多维数据集中高维数据可视化算法研究

张红军

（贺州学院 计算机科学与信息工程学院，广西 贺州 542899）

摘　要：多维数据集中包含海量的复杂高维数据，如何对这些数据进行科学的理解、分析和使用，是数据挖掘领域的一个重点和难点课题.针对传统可视化算法无法解决高维数据多属性的可视化问题，算法复杂、适用性较差的不足，提出基于快速数据聚类的可视化算法研究.算法构建了一种多维数据模型，采用三角多项式建立映射关系；对多维数据集做不同层次的划分，并基于聚类的效果筛选最佳聚类数；最后在平行坐标系基础上，对高维数据进行时序多维分析，提高算法鲁棒性.实验证明提出算法简洁易用，能够较好地实现对多维数据集中高维数据的理解和分析.

关键词：多维；数据集；可视化；平行坐标系

Cube High-Dimensional Data Visualization Algorithm Research

ZHANG Hong-jun

(College of Computer Science & Information Engineering Hezhou University,Hezhou 542899,China)

Abstract：Cube contains vast amounts of complex data, how to carry on the scientific understanding and analysis of the data and use, is an important and difficult topic in the field of data mining. For traditional visualization algorithm can't solve the problem of high-dimensional data visualization of multiple attribute, the deficiencies of the algorithm is less complex, applicability, visualization algorithm based on fast data clustering is proposed. Algorithm firstly builds a kind of multidimensional data model, trigonometric polynomial is adopted to establish the mapping relation; The cube for different levels of division, and based on clustering effect to screen the optimal clustering number; Finally based on parallel coordinates, temporal multi-dimensional analysis of high-dimensional data, improve the robustness of the algorithm. Experiments prove that the proposed algorithm is concise and easy to use, can well realize the cube understanding and analysis of high-dimensional data.

Key words：multidimensional; data set; visualization; parallel coordinates

作者简介：

张红军　男，(1975-)，硕士，讲师.研究方向为计算机算法与分析、软件工程、计算机教育与应用.　E-mail：bspandsh@163.com.