CONTENT

- Visualization in KDD
- Interactions in Visualization
- Visualization in preprocessing
- Visualization in Data Mining
  - 1D 2D 3D data
  - Multi-dimensional data
  - Text
  - Hierarchies and Graphs
- Visualization in Evaluation
- Data Mining and Visualization Discussion
- Bibliography
**Visualization in KDD**

- What is visualization?
- Why do data mining need visualization?
- When do data mining need visualization?

**Interactions in Visualization**

- Filtering
  - Be able to filter unwanted data
- Focus and context
  - Get an overview of data and get detailed data on demand
- Brushing and Linking
  - Feedback to the user when marking data
  - Linking different visualization tools
Visualization in Preprocessing

- Overview of data distribution
- Find outliers
- Box plot
  - Available in WEKA
- Stem and leaf plot
  - Bins with number of instances

Visualization in Data Mining

- Data types
  - Number of attributes
  - Nominal / Categorical
- Visualization techniques
  - Standard 2D/3D displays
  - Icon-based
  - Dense pixel display
1D 2D 3D DATA

- Standard visualizations
  - Plots
  - Histograms
  - Graphs
  - Maps
  - Charts
  - Tree maps

1D 2D 3D DATA

- Scatterplot
  - Works well with icon based methods to visualize multi-dimensional data
  - Available in WEKA
1D 2D 3D DATA

- Scatter matrix
  - Matrix of scatter plots
  - Find correlations
  - Available in WEKA

1D 2D 3D DATA

- Treemap
  - Dense pixel method
  - Size corresponds to an attributes percentage of the whole data set
MULTI-DIMENSIONAL DATA

- More than three attributes
- Harder to visualize on the screen
  - Glyphs
  - Parallel coordinates
  - Table lens

MULTI-DIMENSIONAL DATA

- Glyphs
  - Icon based method
  - Uses an arbitrary glyph to visualize many attributes
    - Size
    - Color
    - Shape
  - Famous glyph: Chernoff faces
MULTI-DIMENSIONAL DATA

- Chernoff faces
  - Many attributes builds up a face
  - Uses the minds recognition of faces to see differences

MULTI-DIMENSIONAL DATA

- Parallel coordinates
  - Shows a profile of all data items in the set
  - One axis per attribute
  - One line per data item
  - Filtering for each axis
MULTI-DIMENSIONAL DATA

- Table lens
  - Dense pixel method
  - Super spreadsheet
  - Details on demand
  - Unmarked items only one pixel width
  - Focus and context

TEXT

- Visualizing text or whole documents is not directly possible using any of the methods named above
- Apply some kind of transformation
  - Word occurrences
  - Word frequencies
TEXT

- Plot similarities between text
  - Landscape
  - Galaxy

HIERARCHIES AND GRAPHS

- Connections and relations important
- Example
  - Relations in a homepage
  - Money transfer
**Hierarchies and Graphs**

- **Hyperbolic tree**
  - A hyperbolic tree is a technique capable of showing many relations at once
  - Fish-eye technique
  - Focus and context

- **Flow map**
  - Using width to indicate an attribute

- **Example**
  - Minard’s flow map of Napoleons march
**Visualization in Evaluation**

- Combine methods
- Customize according to usage
- Example
  - SCB eXplorer (NCVA)

**Data Mining and Visualization Discussion**

- Visualization is a useful tool in the KDD process
  - Understand the data
  - Trust the result
  - Find new results
- Can be negative
  - Hard to describe high-dimensional data
  - Hard to visualize big datasets (clutter)
- Limited visualization tools in WEKA
BIBLIOGRAPHY

- TNM048, Information Visualization, 2010