

Л. В. Гадасина, Л. Ф. Вьюненко (Санкт-Петербург, СПбГУ).
Визуализация как специальный прием в анализе многомерных данных.

УДК 51-77+330.4

Резюме: В работе предлагается подход к визуализации как специальному инструменту анализа многомерных данных. Выделены три типа визуального представления данных: агрегирование, декомпозиция и проекция, каждый из которых требует использования довольно сложного математического аппарата. Для каждого типа визуализации приведены примеры, в частности, возникающие при применении SimDec-технологии, фазовых портретов, используемых при сплайн-анализе экономических данных и многомерного шкалирования в задачах классификации и кластеризации.

Ключевые слова: многомерные данные, визуализация данных, агрегирование, декомпозиция, проекция.

Визуализация традиционно рассматривается как как вспомогательное средство для представления данных в наглядном виде (через таблицы, графики, диаграммы, структурные схемы, информационные панели и т. д.). Однако, во многих современных практических задачах визуализация сама по себе может выступать инструментом интеллектуального анализа данных. Особенностью визуализации в таких задачах является то, что она требует применения довольно сложного математического аппарата к исследуемым данным. Можно выделить три типа визуализации как инструмента анализа данных:

- агрегация,
- декомпозиция,
- проекция.

Типичными примерами визуализации-агрегации являются гистограммы и «ящички с усами», используемые для визуализации вероятностных распределений статистических данных. Декомпозиция подразумевает визуализацию дополнительного признака с помощью, как правило, цветового выделения в исходной визуализации. Третий тип визуального представления данных подразумевает визуализацию проекции n -мерных данных на пространства меньшей размерности с сохранением некоторого заданного признака. Так, например, метод многомерного шкалирования предполагает сохранение при проецировании соотношения расстояний между объектами, визуализируемыми как точки на плоскости (или трехмерном пространстве).

В работе рассмотрены следующие примеры визуализации как инструмента анализа данных:

1. Декомпозированные гистограммы, получаемые в результате применения технологии SimDec [3].
2. Фазовые портреты, используемые в сплайн-анализе экономических данных [4] как инструмент визуализации для обнаружения и анализа циклических конструкций в экономической динамике.
3. Визуализации многомерных данных, получаемые в результате проекции на плоскость, как метод отбора признаков в задаче кластеризации.

СПИСОК ЛИТЕРАТУРЫ

1. *Cox M., Cox T.* Multidimensional Scaling. In: Handbook of Data Visualization. Springer Handbooks Comp. Statistics. Springer, Berlin, Heidelberg, 2008. https://doi.org/10.1007/978-3-540-33037-0_14.
2. *Айвазян С. А., Мхитарян В. С.* Теория вероятностей и прикладная статистика. Т. 1. Прикладная статистика. Основы эконометрики: Учебник для вузов, 2001.// *Ayvazyan S. A., Mkhitaryan V. S.* Probability theory and applied statistics. V. 1. Applied Statistics. Fundamentals of Econometrics: Textbook for universities, 2001 (in Russian).
3. *Kozlova M., Collan M., Luukka P.* Simulation decomposition: New approach for better simulation analysis of multi-variable investment projects. — Fuzzy Econom. Rev., 2016, v. 21(2), p. 3–18.
4. *Винтизенко И. Г., Яковенко В. С.* Экономическая цикломатика. М.: Финансы и статистика, Ставрополь: АГРУС, 2008, 428 с.// *Vintizenko I. G., Yakovenko V. S.* Economic Cyclomatics. M.: Finance and Statistics. Stavropol: AGRUS, 2008, 428 p. (in Russian).

UDC 51-77+330.4

Gadasina L. V., Vyunenko L. F. (St. Petersburg, St. Petersburg State University).
Visualization as a special technique in the multidimensional data analysis

Abstract: In this paper, we propose an approach to visualization as a special tool for analyzing multidimensional data. There are three types of visual representation of data: aggregation, decomposition, and projection, each requires the use of a rather complex mathematical apparatus. For each type of visualization, examples are given, in particular, arising from the SimDec-technology and phase portraits used in spline analysis of economic data, and multidimensional scaling in classification and clustering problems.

Keywords: multidimensional data, data visualization, aggregation, decomposition, projection.