CS855: Data Visualization (Assignment 4)

Information Visualization using Parallel Coordinates and Tree Maps

Akshay Jindal

IMT2012002

IIIT-B

akshay.jindal@iiitb.org

Abstract—**This report explains the algorithms used for visualization of human liver data and cars evaluation data, details the data used and provides the insight of the visualization.**

Keywords—parallel coordinates, tree maps;

I. INTRODUCTION

This assignment does a multivariate data visualization using parallel coordinates on human liver data and draws a tree map of car evaluation data.

II. DATASET

A. Data Procurement

Both the datasets were obtained from UCI archive of ML datasets. The file format used for reading data is *.csv. Each .csv contains multiple attributes and one final classifier.

B. Data Processing

A separate class DataHandler.pde was written to parse the .csv input file and handle all data operations.

III. VISUALIZATION

There are two types of visualizations:

A. Parallel Coordiantes

Every attribute is give a unique axis with its range normalized to provide consistency in visualization. Each row of data is mapped on the axis (using linear interpolation) by drawing a line between each point a given order. Order of axis/attributes can be changed by dragging the axis.

B. Tree Maps

The data structure used to store the tree map is a rectangle class with an ArrayList of rectangles as an attribute. Rectangles are drawn recursively using a simple slic-and-dice algorithm.

C. User Interaction

- Axes in parallel coordinates can be moved by dragging them with mouse.
- Tree Map can be zoomed in/out by clicking on either half of the screen

IV. INSIGHT

- Parallel coordinates can be used to find correlations, trends, outliers and clusters in dataset.
- Tree maps provides an easy to understand visualization of large datasets.

REFERENCES

[1] http://archive.ics.uci.edu/ml/datasets/Liver+Disorders

[2] http://archive.ics.uci.edu/ml/datasets/Car+Evaluation