Ranked data cluster analyzation

Brendan Mikolajczyk, Eric Mustee, Dan Skrodzki



User’s Guide

Table Of Contents

Introduction ..............................................................................1

Startup …………………………………………………………….…2

Loading a File …………………………………………..…..#

Formatting Guidelines …………………………………...#

Opening a Saved Session ………………………….……….#

Settings…...……………… …………………….………….…..……#

Cluster Analyzer……… …….……………………………..#

Data Description ………………...…………………………#

Random Data Generation #

Interpreting the Results ……………………………………………#

Saving/Exporting #

Troubleshooting ……………………………………………….…..#

Glossary …………………………………………………………....#

Introduction

What is a ranking?

Rankings are a relationship between a set of objects such that one object is better, or ranked higher than another. For instance, a list of one's favorite ice cream cones is a ranking.

What does this software do?

This program clusters rankings. This means finding rankings that are similar enough, and grouping them together in a way that makes the most sense. The program computes what are called cluster centers from a set of partial rankings.

Startup

Loading a File

Before running a cluster analysis, a ranked data file must first be loaded in. To load in a file, navigate to the menu bar and select File>Import Rankings> and then browse your computer for an acceptable file type. Only **.txt** files and **.csv** files can be loaded in.

Formatting Guidelines

Acceptable ranked data files must adhere to the following rules:

* File must not contain any non-numeric characters
* Rankings cannot include the number 0 (nonzero integers only)
* Each line represents a pi vector, so no line may contain more than one of the same number
* Each line must stick to one style of appropriate delimiters
  + Appropriate delimiters include:
  + Spaces
  + Commas not followed by spaces
  + Commas followed by spaces

Example of an appropriately formatted file:

1, 2, 3, 4, 5

5,3,-4,1,2

-2 -4 5 3 1

3 5 4

-1 5 4 2

Opening a Saved Session

This program also allows you to save current settings and results into a separate **.rnkr** file. To restore a previously saved session, navigate to the menu bar and select File>Open Session> and then browse your computer for the desired **.rnkr** save file. Once this file is loaded in, all settings and results will be restored from that session.

Settings

Cluster Analyzer

Glossary

**cluster** – group of ranked data that is closely associated with a particular center (σ).

**cluster center (σ)** – a single ranked data element that represents the center of a cluster, which may or may not be an actual element of that cluster.

**complete ranked data** – an ordered set of numbers from 1 to *n*, that represent a particular person’s preferences, where the order in which these numbers determine the ranking of these preferences (first element is top ranked and last element is lowest ranked). These numbers can also be negated to represent an opposition of a particular preference.

**c vector** – list containing cluster weights for the current clusters in the sigma vector.

**data description** – words replacing numbers in ranked data sets to provide concrete meaning to the rankings.

**dispersion parameter** **(λ)** – represents the standard deviation of a particular cluster; in other words, the average distance of every ranked data element to the cluster center.

**partially ranked data** – an incomplete ranked data set; a complete ranked data set that is missing at least one element, where the missing elements represent a neutral preference.

**pi vector** – list containing all the ranked data elements for basis of analyzation.

**q vector** – two-dimensional assignment probability vector that contains probabilities (0 – 1.0) of how closely any one particular ranked data element is associated with a particular cluster.

**sigma vector** – list containing current existing cluster centers (σ).