Package: isokernel (via r-universe)

October 10, 2024

Type Package

Title Isolation Kernel

Version 0.1.0

Maintainer Ye Zhu <ye.zhu@ieee.org>

Description Implementation of Isolation kernel (Qin et al. (2019) <doi:10.1609/aaai.v33i01.33014755>).

License GPL-3

URL https://github.com/zhuye88/isokernel

BugReports https://github.com/zhuye88/isokernel/issues

Depends R (>= 3.5.0)

Imports RANN (>= 2.6.1), Matrix (>= 1.3-4)

Encoding UTF-8

Language en-US

LazyData true

RoxygenNote 7.1.2

Repository https://zhuye88.r-universe.dev

RemoteUrl https://github.com/zhuye88/isokernel

RemoteRef HEAD

RemoteSha 33d1b071159522f3e2e3b551c048233a71ff4249

Contents

1

IKFeature	*Build Isolation Kernel feature vector representations via the feature map for a given dataset.*

Description

Isolation kernel is a data dependent kernel measure that is adaptive to local data distribution and has more flexibility in capturing the characteristics of the local data distribution. It has been shown promising performance on density and distance-based classification and clustering problems.

This version uses Voronoi diagrams to split the data space and calculate Isolation kernel Similarity, following the paper: Qin, X., Ting, K.M., Zhu, Y. and Lee, V.C., 2019, July. Nearest-neighbour-induced isolation similarity and its impact on density-based clustering. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, pp. 4755-4762). Based on this implementation, the feature in the Isolation kernel space is the index of the cell in Voronoi diagrams. Each point is represented as a binary vector such that only the cell the point falling into is 1.

Usage

```
IKFeature(data, Sdata = data, psi = 64, t = 200, Sp = TRUE)
```

Arguments

data	A dataset used for applying Isolation kernel function. The data is a n by d matrix, where n is the data size, d is the dimensionality.
Sdata	The dataset use for generating Voronoi diagrams, it can be the same as the input data.
psi	The number of cells in each Voronoi diagram, it should be large if there are more clusters or more complex structures in the data. It could be [2, 4, 8, 16, 32, 64, 128, 256, 512, 1024].
t	The number of Voronoi diagrams, the higher the more stable the result.
Sp	Indicating whether return the features as a sparse matrix.

Value

The finite binary features based on the kernel feature map. The features are organised as a n by psi*t matrix.

Examples

```
library(isokernel)
df <- matrix(1:50, nrow = 5, ncol = 10)
IKFeatures <- IKFeature(data=df,psi=4,t=200)
```

| IKSimilarity | *Calculate pairwise Isolation Kernel Similarity for a given dataset* |

Description

Isolation kernel is a data dependent kernel measure that is adaptive to local data distribution and has more flexibility in capturing the characteristics of the local data distribution. It has been shown promising performance on density and distance-based classification and clustering problems.

This version uses Voronoi diagrams to split the data space and calculate Isolation kernel Similarity, following the paper: Qin, X., Ting, K.M., Zhu, Y. and Lee, V.C., 2019, July. Nearest-neighbour-induced isolation similarity and its impact on density-based clustering. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, pp. 4755-4762). Based on this implementation, the higher the probability of two data points (x and y) falling into the same cell of a Voronoi diagram, the more the similar between the two points. Therefore, Isolation kernel is adaptive to the local density, i.e., two points are less likely to fall into the same cell unless they are very close in a dense region, because more cells are generated in the dense region than in the sparse region.

Usage

```
IKSimilarity(data, Sdata = data, psi = 64, t = 200)
```

Arguments

data	A dataset used for applying Isolation kernel function. The data is a n by d matrix, where n is the data size, d is the dimensionality.
Sdata	The dataset use for generating Voronoi diagrams, it can be the same as the input data.
psi	The number of cells in each Voronoi diagram, it should be large if there are more clusters or more complex structures in the data. It could be [2, 4, 8, 16, 32, 64, 128, 256, 512, 1024].
t	The number of Voronoi diagrams, the higher the more stable the result

Value

A n by n similarity matrix based on Isolation kernel. The similarity matrix is the inner products between all pairs of data in the feature space. The feature vectors in the Isolation kernel space are built by IKFeature function.

Examples

```
### 1. calculate the pairwise Isolation kernel similarity in the iris dataset
library(isokernel)
df <- iris
SimMatrix <- IKSimilarity(data=df[,1:4],psi=4,t=200)

### 2. calculate the Isolation kernel similarity between A and B
library(isokernel)
```

```
A <- iris[1:10,1:4]
B <- iris[21:40,1:4]
S <- rbind(A,B)
t <- 200
FA <- IKFeature(A,S,psi=4,t=200) # Kernel space features for A
FB <- IKFeature(B,S,psi=4,t=200) # Kernel space features for B
SimAB <- FA%*%t(as.matrix(FB))/t  # dot product on FA and FB
```

Index