A Survey on Multidimensional Visual Analysis Techniques

Nan Cao @ HKUST
Sep. 2011
Introduction – Motivation

- Real world data contain multiple dimensions
Introduction – Motivation

Data mining techniques

Classification
Clustering
Dimension Reduction

Data table

Clustering
Classification
Dimension Reduction

3
Visualizations aid on multidimensional data

1. Interpretation & comparison
2. Visual pattern detection
3. Analysis
   1. Testing assumptions
   2. Detecting outliers
   3. Selecting analysis factors, models and estimators
Introduction – Challenges

• How to represent the multidimensional data to facilitate interpretation and comparison?
• How to detect and represent the visual patterns?
• How to aid on data analysis by considering user feedback?
Overview – Taxonomy

• Surveyed 125 articles including papers and book chapters

• Item packing techniques
  – Encoding each data item or attribute to a visual primitive (e.g. node, link) or a visual feature (e.g. color, size, shape)

• Statistical embedding techniques
  – Encoding the statistical features (e.g. means, frequency, variance) of a subset of aggregated items to visual primitives or visual features
Outline

• Item packing techniques
  – Geometrical plots
  – Parallel coordinates
  – Pixel oriented techniques
  – Icon based techniques
• Statistical embedding techniques
  – Hierarchical aggregation
  – Distribution visualizations
  – Aggregated traditional design
• Visual analysis systems
  – Systems for dimension reduction
  – Systems for visual diagnostics
  – Systems for visual cluster analysis
Outline

• **Item packing techniques**
  – Geometrical plots
  – Parallel coordinates
  – Pixel oriented techniques
  – Icon based techniques

• Statistical embedding techniques
  – Hierarchical aggregation
  – Distribution visualizations
  – Aggregated traditional design

• Visual analysis systems
  – Systems for dimension reduction
  – Systems for visual diagnostics
  – Systems for visual cluster analysis
(1) Scatter Plots

2D Cartesian coordinate

$Y$ vs $X$

$n (n - 1) / 2$ views for an $n$-dimensional data
A projection \((X \rightarrow Y)\) maps points \(\{x_1, x_2, \ldots, x_m\}\) in an \(n\)-dimensional space into a \(p\)-dimensional space as \(\{y_1, y_2, \ldots, y_m\}\) \((p << n)\) while preserving distance measures of data items.

- Linear projections
- Nonlinear projections
Linear Projections

• General form (Koren & Carmer)

\[
\max \sum_{i<j} \omega_{ij} d_p(x_i, x_j)^2
\]

subject to

\[
V_i^T X^T X V_j = \theta_{i,j}
\]

\[|v_i| = 1\]

• Preserving the variance of the data
• \(w_{ij}\) : the importance of distinguish point i and point j
• Project the points to the plane determined by \(v_1\) and \(v_2\) (\(p = 2\))
Linear Projections

$$\max \sum_{i<j} \omega_{ij} d_p(x_i, x_j)^2$$
subject to
$$V_i^T X^T X V_j = \theta_{ij}$$

PCA : a special case

**standard PCA:** \(\theta_{ij} = 0\), \(\omega_{ij} = 1\)

**weighted PCA:** \(\omega_{ij}\) is not a constant

**normalized PCA:** \(\omega_{ij} = 1 / d_m(x_i, x_j)\)
Nonlinear Projections

Multidimensional Scaling

\[ \min \sum_{i<j} \mu_{ij} (d_p(x_i, x_j) - d_m(f(x_i), f(x_j)))^2 \]

(Sammon Jr., 1969; Jourdan & Melangon, 2004; Morrison & Chalmers, 2004; Williams & Munzner, 2004 etc.)

ThemeSpace (Wise, 1995)
Nonlinear Projections

InfoSky (Andrews et al, 2002)

(Iwata et al., 2008)

(Chen et al., 2009)
(3) Parallel Coordinates – Visual Designs

- Dimensions as parallel axes
- Data items as line segments
- Intersections on the axes indicates the values of the corresponding attributes
Parallel Coordinates – Examples

Cars dataset:
392 data items, 7 dimensions

(Homan, 1977)

(Fanea et al., 2005)

(McDonnell & Mueller, 2008)

(Yuan et al., 2009)

(Rubel et al., 2006)
Parallel Coordinates – Clutter Reduction

- Clustering and filtering approaches
- Dimension reordering approaches
- Visual enhancement approaches
Clutter Reduction – Clustering and Filtering

Filtering on Density & Frequency

Clustering

Sampling

Visual Clustering

(Artero et al., 2004)  (Novotny et al., 2004)  (Ellis & Dix, 2006)  (Zhou et al., 2008)
**Clutter Reduction – Dimension Reordering**

Original order

(Wang et al, 2004)

(Ainkerst et al, 2004)

(Ferdosi & Roerdink, 2010)
Clutter Reduction – Visual Enhancement

(Theisel, 2000)

(Graham & Kennedy, 2003)

(Johansson et al., 2005)

(Novotny & Hauser, 2006)
(4) **Pixel Oriented Techniques – Encoding**

A multidimensional data item contains 6 attributes

Splitting the display into 6 regions (one for each dimension)

An item is represented by multiple pixels in a split manner
Pixel Oriented Techniques

- Database visualization (10,000 items, 6 dimensions)

(Keim & Kriegel, 1994; 1996)
**Pixel Oriented Techniques – Display Region**

- Ways of splitting the display region

 Pixel oriented bar charts

(Keim et al., 2001; 2002)  

(Yang et al., 2006)
(5) Icon Based Techniques – Chernoff Faces

(Chernoff, 1973)
Icon Based Techniques – Star Glyph

attr1

attr2

attr3

attr4

attr5

attr6

attr7

attr8

Cluster

Outlier
## Comparison

<table>
<thead>
<tr>
<th>Visualizations</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scatter Plot Matrices</td>
<td>Simple</td>
<td>1. Visual clutter 2. Unclear patterns</td>
</tr>
<tr>
<td>Parallel Coordinates</td>
<td>Clear visual patterns</td>
<td>Visual Clutter</td>
</tr>
<tr>
<td>Pixel Oriented</td>
<td>No visual clutters Clear visual patterns</td>
<td>Hard for interpretation</td>
</tr>
<tr>
<td>Icon Based</td>
<td>Facilities visual comparison</td>
<td>Support limited numbers of dimensions</td>
</tr>
</tbody>
</table>
## Comparison

<table>
<thead>
<tr>
<th>Visualizations</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scatter Plot Matrices</strong></td>
<td>Simple</td>
<td>1. Visual clutter</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Unclear patterns</td>
</tr>
<tr>
<td><strong>Scalability</strong></td>
<td></td>
<td>3. Obscured semantics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4. Loss of information</td>
</tr>
<tr>
<td><strong>Parallel Coordinates</strong></td>
<td>Clear visual patterns</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5. Visual Clutter</td>
</tr>
<tr>
<td><strong>Pixel Oriented</strong></td>
<td>Clear visual patterns</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>6. Hard for interpretation</td>
</tr>
<tr>
<td><strong>Icon Based</strong></td>
<td>Facilities visual comparison</td>
<td>Support limited numbers of dimensions</td>
</tr>
</tbody>
</table>
Overview – Outline

• Item packing techniques
  – Geometrical plots
  – Parallel coordinates
  – Pixel oriented techniques
  – Icon based techniques

• **Statistical embedding techniques**
  – Hierarchical aggregation
  – Distribution visualizations
  – Aggregated traditional design

• Visual analysis systems
  – Systems for dimension reduction
  – Systems for visual diagnostics
Hierarchical Aggregation

Building a tree of aggregated data items either bottom-up or top-down

(Elmqvist & Fekete, 2010)
(1) Distribution Visualizations

Histogram (Pearson, 1895)

Heatmap (Wilkinson & Friendly, 2009)

InfoCube (Stolte et al., 2003)

(Lin et al., 2010)
(2) Aggregation Based on Traditional Designs

- Scatter Plots
  - (Elmqvist & Fekete, 2010)

- Parallel Coordinates
  - (Kosara et al., 2006)

- Star Plots
  - (Fua et al. 1999)
Overview – Outline

• Item packing techniques
  – Geometrical plots
  – Parallel coordinates
  – Pixel oriented techniques
  – Icon based techniques

• Statistical embedding techniques
  – Hierarchical aggregation
  – Distribution visualizations
  – Aggregated traditional design

• **Visual analysis systems**
  – *Systems for dimension reduction*
  – *Systems for visual diagnostics*
  – *Systems for clustering analysis*
Visual Analysis System

Standard Pipeline for Explorative Visual Analysis
(1) Dimension Reduction Systems

Value & Relation View

Visual attributes in each dimension as an pixel icon

Project the icons based on MDS

Similar dimensions are clustered

Merge the similar dimensions
Delete the outliers

(Yang et al., 2007)
(2) Visual Diagnostics

• Visual diagnostics:
  – Estimates multidimensional visualization views based on a set of measures
  – Recommends the views with interesting visual patterns based on their measurements

• Applications:
  – Scagnostics (*scatter plots + diagnostics*)
  – Pixnostics (*pixel + diagnostics*)
  – Pragnostics (*parallel coordinates + diagnostics*)
**Scagnostics** *(Scatter plots + Diagnostics)*

- Scatter plots diagnostics estimates scatter plot views based on a set of pre-defined measurements.
- Rank views based on these measurements.

![Scagnostics Diagram](image-url)
Pixnistics (pixel + diagnostics)

Examples of pixel measures

Relevant score:

0.8 > 0.5

(schneidewind et al., 2006)
**Pragnostics** *(parallel coordinates + diagnostics)*

<table>
<thead>
<tr>
<th>Number of Crossings</th>
<th>Angle of Crossings</th>
<th>Parallelism</th>
<th>Mutual Information</th>
<th>Convergence Divergence</th>
<th>Over Plotting</th>
</tr>
</thead>
<tbody>
<tr>
<td>NumCrossings</td>
<td>AngleOfCrossings</td>
<td>Parallelism</td>
<td>MutualInformation</td>
<td>Convergence_Divergence</td>
<td>OverPlotting</td>
</tr>
</tbody>
</table>

Ranking the order of dimension pairs  
(Dasgupta & Kosara, 2010)
Visual Cluster Analysis Systems

Hieratical Cluster Explorer

(Seo & Shneiderman, 2002)
Conclusion & New Challenges

• Conclusion
  – Item packing techniques
  – Statistical embedding techniques
  – Visual analysis systems

• New challenges
  – Visualizing high dimensional dataset in a large scale
  – Emerging of new application domains lead to new problems
  – Hybrid research areas - interactive learning, incremental mining
My Research Targets

• Design novel statistical embedding visualizations for new applications
  – Multidimensional visualization for social media data
  – Multidimensional visualization for healthcare applications
  – Multidimensional visualization for text visualizations

• Visual analysis systems for uncovering multidimensional cluster patterns.
Thank you!