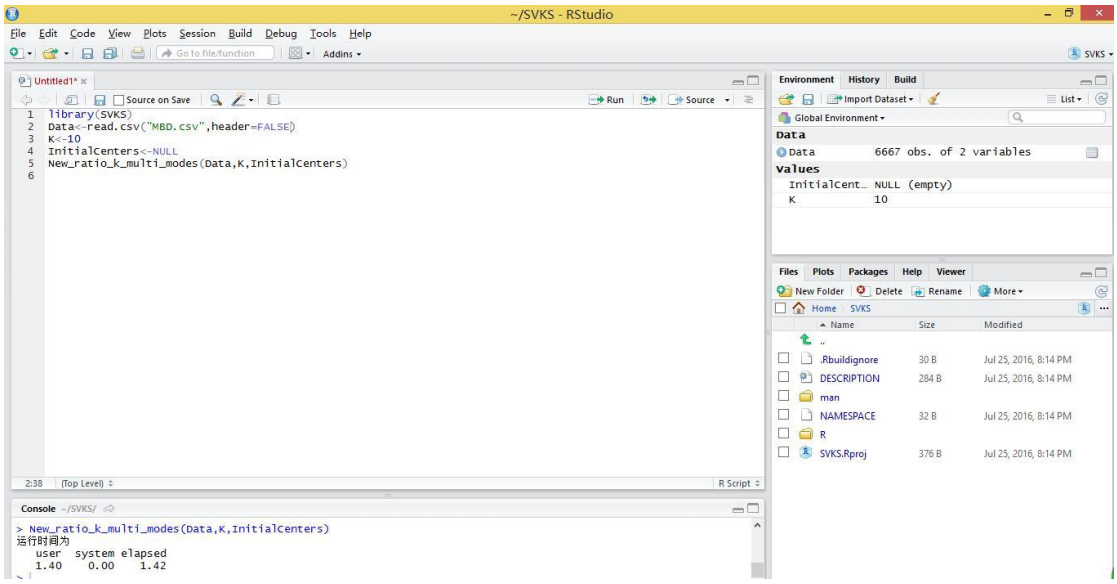

Figure 11 Example of the SV-K-modes -S function

3.2.2 SV-K-modes-M

Description

Implement the SV-K-modes algorithm for multi-threaded.

Usage

New-ratio-k-multi-modes-M (Data, K, InitialCenters)

Return (time)

Arguments

Inputs	Data	Data set to be clustered
	K	Number of clusters
	InitialCenters	Initial cluster centers
Outputs	time	Computational cost

Details

Please refer to the description mentioned above.

Reference

Fuyuan Cao, Joshua Zhexue Huang, Jiye Liang, Xingwang Zhao and Yinfeng Meng. “SV-k-modes: an Algorithm for Clustering Categorical Data with Set-valued Attributes,” Neural Networks and Learning System. IEEE Transaction, vol.13.No.9,April 2015

Example

>library (SVKM)

>Data=read.csv (“ MBD.csv ” , header=FALSE)

>K=10

>InitialCenters=NULL

>New-ratio-k-multi-modes (Data, K, InitialCenters)

Results

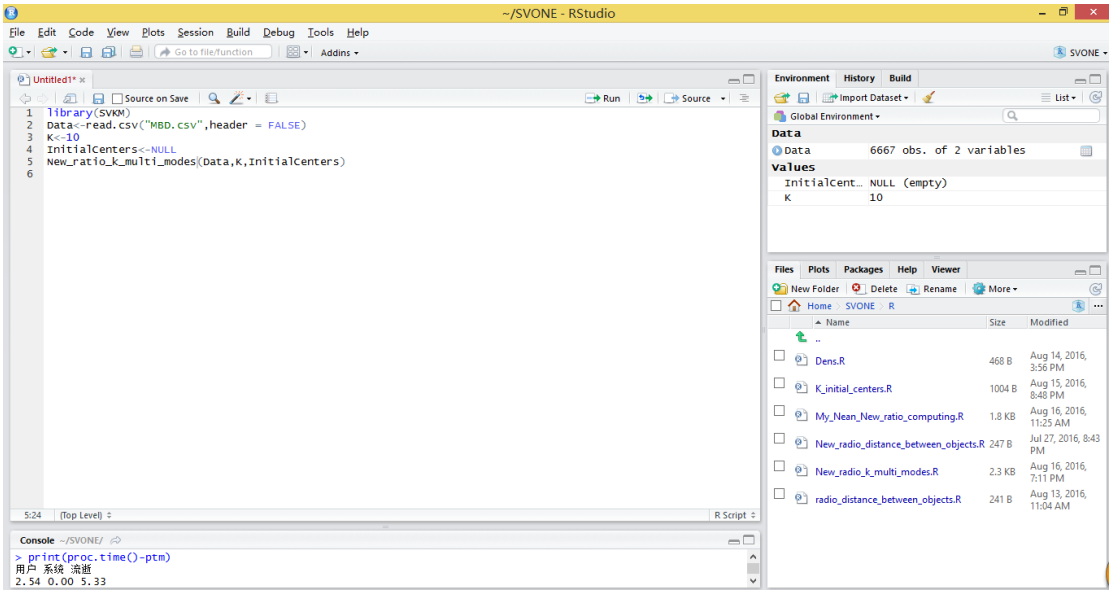

Figure 12 Example of the SV-K-modes -M function

3.2.3 SV-K-modes-D

Description

Implement the SV-K-modes algorithm for distribute computation.

Usage

New-ratio-k-multi-modes-D (Data, K, InitialCenters)

Return (nj)

Arguments

Inputs	Data	Data set to be clustered
	K	Number of clusters
	InitialCenters	Initial cluster centers
Outputs	nj	Clustering result

Details

Please refer to the details mentioned above.

Reference

Fuyuan Cao, Joshua Zhexue Huang, Jiye Liang, Xingwang Zhao and Yinfeng Meng. “SV-k-modes: an Algorithm for Clustering Categorical Data with Set-valued Attributes,” Neural Networks and Learning System. IEEE Transaction, vol.13.No.9,April 2015

Example

```
>library ( SVKD )
>Data=read.csv ( “ MBD.csv ” , header=FALSE )
>K=10
>InitialCenters=null
>New-ratio-k-multi-modes ( Data, K, InitialCenters )
```

Results

```
 [1,] 1 5
 [2,] 1 3
 [3,] 1 8
 [4,] 1 2
 [5,] 1 7
 [6,] 1 1
 [7,] 1 9
 [8,] 2 5
 [9,] 2 3
[10,] 2 8
[11,] 2 2
[12,] 2 7
[13,] 2 1
[14,] 2 9
[15,] 2 9
[16,] 3 5
[17,] 3 3
[18,] 3 8
[19,] 3 2
[20,] 3 7
[21,] 3 1
[22,] 3 9
[23,] 3 9
```

Figure 13 Example of the SV-K-modes -D function

3.2.4 F-SV-K-modes-S

Description

Implement the fuzzy SV-K-modes algorithm for single-threaded.

Usage

Fuzzy-ratio-k-multi-modes (Data, K, InitialCenters, a)

return (iter, cid, time)

Arguments

Inputs	data	Data set to be clustered
	K	Number of clusters
	InitialCenters	Initial cluster centers
	a	Index
Outputs	cid	Clustering result
	time	Computational cost
	iter	Iteration number

Details

Fuzzy SV-k-modes algorithm use Jaccard coefficient to compute the similarty measure between two objects and represent the center of a cluster with set-valued mode. A kind of update ways of cluster prototype is developed for the fuzzy partition matrix. These extensions made the fuzzy SV-k-modes algorithm can cluster categorical data with mixed single-valued and set-valued attributes.

Reference

Fuyuan Cao, Joshua Zhexue Huang, Xingwang Zhao and Jiye Liang, A Fuzzy SV-k-modes Algorithm for Clustering Categorical Data with Set-valued Attributes

Example

>library (FSVKS)

>Data=read.csv (" MBD.csv ", header=FALSE)

>K=10

>InitialCenters=NULL

>a=1.1

>Fuzzy-ratio-k-multi-modes (Data, K, InitialCenters, a)

Results

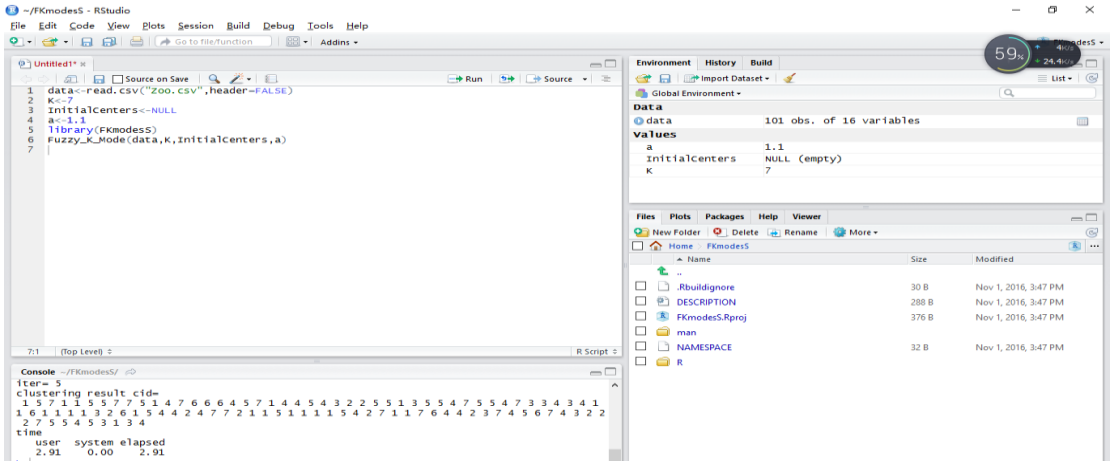

Figure 14 Example of the Fuzzy-SV-K-modes -S function

3.2.5 F-SV-K-modes-M

Description

Implement the fuzzy SV-K-modes algorithm for multi-threaded.

Usage

Fuzzy-ratio-k-multi-modes-M (Data, K, InitialCenters, a)

Return (time)

Arguments

Inputs	data	Data set to be clustered
	K	Number of clusters
	InitialCenters	Initial cluster centers
	a	Index
Outputs	cid	Clustering result
	time	Computational cost
	iter	Iteration number

Details

Please refer to the description mentioned above.

Reference

Fuyuan Cao, Joshua Zhexue Huang, Xingwang Zhao and Jiye Liang, A Fuzzy SV-k-modes Algorithm for Clustering Categorical Data with Set-valued Attributes

Example

>library (FSVKM)

>Data=read.csv (" MBD.csv " ,header=FALSE)

>K=10

>InitialCenters=NULL

>a=1.1

Fuzzy-ratio-k-multi-modes-M (Data, K, InitialCenters,a)

Results

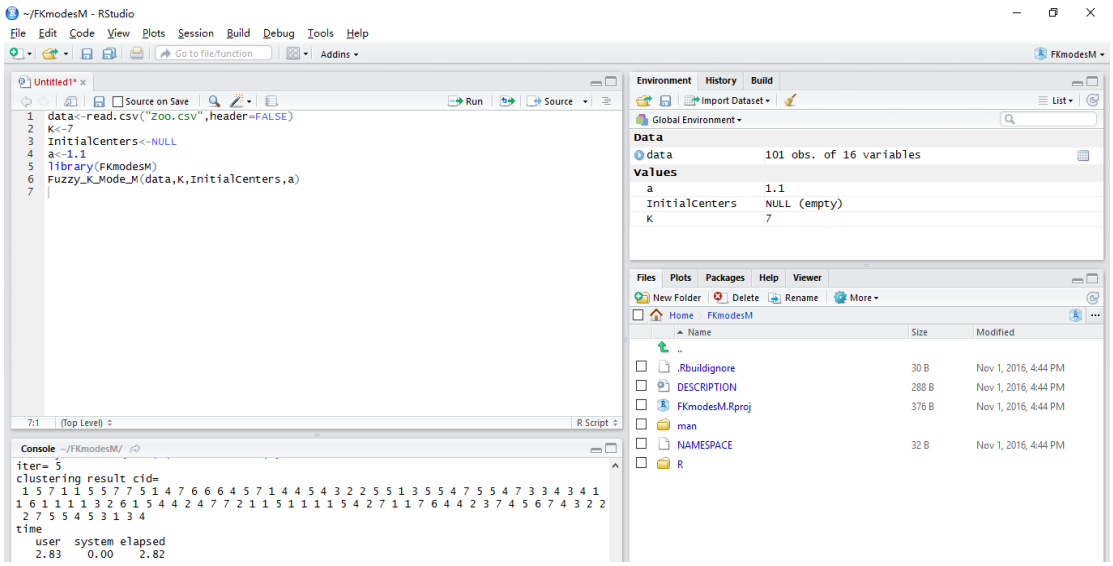

Figure 15 Example of the Fuzzy-SV-K-modes-M function

3.2.6 F-SV-K-modes-D

Description

Implement the fuzzy SV-K-modes algorithm for distribute computation.

Usage

Fuzzy-ratio-k-multi-modes-D (Data, K, InitialCenters)

Return (nj)

Arguments

Inputs	Data	Data set to be clustered
	K	Number of clusters
	InitialCenters	Initial cluster centers
	a	index
Outputs	nj	Clustering result

Details

Please refer to the description mentioned above.

Reference

Fuyuan Cao, Joshua Zhexue Huang, Jiye Liang, Xingwang Zhao and Yinfeng Meng,SV-k-modes: an Algorithm for clustering Categorical Data with Set-Valued Attributes, IEEE Transactions on meural networks and Learning system. Vol. 13. NO. 9. April. 2015

Example

```
>library ( FSVKD )
>Data=read.csv ( " MBD.csv " , header=FALSE )
>K=10
>InitialCenters=NULL
>a==1.1
>Fuzzy-ratio-k-multi-modes-D ( Data, K, InitialCenters, a)
```

Results

```
 [1,]  1  1
 [2,]  1  2
 [3,]  1  3
 [4,]  1  4
 [5,]  1  5
 [6,]  1  6
 [7,]  2  7
 [8,]  2  1
 [9,]  2  8
[10,]  2  9
[11,]  2  3
[12,]  2  2
[13,]  3 10
[14,]  3  8
[15,]  3 11
[16,]  3  9
[17,]  3  4
[18,]  3  5
```

Figure 16 Example of the Fuzzy-SV-K-modes -D function

3.3 Initial class center selection function

3.3.1 K-modes-k-initial-center

Description

The initialization clustering algorithm to select the initial cluster centers for the k-modes algorithm.

Usage

k-initial-center<-function(data,K)

return(DDist)

Arguments

Input	data	A vector in data matrix
	K	Number of clusters
Output	DDist	Initial cluster centers

Details

The initialization clustering algorithm for categorical data is given to choose first class center by calculating the average density of the objects. If selecting more initial centers, it needs to consider the density of an object and the distance between the object and the selected initial cluster centers simultaneously.

Reference

Fuyuan Cao,Jiye Liang,Liang Bai. A new initialization method for categorical data clustering, Expert Systems with Applications, 2009, 36(7):10223-10228

Example

>Data=read.csv("Zoo.csv",header=FALSE)

>K=7

Result

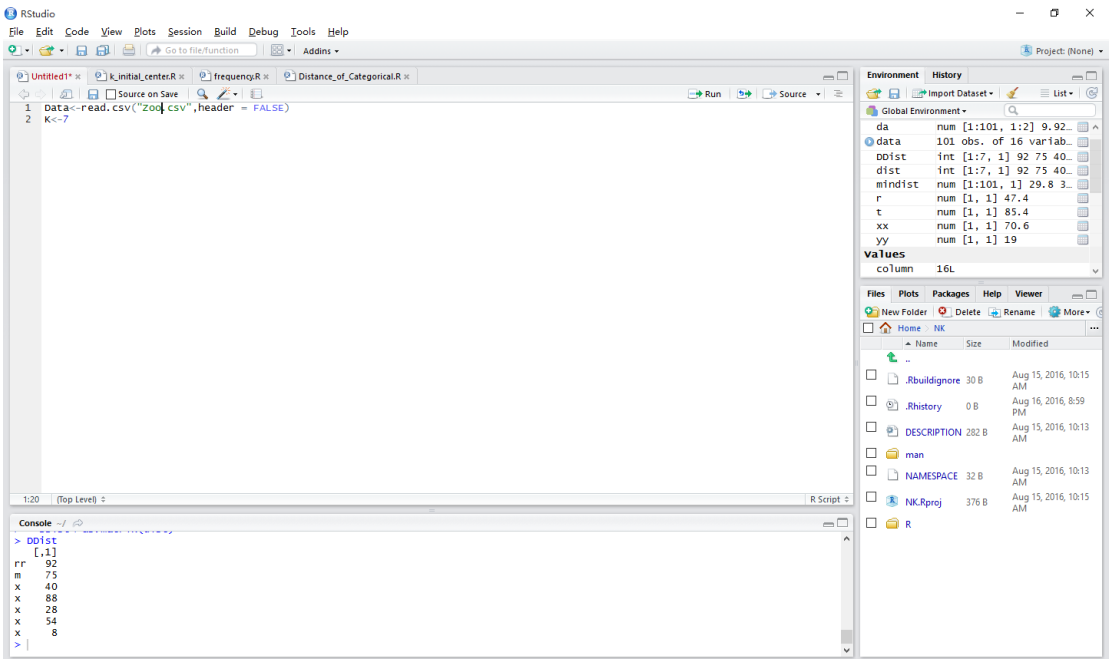

Figure 17 Example of the K-modes-k-initial-center function

3.3.2 SV-K-modes-k-initial-center

Description

The initialization algorithm to select the initial cluster centers for the SV-k-modes algorithm.

Usage

K-initial-centers<-function(Data,K)

Return(dist)

Arguments

Input	Data	A vector in data matrix
	K	Number of clusters
Output	dist	Initial cluster centers

Details

The initialization algorithm is to select the initial cluster centers for the SV-k-modes algorithm by extending the initialization method. If selecting more initial centers, it needs to consider the density of an object and the distance between the object and the selected initial cluster centers simultaneously.

Reference

Fuyuan Cao, Joshua Zhexue Huang, Jiye Liang, Xingwang Zhao and Yinfeng Meng,SV-k-modes: an Algorithm for clustering Categorical Data with Set-Valued Attributes, IEEE Transactions on meural networks and Learning system. Vol. 13. NO. 9. April. 2015

Example

>Data=read.csv("MBD2.csv", header=FALSE)

>K=3

Result

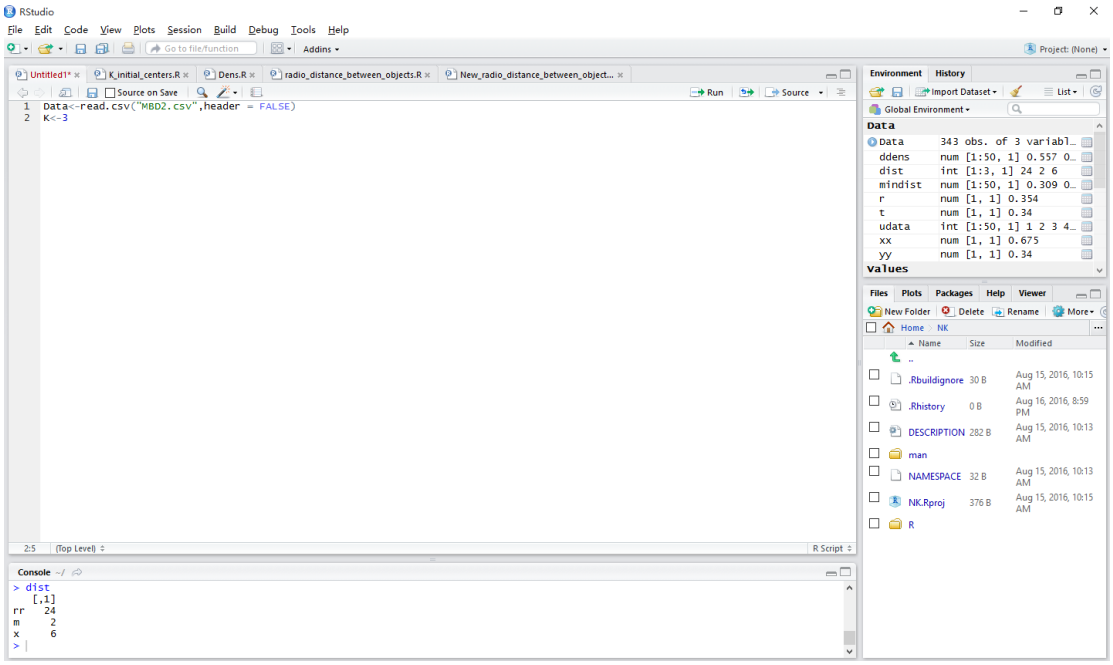

Figure 18 Example of the K-modes-k-initial-center function

3.4 Sub function

3.4.1 K-dis

Description

Calculate the distance between each object to the class center.

Usage

Distance-of-Categorical (Data1, Data2)

Return (distance)

Arguments

Inputs	Data1	A vector in data matrix
	Data2	A vector in data matrix
Output	distance	The distance between Data1 and Data2

Details

A simple 0 - 1 matching method is used to calculate the distance between the two attributes of different objects in the same classification.

Example

>Data=read.csv (" Zoo.csv " , header=FALSE)

>Distance-of-Categorical (Data [1:10,] , Data2 [10:20 ,])

Results

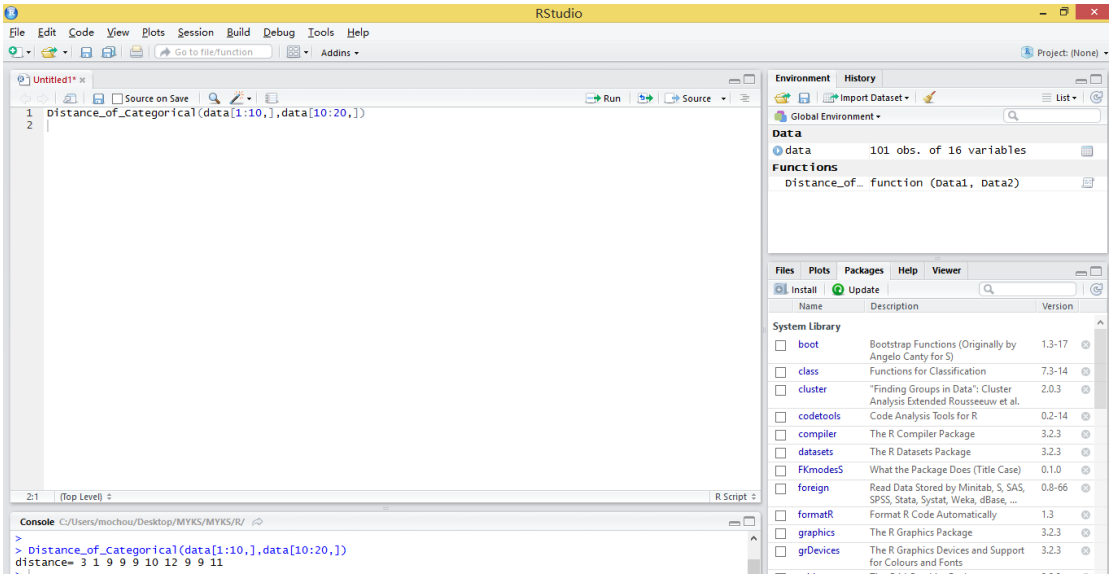

Figure 19 Example of the K-dis function

3.4.2 K-center

Description

Calculate the cluster centers for each class.

Usage

Find-Mode (Data)

Return (NNew-Mode)

Arguments

Input	Data	A vector in data matrix
Output	NNew-Mode	The cluster centers for Data

Details

Calculate the maximum number of frequencies in each attribute as the center of each class.

Example

>Data=read.csv (" Zoo.csv ", header=FALSE)

>Find-Mode (Data [1:30,])

Results

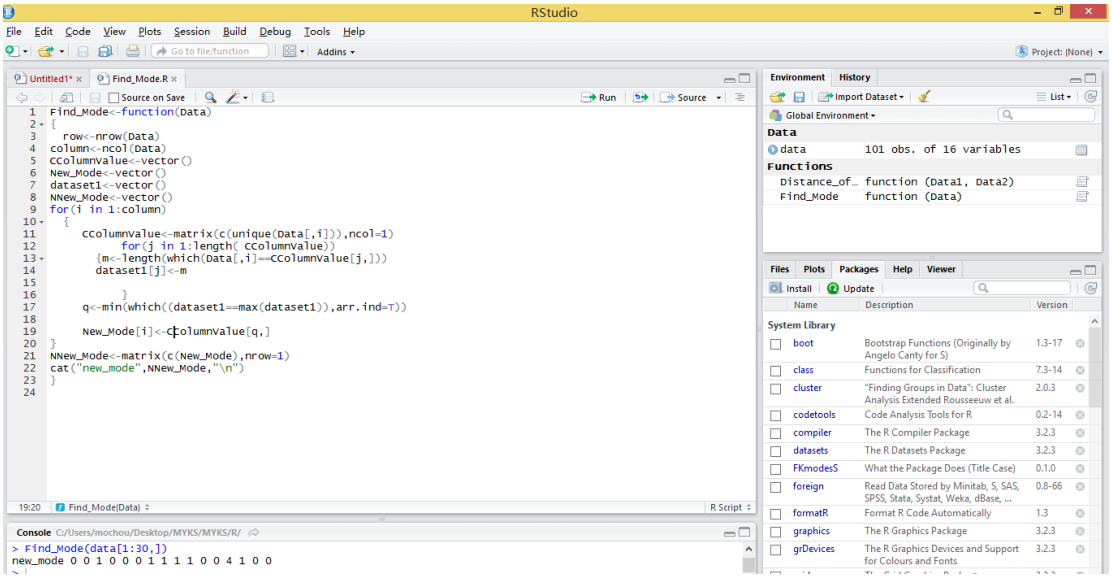

Figure 20 Example of the K-center function

3.4.3 SV-K-dis

Description

The distance function of SV-K-modes algorithm.

Usage

New-ratio-distance-between-objects (DData1, DData2)

Return (distance)

Arguments

Inputs	DData1	A vector in data matrix
	DData2	A vector in data matrix
Output	distance	The distance between Data1 and Data2

Details

Calculate the distance between each object to the center of each class.

Reference

Fuyuan Cao, Joshua Zhexue Huang, Jiye Liang, Xingwang Zhao and Yinfeng Meng. “SV-k-modes: an Algorithm for Clustering Categorical Data with Set-valued Attributes,” Neural Networks and Learning System. IEEE Transaction, vol.13.No.9,April 2015

Example

>DData1=c(1,3,5,6,8,9,2)

>DData2=c(2,4,5,7,8,1,9)

>New-ratio-distance-between-objects(DData1,DData2)

Results

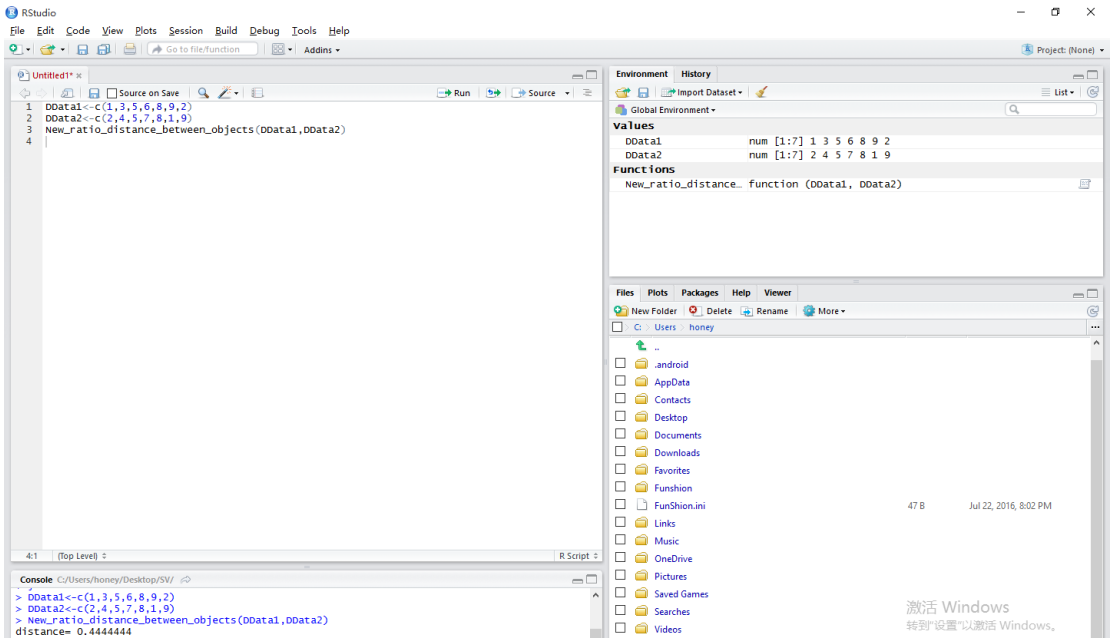

Figure 21 Example of the SV-K-dis function

3.4.4 SV-K-center

Description

The find mode function of SV-K-modes algorithm.

Usage

My-Nean-New-ratio-computing-modes(data)

return(Mode)

Arguments

Input	data	A vector in data matrix
Output	Mode	The cluster centers for data

Reference

Fuyuan Cao, Joshua Zhexue Huang, Jiye Liang, Xingwang Zhao and Yinfeng Meng,SV-k-modes: an Algorithm for clustering Categorical Data with Set-Valued Attributes, IEEE Transactions on meural networks and Learning system. Vol. 13. NO. 9. April. 2015

Example

>Data=read.csv("MBD1.csv",header=FALSE)

>My-Nean-New-ratio-computing-modes(data)

Results

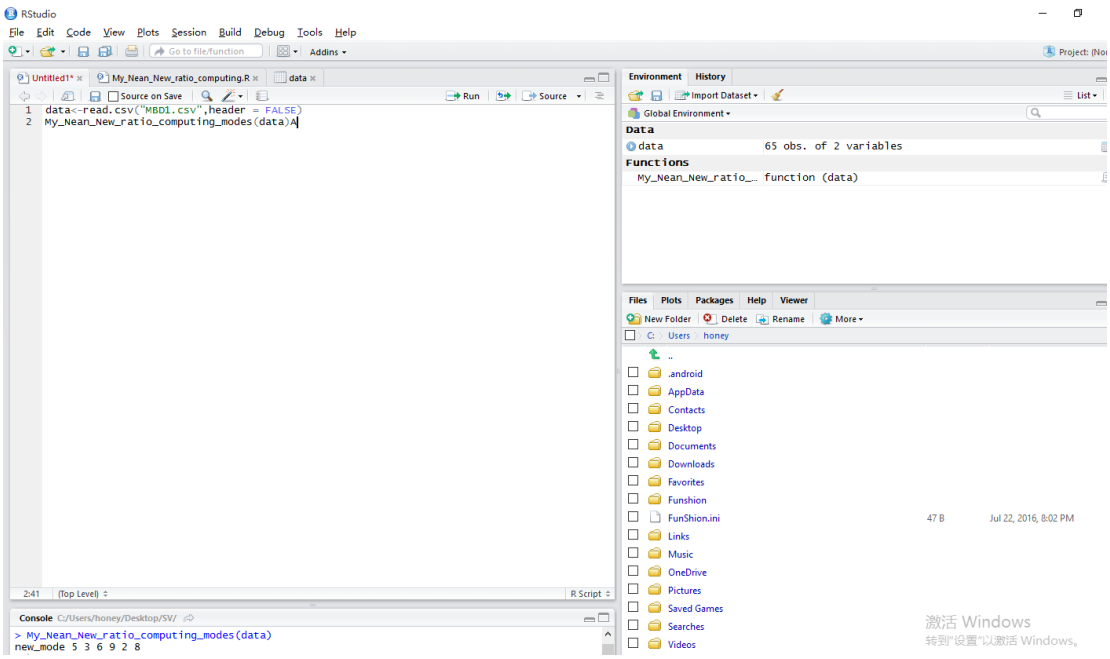

Figure 22 Example of the SV-K-center function