

hmds: An R Package for Heuristic High and Multi Dimensional Scaling

A Vignette

Masashi Okada

Okada Algorithm Private Invention Research Laboratory

okadaalgorithm@gmail.com

hmds version 1.0 2016-09-27

Abstract

In this paper, we propose the heuristic in order to calculate the coordinates in high dimensions, if the similarities or distances between the two points and dimensions in the space are given, this heuristic calculates the approximate coordinates in high dimensions. And if the similarities or distances have a contradict in metric space, the method calculates the approximate coordinates. Its coordinates are available for lots of analysis. This heuristic proposed by R package.

Introduction

Multi-Dimensional Scaling(MDS)(Borg 2005) is a statistical method in order to find a data structure. If the similarities or distances between the two points are given, MDS can map the data to two or three dimension coordinate space. In this package, I proposed the heuristic in order to calculate coordinates in high dimensional space from the data of similarities or distance between two points. This heuristic calculates the approximate coordinates in the dimension determined by user. And if the similarities or distances have a contradict in metric space, the method calculates the approximate coordinates. And several important methods like Clustering(Liu 2007), Nearest Neighbor Search(Zezula 2006) and Data Visualization(Fry 2007) require the coordinates in high dimensions as metric space. And this heuristic acts as below. First of all, randomly map the points in the high dimensional space. Then a distance between two points compare with the given data in order. If distance is longer than the data, the distance is made shorter by moving coordinates. If distance is shorter than the data, the distance is made longer. This iteration continues until a sum of distance becomes smaller than approximate rate or a limit of iteration count. As a result, an approximate coordinate points in all data are acquired.

Installation

```
> install.packages("hmds")
```

If download from Github you can use devtools by the commands:

```
> library(devtools)
> install_github("jirotubuyaki/hmds")
```

Once the packages are installed, it needs to be made accessible to the current R session by the commands:

```
> library(hmds)
```

For online help facilities or the details of a particular command (such as the function hmds) you can type:

```
> help(package="hmds")
```

Method

This package has only one method. And it is excused by:

```
> output <- hmds(data = input, dim=20, approx=1.2, itera=10000)
```

Let's args be

- data is a numeric symmetric matrix of input data. it describe the distance between two points.
- dim describe dimentions of distance space between two poitns.
- approx is approximate rate of difference between input distance and output distanced. If the difference between input and output are shoter than approximate rate, iterations are halt.
- itera is iteration numbers to search space between all each two points.

Then let's return be

- output is a numeric matrix of points in a distance space. row is items. col is dimentions.

Data

This package includes a sample dataset. The dataset containing a matrix of similarity between two points. The dataset is generated by R. Please check the data like this:

```
> data(package="hmds")
```

Conclusions

The heuristic for Multi dimentional Scaling is described and explain how to use. This package can produce the approximate coordinates in high dimentions. And several improvements are planed. Please send suggestion and report bugs to okadaalgorithm@gmail.com.

References

- Borg, Groenen, I. 2005. "Modern Multidimensional Scaling." *Springer-Verlag*.
- Fry, Ben. 2007. "Visualizing Data Exploring and Explaining Data with the Processing Environment." *O'Reilly Media*.
- Liu, Bingh. 2007. "Web Data Mining Exploring Hyperlinks, Contents, and Usage Data." *Springer-Verlag* pp. 117-146,
- Zezula, Amato, P. 2006. "Similarity Search." *IEEE Transactions on Pattern Analysis and Machine Intelligence*.