Package 'metaCluster'

October 13, 2022

Type Package

Title Metagenomic Clustering

Version 0.1.0

Maintainer Dipro Sinha <diprosinha@gmail.com>

Description Clustering in metagenomics is the process of grouping of microbial contigs in species specific bins. This package contains functions that extract genomic features from metagenome data, find the number of clusters for that given data and find the best clustering algorithm for binning.

License GPL-3

Encoding UTF-8

LazyData true

RoxygenNote 7.1.2

Imports factoextra, cluster, dbscan, dplyr, seqinr, Biostrings

Depends R (>= 3.6)

NeedsCompilation no

Author Dipro Sinha [aut, cre],
Sayanti Guha Majumdar [aut],
Anu Sharma [aut],
Dwijesh Chandra Mishra [aut]

Repository CRAN

Date/Publication 2021-09-30 08:10:02 UTC

R topics documented:

clust.suite *Determination of Suitable Clustering Algorithm for Metagenomics Data*

Description

This function will give the best clustering algorithm for a given metagenomics data based on silhouette index for kmeans clustering, kmedoids clustering, fuzzy kmeans clsutering, DBSCAN clustering and hierarchical clsutering.

Usage

```
clust.suite(data, k, eps, minpts)
```

Arguments

data	Feature matrix consisting of different genomic features.Each row represents features corresponding to a particular individual or contig and each column represents different genomic features.
k	Optimum number of clusters
eps	Radius value for DBSCAN clustering
minpts	Minimum point value of DBSCAN clustering

Value

kmeans	Output of kmeans clustering
kmedoids	Output of kmedoids clustering
fkmeans	Output of fuzzy kmeans clustering
dbscan	Output of dbscan clustering
hierarchical	Output of hierarchical clustering
silhouette.kmeans	Silhouette plot of kmeans clustering
silhouette.kmedoids	Silhouette plot of kmedoids clustering
silhouette.fkmeans	Silhouette plot of fuzzy kmeans clustering
silhouette.dbscan	Silhouette plot of dbscan clustering
silhouette.hierarchical	Silhouette plot of hierarchical clustering
best.clustering.method	Best clustering algorithm based on silhouette index
silhouette.summary	Average silhouette width of each clustering algorithm

Author(s)

Dipro Sinha <<diprosinha@gmail.com>>,Sayanti Guha Majumdar, Anu Sharma, Dwijesh Chandra Mishra

Examples

```
library(metaCluster)
data(metafeatures)
result <- clust.suite(metafeatures[1:200,],8,0.5,10)
```

GC.content *Calculation of GC content*

Description

This function will calculate GC content from each sequence or contigs of a FASTA file.

Usage

```
GC.content(fasta_file)
```

Arguments

fasta_file	Name of the fasta or multifasta file

Value

Value of the GC content of each sequence or contig.

Author(s)

Dipro Sinha <<diprosinha@gmail.com>>,Sayanti Guha Majumdar, Anu Sharma, Dwijesh Chandra Mishra

Examples

```
library(metaCluster)
library(seqinr)
sample_data <- read.fasta(file = system.file("extdata/sample1.fasta", package = "metaCluster"),
seqtype = "DNA")
gc <- GC.content(sample_data)
```

`metafeatures` *Metagenomic data*

Description

Feature matrix consisting of different genomic features.Each row represents features corresponding to a particular individual or contig and each column represents different genomic features.

Usage

```
data("metafeatures")
```

Format

A data frame with 1196 observations on the following 8 variables.

`class` a factor with levels `contig-0` `contig-1000000` `contig-10000000` `contig-100000000` `contig-1000000000` `contig-1001000000` `contig-1002000000` `contig-1003000000` `contig-1004000000` `contig-1005000000` `contig-1006000000` `contig-1007000000` `contig-1008000000` `contig-1009000000` `contig-101000000` `contig-1010000000` `contig-1011000000` `contig-1012000000` `contig-1013000000` `contig-1014000000` `contig-1015000000` `contig-1016000000` `contig-1017000000` `contig-1018000000` `contig-1019000000` `contig-102000000` `contig-1020000000` `contig-1021000000` `contig-1022000000` `contig-1023000000` `contig-1024000000` `contig-1025000000` `contig-1026000000` `contig-1027000000` `contig-1028000000` `contig-1029000000` `contig-103000000` `contig-1030000000` `contig-1031000000` `contig-1032000000` `contig-1033000000` `contig-1034000000` `contig-1035000000` `contig-1036000000` `contig-1037000000` `contig-1038000000` `contig-1039000000` `contig-104000000` `contig-1040000000` `contig-1041000000` `contig-1042000000` `contig-1043000000` `contig-1044000000` `contig-1045000000` `contig-1046000000` `contig-1047000000` `contig-1048000000` `contig-1049000000` `contig-105000000` `contig-1050000000` `contig-1051000000` `contig-1052000000` `contig-1053000000` `contig-1054000000` `contig-1055000000` `contig-1056000000` `contig-1057000000` `contig-1058000000` `contig-1059000000` `contig-106000000` `contig-1060000000` `contig-1061000000` `contig-1062000000` `contig-1063000000` `contig-1064000000` `contig-1065000000` `contig-1066000000` `contig-1067000000` `contig-1068000000` `contig-1069000000` `contig-107000000` `contig-1070000000` `contig-1071000000` `contig-1072000000` `contig-1073000000` `contig-1074000000` `contig-1075000000` `contig-1076000000` `contig-1077000000` `contig-1078000000` `contig-1079000000` `contig-108000000` `contig-1080000000` `contig-1081000000` `contig-1082000000` `contig-1083000000` `contig-1084000000` `contig-1085000000` `contig-1086000000` `contig-1087000000` `contig-1088000000` `contig-1089000000` `contig-109000000` `contig-1090000000` `contig-1091000000` `contig-1092000000` `contig-1093000000` `contig-1094000000` `contig-1095000000` `contig-1096000000` `contig-1097000000` `contig-1098000000` `contig-1099000000` `contig-11000000` `contig-110000000` `contig-1100000000` `contig-1101000000` `contig-1102000000` `contig-1103000000` `contig-1104000000` `contig-1105000000` `contig-1106000000` `contig-1107000000` `contig-1108000000` `contig-1109000000` `contig-111000000` `contig-1110000000` `contig-1111000000` `contig-1112000000` `contig-1113000000` `contig-1114000000` `contig-1115000000` `contig-1116000000` `contig-1117000000` `contig-1118000000` `contig-1119000000` `contig-112000000` `contig-1120000000` `contig-1121000000` `contig-1122000000` `contig-1123000000` `contig-1124000000` `contig-1125000000` `contig-1126000000` `contig-1127000000` `contig-1128000000` `contig-1129000000` `contig-113000000` `contig-1130000000` `contig-1131000000` `contig-1132000000` `contig-1133000000` `contig-1134000000` `contig-1135000000`

contig-1136000000 contig-1137000000 contig-1138000000 contig-1139000000 contig-114000000
contig-1140000000 contig-1141000000 contig-1142000000 contig-1143000000 contig-1144000000
contig-1145000000 contig-1146000000 contig-1147000000 contig-1148000000 contig-1149000000
contig-115000000 contig-1150000000 contig-1151000000 contig-1152000000 contig-1153000000
contig-1154000000 contig-1155000000 contig-1156000000 contig-1157000000 contig-1158000000
contig-1159000000 contig-116000000 contig-1160000000 contig-1161000000 contig-1162000000
contig-1163000000 contig-1164000000 contig-1165000000 contig-1166000000 contig-1167000000
contig-1168000000 contig-1169000000 contig-117000000 contig-1170000000 contig-1171000000
contig-1172000000 contig-1173000000 contig-1174000000 contig-1175000000 contig-1176000000
contig-1177000000 contig-1178000000 contig-1179000000 contig-118000000 contig-1180000000
contig-1181000000 contig-1182000000 contig-1183000000 contig-1184000000 contig-1185000000
contig-1186000000 contig-1187000000 contig-1188000000 contig-1189000000 contig-119000000
contig-1190000000 contig-1191000000 contig-1192000000 contig-1193000000 contig-1194000000
contig-1195000000 contig-1196000000 contig-1197000000 contig-1198000000 contig-1199000000
contig-12000000 contig-120000000 contig-1200000000 contig-1201000000 contig-1202000000
contig-1203000000 contig-1204000000 contig-1205000000 contig-1206000000 contig-1207000000
contig-1208000000 contig-1209000000 contig-121000000 contig-1210000000 contig-1211000000
contig-1212000000 contig-1213000000 contig-1214000000 contig-1215000000 contig-1216000000
contig-1217000000 contig-1218000000 contig-1219000000 contig-122000000 contig-1220000000
contig-1221000000 contig-1222000000 contig-1223000000 contig-1224000000 contig-1225000000
contig-1226000000 contig-1227000000 contig-1228000000 contig-1229000000 contig-123000000
contig-1230000000 contig-1231000000 contig-1232000000 contig-1233000000 contig-1234000000
contig-1235000000 contig-1236000000 contig-1237000000 contig-1238000000 contig-1239000000
contig-124000000 contig-1240000000 contig-1241000000 contig-1242000000 contig-1243000000
contig-1244000000 contig-1245000000 contig-1246000000 contig-1247000000 contig-1248000000
contig-1249000000 contig-125000000 contig-1250000000 contig-1251000000 contig-1252000000
contig-1253000000 contig-1254000000 contig-1255000000 contig-1256000000 contig-1257000000
contig-1258000000 contig-1259000000 contig-126000000 contig-1260000000 contig-1261000000
contig-1262000000 contig-1263000000 contig-1264000000 contig-1265000000 contig-1266000000
contig-1267000000 contig-1268000000 contig-1269000000 contig-127000000 contig-1270000000
contig-1271000000 contig-1272000000 contig-1273000000 contig-1274000000 contig-1275000000
contig-1276000000 contig-1277000000 contig-1278000000 contig-1279000000 contig-128000000
contig-1280000000 contig-1281000000 contig-1282000000 contig-1283000000 contig-1284000000
contig-1285000000 contig-1286000000 contig-1287000000 contig-1288000000 contig-1289000000
contig-129000000 contig-1290000000 contig-1291000000 contig-1292000000 contig-1293000000
contig-1294000000 contig-1295000000 contig-1296000000 contig-1297000000 contig-1298000000
contig-1299000000 contig-13000000 contig-130000000 contig-1300000000 contig-1301000000
contig-1302000000 contig-1303000000 contig-1304000000 contig-1305000000 contig-1306000000
contig-1307000000 contig-1308000000 contig-1309000000 contig-131000000 contig-1310000000
contig-1311000000 contig-1312000000 contig-1313000000 contig-1314000000 contig-1315000000
contig-1316000000 contig-1317000000 contig-1318000000 contig-1319000000 contig-132000000
contig-1320000000 contig-1321000000 contig-1322000000 contig-1323000000 contig-1324000000
contig-1325000000 contig-1326000000 contig-1327000000 contig-1328000000 contig-1329000000
contig-133000000 contig-1330000000 contig-1331000000 contig-1332000000 contig-1333000000
contig-1334000000 contig-1335000000 contig-1336000000 contig-1337000000 contig-1338000000
contig-1339000000 contig-134000000 contig-1340000000 contig-1341000000 contig-1342000000
contig-1343000000 contig-1344000000 contig-1345000000 contig-1346000000 contig-1347000000
contig-1348000000 contig-1349000000 contig-135000000 contig-1350000000 contig-1351000000

contig-1352000000 contig-1353000000 contig-1354000000 contig-1355000000 contig-1356000000
contig-1357000000 contig-1358000000 contig-1359000000 contig-136000000 contig-1360000000
contig-1361000000 contig-1362000000 contig-1363000000 contig-1364000000 contig-1365000000
contig-1366000000 contig-1367000000 contig-1368000000 contig-1369000000 contig-137000000
contig-1370000000 contig-1371000000 contig-1372000000 contig-1373000000 contig-1374000000
contig-1375000000 contig-1376000000 contig-1377000000 contig-1378000000 contig-1379000000
contig-138000000 contig-1380000000 contig-1381000000 contig-1382000000 contig-1383000000
contig-1384000000 contig-1385000000 contig-1386000000 contig-1387000000 contig-1388000000
contig-1389000000 contig-139000000 contig-1390000000 contig-1391000000 contig-1392000000
contig-1393000000 contig-1394000000 contig-1395000000 contig-1396000000 contig-1397000000
contig-1398000000 contig-1399000000 contig-14000000 contig-140000000 contig-1400000000
contig-1401000000 contig-1402000000 contig-1403000000 contig-1404000000 contig-1405000000
contig-1406000000 contig-1407000000 contig-1408000000 contig-1409000000 contig-141000000
contig-1410000000 contig-1411000000 contig-1412000000 contig-1413000000 contig-1414000000
contig-1415000000 contig-1416000000 contig-1417000000 contig-1418000000 contig-1419000000
contig-142000000 contig-1420000000 contig-1421000000 contig-1422000000 contig-1423000000
contig-1424000000 contig-1425000000 contig-1426000000 contig-1427000000 contig-1428000000
contig-1429000000 contig-143000000 contig-1430000000 contig-1431000000 contig-1432000000
contig-1433000000 contig-1434000000 contig-1435000000 contig-1436000000 contig-1437000000
contig-1438000000 contig-1439000000 contig-144000000 contig-1440000000 contig-1441000000
contig-1442000000 contig-1443000000 contig-1444000000 contig-1445000000 contig-1446000000
contig-1447000000

`Dim.1` a numeric vector

`Dim.2` a numeric vector

`Dim.3` a numeric vector

`Dim.4` a numeric vector

`Dim.5` a numeric vector

`Dim.6` a numeric vector

`gc` a numeric vector

`oligo.freq` *Oligonucleotide Frequency*

Description

This function will calculate oligonucleotide frequency of each sequence or contig from a FASTA file.

Usage

```
oligo.freq(fasta_file, f)
```

Arguments

`fasta_file`	Name of the fasta or multifasta file
`f`	Length of the oligonucleotide

Value

Frequency value of each oligonucleotide of length specified by the user

Author(s)

Dipro Sinha <<diprosinha@gmail.com>>,Sayanti Guha Majumdar, Anu Sharma, Dwijesh Chandra Mishra

Examples

```
library(metaCluster)
freq <- oligo.freq(fasta_file = system.file("extdata/sample1.fasta", package = "metaCluster"),4)
```

---

opt.clust.num *Finding Optimum Number of Cluster for Metagenomics Data*

Description

This function will give optimum number of clusters based on Within Sum of Squares (wss) plot.

Usage

```
opt.clust.num(data, nc, seed = 1234)
```

Arguments

data	Feature matrix consisting of different genomic features.Each row represents features corresponding to a particular individual or contig and each column represents different genomic features.
nc	Probable number of clusters
seed	Seed value for iteration

Value

WSS plot

Author(s)

Dipro Sinha <<diprosinha@gmail.com>>,Sayanti Guha Majumdar, Anu Sharma, Dwijesh Chandra Mishra

Examples

```
library(metaCluster)
data(metafeatures)
wss_plot <- opt.clust.num(metafeatures[1:200,], nc=10, seed = 1234)
```

Index