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 2016-11-24

Abstract

In this document, I propose a heuristic to calculate the coordinates in high dimensions. If the similarities or distances between two objects and dimensions in the coordinate space are given, The heuristic calculates approximate coordinates in high dimensions. And if the similarities or distances have contradiction in metric space, the heuristic can calculate approximate coordinates. The coordinates are available for lots of analysis. The heuristic is proposed by R package.

Introduction

Multi-Dimensional Scaling(MDS) is a statistical method in order to put objects at coordinates. If the similarities or distances between two objects are given, MDS can put objects into two or three dimensional coordinate space. In this package, I propose a heuristic in order to calculate coordinates in high dimensional space from the data of similarities or distances between two objects. The heuristic calculates approximate coordinates in the dimensions given by user. And if the similarities or distances have contradiction in metric space, the method can calculate approximate coordinates. And several important methods like Clustering(Liu 2007) and Data Visualization(Fry 2007) require coordinates in high dimensions. And the heuristic acts as follows. First of all, the heuristic randomly puts the objects in the high dimensional space. The number of dimensions is given by user. Then the distances between two objects are compared with the given data in turns. If the distance is longer than the distance of two objects in the data, the distances is made shorter by moving the objects in coordinate space. If the distance is shorter than the data, the distance is made longer. The iteration continues until the sum of distances is less than an approximate rate. And if the sum of distances is not less than the rate, the program exits by the limit of iteration count. As a result, approximate coordinate points of all objects are acquired.

Installation

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 pakage 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 similarities or distances between two objects.
- dim describes dimensions of coordinate space.
- approx is approximate rate between the sum of input distances and the sum of output distances. If the rate between input and output are less than approximate rate, iterations are halt.
- itera is iteration numbers to move points in coordinate space.

Then let's return be

- output is a numeric matrix of points in coordinate space. Row is objects. Col is dimensions.

Data

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

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

Conclusions

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

Acknowledgments

This activity would not have been possible without the support of my family and friends. To my family, thank you for lots of encouragement for me and inspiring me to follow my dreams. I am especially grateful to my parents, who supported me all aspects.

References

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,