

Практическая работа №6: Исследование методов многомерного шкалирования

Цель работы

Ознакомиться с методами многомерного шкалирования на основе пакета Statistica.

Основные теоретические положения

Многомерное шкалирование (МНШ) можно рассматривать как альтернативу факторному анализу. Целью последнего, вообще говоря, является поиск и интерпретация «латентных (т.е. непосредственно не наблюдаемых) переменных», дающих возможность пользователю объяснить сходства между объектами, заданными точками в исходном пространстве признаков. Для определенности и краткости, далее, как правило, будем говорить лишь о сходствах объектов, имея ввиду, что на практике это могут быть различия, расстояния или степени связи между ними. В факторном анализе сходства между объектами (например, переменными) выражаются с помощью матрицы (таблицы) коэффициентов корреляций. В методе МНШ дополнительно к корреляционным матрицам, в качестве исходных данных можно использовать произвольный тип матрицы сходства объектов. Таким образом, на входе всех алгоритмов МНШ используется матрица, элемент которой на пересечении её i -й строки и j -го столбца, содержит сведения о попарном сходстве анализируемых объектов (объекта i и объекта j). На выходе алгоритма МНШ получают числовые значения координат, которые приписываются каждому объекту в некоторой новой системе координат (во «вспомогательных шкалах», связанных с латентными переменными, откуда и название МНШ), причем размерность нового пространства признаков существенно меньше размерности исходного (за это собственно и идет борьба).

Логику МНШ можно проиллюстрировать на следующем простом примере. Предположим, что имеется матрица попарных расстояний (т.е. сходства некоторых признаков) между крупными американскими городами. Анализируя матрицу, стремятся расположить точки с координатами городов в двумерном пространстве (на плоскости), максимально сохранив реальные расстояния между ними. Полученное размещение точек на плоскости впоследствии можно использовать в качестве приближенной географической карты США.

В общем случае метод МНШ позволяет таким образом расположить «объекты» (города в данном примере) в пространстве некоторой небольшой размерности (в данном случае она равна двум), чтобы достаточно адекватно воспроизвести наблюдаемые расстояния между ними. В результате можно «измерить» эти расстояния в терминах найденных латентных переменных. Так, в данном примере можно объяснить расстояния в терминах пары географических координат Север/Юг и Восток/Запад.

Общая формулировка задачи

- Ознакомиться с теорией многомерного шкалирования.
- Провести исследование на имеющихся данных.

- Оформить результаты в виде отчёта.

Порядок выполнения работы

1. Запустить пакет Statistica. Войти в меню “Statistics” → “Advanced/Multivariate” → “Mult/Exploratory”:
 1. в выпадающем меню выбрать “Cluster”;
 2. с помощью одного из методов кластер-анализа получить матрицу расстояний, задав следующие параметры для кластеризации в панели меню “Advanced”:
 - выбрать все вектора выборки(cases). В случае выдачи ошибки о слишком большом числе векторов разделить выборку на 2 части, отобрав по половине векторов из каждого класса. Работайте только с одной из частей;
 - установить параметр Cluster = cases;
 - выбрать метод связывания;
 - выбрать метод измерения расстояния;
 - запустить процедуру кластеризации.
 3. Сохранить матрицу расстояний.
2. Запустить модуль статистики “Multidimensional Scaling”:
 1. в меню “Quick” выбрать переменные (all variables);
 2. в этом же меню ввести значение размерности для анализа;
 3. в меню “Options” выбрать конфигурацию “Standard Guttman-Lingoes”, OK;
 4. просмотреть данные в окне Parameter Estimation, OK;
 5. в диалоговом окне “Results” (Результаты) проанализировать результаты многомерного шкалирования:
 - Final Configuration;
 - D-hat Values;
 - D-star Values;
 - Distance Matrix;
 - Summary.
 6. сохранить координаты заключительной конфигурации в стандартном файле данных;
 7. построить график заключительной конфигурации “Graphs final configuration - 2D, 3D”;
 8. построить диаграмму Шепарда зависимости D-hats от расстояний “Graph D-hat vs.Distances”;
 9. построить диаграмму Шепарда зависимости D-stars от расстояний “Graph D-star vs.Distances”.
3. Вернуться к исходной выборке (неразделенной):
 1. Прodelайте п.3-4 для признаков (variables), то есть при формировании матрицы расстояний выбрать все переменные, кроме целевой, и установить параметр Cluster = variables.
 2. При проведении многомерного шкалирования - выбрать все переменные (all variables) и ввести значение размерности для анализа = 2.
 3. Сохранить графики и таблицы результатов исследования.
 4. Сделать сравнительные выводы по проведенным исследованиям.
 5. Сравните результаты шкалирования выборки по признакам с результатами факторного анализа, полученными в предыдущей работе.
4. Оформить результаты в виде отчета.

From:

<http://se.moevm.info/> - **se.moevm.info**

Permanent link:

http://se.moevm.info/doku.php/courses:data_analysis_and_interpretation:task6?rev=1547547014 

Last update: **2022/12/10 09:08**