Multi-dimensional Visualization based on Bidimensional Mapping

Rosane Minghim

Instituto de Ciências Matemáticas e de Computação
USP-São Carlos

Projection Techniques

\[ X \in \mathbb{R}^m \quad f \quad Y \in \mathbb{R}^{p \in \{1,2,3\}} \]

- \( \delta: x_i, x_j \to \mathbb{R}, x_i, x_j \in X \)
- \( d: y_i, y_j \to \mathbb{R}, y_i, y_j \in Y \)
- \( f: X \to Y, |\delta(x_i, x_j) - d(f(x_i), f(x_j))| \leq 0, \forall x_i, x_j \in X \)

Ex: Mapping to plane of patents

surgery, drugs, molecular bio

Problems PCA

390 dimensions
Problems PCA

Ex: Sammon Mapping

- Let $X$ be the points in the original space $\mathbb{R}^n$, we apply a distance measure $d_{ij}^*$ between $X_i$ and $X_j$, and find $Y$, the projected point, e.g. $\mathbb{R}^2$ and $d_{ij}$ the Euclidean distance between them.

- Sammon's method applies an error function to measure the target.

Force Based Point Placement
Force Scheme [Tejada et al., 2003]

Force Based Point Placement

Force Scheme [Tejada et al., 2003]
Let $V_i = \{p_1, \ldots, p_n\}$ be a neighborhood of a point $p_i$ and let $c_i$ be the coordinates of $p_i$ in $\mathbb{R}^p$

$$c_i - \frac{1}{ki} \sum_{p_j \in V_i} c_j = 0$$

Each $p_i$ is the centroid of points in $V_i$

$LSP$: Laplacian Matrix

$Lx_1 = 0, Lx_2 = 0, \ldots, Lx_p = 0$
Where $x_1, x_2, \ldots, x_p$ are vectors containing the coordinates of the points and $L$ is the matrix given by:

$$L_{ij} = \begin{cases} 
1 & i = j \\
\frac{1}{ki} & p_j \in V_i \\
0 & \text{otherwise}
\end{cases}$$
LSP: Adicionando os Pontos de Controle

\[ A = \begin{pmatrix} L \\ C \end{pmatrix} \quad C_{ij} = \begin{cases} 1 & \text{if } p_j \text{ is a control point} \\ 0 & \text{otherwise} \end{cases} \]

\[ b_i = \begin{cases} 0 & i \leq n \\ x_{p_i} & n < i \leq n + nc \end{cases} \]

Choosing the Control Points

• In order to select the control points
  • the space \( \mathbb{R}^m \) is split into \( nc \) clusters using k-medoids.
  • the control points are the medoids of each cluster

Choosing the Control Points

• Once the control points are chosen, these points are projected onto \( \mathbb{R}^d \) through a fast dimensionality reduction method
  • Fast Projection (Fastmap or NNP)
  • Force Placement
Content – based by Projections

(1)  
(2)  
(3)  

Projection Example: IDH
Projection Example: voting in US Senate

Point Placement by Phylogenetic Tree Construction Algorithms (N-J Trees)

\[ d_{AB} + d_{CD} \leq \max (d_{AC} + d_{BD}, d_{AD} + d_{BC}) \]

\[ d_{AB} = a + b \quad d_{AC} = a + e + c \quad d_{AD} = a + e + d \]
\[ d_{CD} = c + d \quad d_{BD} = b + e + d \quad d_{BC} = b + e + a \]
• Alternate view (N-J Tree)

Exploration

• Finding Relationships
• Building a Surface

RSS News Flash

Bird and Flu

Palestinian
Application 1: Visual Text Mapping

- Approach 1: Relationship Based (Metadata)

- Approach 2: Content based

Relationships: Topic Bursts and co-word

(Mane and Borner) 2004
Relationships: Citation and Co-citation

(Borner) (2003)

Content-based Text Mapping

- Approach 1: Pre-clustering & View
- Approach 2: Dimension reduction (Projections)

Content-based

(Skupin) (2002) (abstracts) SOM

Content-based

(Dimensional Reduction) News flash IN-SPHERE (PNL)
Content-based

(Surface View) IN-SPiRE

Mapeamento para o plano permitindo a exploração. Ex: Patents surgery, drugs, molecular bio

SOM based

• Self-Organization Maps (SOMs) cartográficos (ex. Skurpin 2002)

Exemplos de Mapas
Exemplos de Mapas

• Detailing topics
Time Series – Streamflow in Hidroelectrics

Text from attributes

- Cattle performance data
  - Translated to text from categorical information, e.g.,
    - Ranges of weight to words such as:
      - \{weight\_below\_fifty\_percent;
        weight\_between\_fifty\_seventy\_five; etc..\}
  - 9135 individuals

Cattle performance data

Colored by word 'top'
Images?

Cattle performance data

Colored by female

Colored by farm

Pipeline

Image Data Set → Feature Acquisition → Feature Selection

Interaction

Classification → Visualization → Similarity Calculation
Comparison of Distance Metrics

- Euclidean
- City Block
- Cosine

512 MRI medical images
12 classes

Comparison of Feature Space (1)

- 16 Gabor Filters
- Fourier, Mean and Deviation
- 72 co-occurrence matrices
- All combined

512 MRI medical images
12 classes
Comparison of Feature Space (1)

- 16 Gabor Filters
- Fourier, Mean and Deviation
- 72 co-occurrence matrices
- All combined

512 MRI medical images
12 classes

Comparison of Feature Space (2)

- All combined
- 1024 Wavelet Features

1000 X-Ray images from ImageCLEF
116 classes

Comparison of Feature Space (2)

- All combined
- 1024 Wavelet Features

1000 X-Ray images from ImageCLEF
116 classes

Detailed Inspection
Detailed Inspection

ImageCLEF Training Data Set (1)

9000 X-Ray images
116 classes

ImageCLEF Training Data Set (2)

Class 108
Class 111

Further Examples on Text

• RSS Patent Data, recovered from the Web http://www.freepatentsonline.com/
• Case 1:
  • 170 files
  • Graphics processing, printer, database, document, ai
Patents – case 2

- http://www.freepatentsonline.com/
- 172 files
- surgery (2), drugs(2), molecular biology
Patents surgery, drugs, molecular bio
stopwords selection

Patents surgery, drugs, molecular bio
topics

Patents surgery, drugs, molecular bio

Patents surgery, drugs, molecular bio
**Projection Explorer (PEx)**

http://infoserver.lcad.icmc.usp.br/

---

**Collaborators**

Alneu de Andrade Lopes – Mineração de textos
alneu@

Haim Levkowitz – Visualization
haim@cs.uml.edu

João E. S. Batista Neto – Imaging
jbatista@

---

**Visualization Group**

Maria Cristina F. Oliveira
cristina@

Fernando Vieira
Paulovich

Rosane Minghim

Luis Gustavo
Nonato

---

**Doutorandos**

Danilo Medeiros Eler
Aretha Barbosa
Kátia Felizardo

---

**Mestrandos**

Jorge Poco Medina
Christian
Tácito Neves
Renato Oliveira
Gabrial Andery
Other Partnerships

Sérgio Furuie (Poli – USP), Brazil
Lars Linsen (Jacobs University Bremen), Germany
Charl Botha (TU Delft); Anton Heijs (Treparel Inc.), The Netherlands

Link

- infoserver.lcad.icmc.usp.br (Pex, Pex-WEB, Pex-Temporal, Pex-Image).

Referências