Visualisation d’information

Jean-Yves Antoine

http://www.info.univ-tours.fr/~antoine/
Visualisation d’information

Chapitre 1.3 – Techniques de visualisation avancées

... ou comment quitter les histogrammes et les camemberts Excel pour entrer dans la 5e dimension
PRINCIPES - Objectifs

1.3.1. Notions

1.3.1.1. Information et incertitude de données
1.3.1.2. Représentation des données multi-dimensionnelles
1.3.1.3. Représentations avancées pour les données hiérarchiques

1.3.2. Pratiques

1.3.2.1. Représentation de l'incertitude
UNCERTAINTY REPRESENTATION

Uncertainty: multiple sources and causes

<table>
<thead>
<tr>
<th>Source: kind of uncertainty</th>
<th>Possible cause</th>
</tr>
</thead>
<tbody>
<tr>
<td>Limited accuracy</td>
<td>Limitations of measuring instruments or computational processes</td>
</tr>
<tr>
<td>Missing data</td>
<td>Non-representative sample size</td>
</tr>
<tr>
<td>Inconsistency</td>
<td>Conflict between data from several sources</td>
</tr>
<tr>
<td>Noise</td>
<td>Random noise</td>
</tr>
<tr>
<td>Bias</td>
<td>Personal idiosyncrasy, systematic measure error</td>
</tr>
<tr>
<td>Ambiguity</td>
<td>Linguistic data</td>
</tr>
</tbody>
</table>

Managing uncertainty

- Direct influence on data understanding and decision making
- A complete correction / reduction of uncertainty is impossible
- Modeling (statistics, probability, fuzzy set…) and reasoning with uncertainty
- Few works on the visualization of uncertainty – needs for:
  - Seamless integration of information and uncertainty (restricted cognitive load)
  - Flexible representation of uncertainty: handling easily multiples uncertainty models
UNCERTAINTY MODELS

Probability

- Likehood of an event to occur / a data to be observed

**Example** - Bayesian probability $P(\text{StateA} \mid \text{ObservationB})$
- Descriptive probability $P(x = A)$
  - Paramétric model: normal distribution (mean, standard deviation)

Provability

- Certitude Facteurs: $-1 =$ surely false, $1 =$ surely true
- Belief function: Dempster-Shafer Theory
  **Example**: BF(A), BF(B), BF(C), BF({A,C}), BF({A,B}), etc…

Possibility

- Alternative matches for an observation
  **Example**: range of errors on an measure

Membership

- Fuzzy sets: degree of membership to a specific set
UNCERTAINTY REPRESENTATION

- Visualization techniques better adapted to specific uncertainty models
- Two main approaches: additional information vs. integrated information

**Additional information: extrinsic representation**

- Additional object to represent a specific parameter/data
- **Error bar** – represent standard deviation or a range of values (confidence interval).
  - works with point chart or bar charts

![Error bar example](image1)

**Turner Syndrome 1999-2004**

![Case distribution by mother's age](image2)
Additionnal information: extrinsic representation

- **Boxplot, or box and whisker plot** (*boite à moustache*) – to represent visually a distribution (quartile or decile, extreme values)

  ![Boxplot Example](image1)

  - 1st quartile
  - median
  - 3rd quartile

- **Summary plot** – explicit representation of the distribution (thickness of the line)

  ![Summary Plot Example](image2)

[Potter and al. 2011]
UNCERTAINTY REPRESENTATION

**Integrated information: intrinsic representation**

One visual feature dedicated to uncertainty representation

**Blur**

- Easy interpretation of the uncertain nature of the data
- Impossibility to estimate the degree of uncertainty

[Hue, luminance, saturation, texture](Grigoryan, Rheingans 2004)
N-DIMENSIONAL DATA

Scatter Plot Matrix (SPLOM)

Principle – use small multiples of scatter plots showing a set of pair-wise relations among variables

[Heer and al. 2010]

![Scatter Plot Matrix Example](image)

- **horsepower**
- **weight**
- **acceleration**
- **displacement**

Legend:
- United States
- European Union
- Japan
DONNEES A N-DIMENSIONS

Murs perspectifs

documents arranged on a perspective wall
DONNEES A N-DIMENSIONS

Parallel coordinates

Principle – Variables are plotted on parallel axes. A poly-line = a row of the DB
Interactive visualization – You can change the order of the variables
DONNEES HIERARCHIQUES

Deux propriétés visuelles pour guider la visualisation

• Connectivité
• Clôture

Connectivité : arbres hiérarchiques

• Représentation la plus courante
• Des limitations toutefois [Card et al. 1999]
  - peu utilisable à mesure que l’arbre croît
  - arbre non équilibré occupation non optimale de l’espace
  - valeurs associées au nœuds : attributs quantitatifs non intégrables

Clôture : treemaps

• Technique de remplissage d’espace
• Treemap : principes généraux [Schneidermann 1992]
  - relations hiérarchiques représentées par une inclusion d’espaces
  - aire des rectangles proportionnelles aux attributs

• Intérêt
  - attributs quantitatifs intégrés en propre
  - espace consommé indépendant de la profondeur de la hiérarchie
Approche surfacique : *Treemaps*

- **Partitionnement d’espace** – chaque rectangle représente un nœud de taille proportionnelle à sa valeur d’attribut. Sous-nœuds : partition du rectangle père.

- Plusieurs algorithmes de découpage : propriétés différentes
  - Ratio – rectangles plus ou moins proche du carré (le plus lisible)
  - Ordre – ordre des sous-rectangles proche de celui des données initiales
  - Stabilité – stabilité du découpage quand les données changent.

<table>
<thead>
<tr>
<th>Algorithme</th>
<th>Ratio</th>
<th>Ordre</th>
<th>Stabilité</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arbre binaire</td>
<td>++</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>Mixed treemap</td>
<td>-</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>Ordered</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Slice &amp; Dice</td>
<td>++</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>Squarified</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Strip</td>
<td>+</td>
<td>++</td>
<td>+</td>
</tr>
</tbody>
</table>

- **Critères perceptifs**
  - Attention au ratio largeur / longueur (se rapprocher du carré)
  - Stabilité – stabilité du découpage quand les données changent.
DONNEES HIERARCHIQUES

Approche surfacique : Treemaps

- Recommandations ergonomiques
  - Ratio longueur / largeur : éviter les zones trop minces
  - Couleurs – Pas plus de deux couleurs, la perception devient vite trop dure sinon
  - Adjacences – Ne pas utiliser les treemaps si adjacences non interprétables

[Schneiderman 1992]

SmartMoney (1999)
DONNEES HIERARCHIQUES

Ellimaps  
[Otjacques et al. 2008]

- Même principe que les *squarified treemaps* mais avec des ellipses
- Gestalt : meilleur respect du principe contour et de continuité
DONNEES HIERARCHIQUES

Ellimaps : algorithme de construction (version initiale)

Adaptation aux ellipses des *squarified treemaps*
- Découpage du rectangle inscrit dans l’ellipse de ratio $\phi = a/b$
- Aire des ellipses proportionnelle à la valeur de l’attribut
- Découpage vertical ou horizontal suivant le plus grand côté du rectangle)

Evaluation

- Trois tâches de recherche d’information
- Temps et erreurs
DONNEES HIERARCHIQUES

Adjacency diagrams

- Partial consideration of the closure principle
- Represent both hierarchy and adjacency

[Heer and al. 2010]
DONNEES MULTI-DIMENSIONNELLES

Entrepôts de données et cubes de données

Cubes de données: intérêt

• Cubes de données: ensemble des agrégations réalisables sur les données
• Pré-calcu réduisant le requêtage OLAP

Cubes de données: limitations

• Exploration fastidieuse des cubes de données du fait de leurs très grandes dimensions

   Exemple – 1 000 000 tuples $\Rightarrow > 100 000 000$ cubes

• A l’heure actuelle, exploration avec des techniques inadaptées (SQL, représentation graphiques simples) ou des requêtes OLAP pré-programmées : aucune interactivité

Solution : visualisation de données

• Techniques de visualisation avancées permettant une réelle exploration interactive
DONNEES MULTI-DIMENSIONNELLES

Semi-treillis des valeurs des cuboïdes

- Graphe acyclique orienté permettant de représenter les attributs
- Taille des nœuds proportionnelle à la valeur de la variable d’agrégation

<table>
<thead>
<tr>
<th>Nom</th>
<th>Equipe</th>
<th>Poste</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALL</td>
<td>ALL</td>
<td>ALL</td>
<td>3</td>
</tr>
<tr>
<td>Parker</td>
<td>ALL</td>
<td>ALL</td>
<td>1</td>
</tr>
<tr>
<td>Diaw</td>
<td>ALL</td>
<td>ALL</td>
<td>1</td>
</tr>
<tr>
<td>Batum</td>
<td>ALL</td>
<td>ALL</td>
<td>1</td>
</tr>
<tr>
<td>ALL</td>
<td>Spurs</td>
<td>ALL</td>
<td>2</td>
</tr>
<tr>
<td>ALL</td>
<td>Blazers</td>
<td>ALL</td>
<td>1</td>
</tr>
<tr>
<td>ALL</td>
<td>ALL</td>
<td>Meneur</td>
<td>1</td>
</tr>
<tr>
<td>ALL</td>
<td>ALL</td>
<td>Ailier</td>
<td>2</td>
</tr>
<tr>
<td>Parker</td>
<td>Spurs</td>
<td>ALL</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td>.....</td>
<td>.....</td>
<td>.....</td>
</tr>
<tr>
<td>ALL</td>
<td>Blazers</td>
<td>Ailier</td>
<td>1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>.....</td>
<td>.....</td>
</tr>
<tr>
<td>Batum</td>
<td>Blazers</td>
<td>Ailier</td>
<td>1</td>
</tr>
</tbody>
</table>

- **Limitation** – Visualisation difficile au-delà de 1000 nœuds
Projections multiples

- Projection 2D, chaque dimension d’affichage correspondant à des agrégations
- Plusieurs formes de représentation (courbes, histogramme, camemberts)
- Plusieurs projections affichables simultanément.

Limitation – Efficace, mais le nombre de projection n’est pas infini: exploration orientée par les possibilités d’affichage, nécessite une étude préalable fine des besoins
DONNEES MULTI-DIMENSIONNELLES

**Visualisation orientée pixel**

- VisDB [Keim, Kriegel, 2000]
- d3Viz [Novelli, Auber, 2008]

- Visualisation très dense de l’information : un pixel = une donnée
- Jeu sur la couleur et la position des pixels pour interpréter l’image visuelle présentée
- Fouille visuelle sans perte de contexte: zooms de type glyphs fish-eye
DONNEES MULTI-DIMENSIONNELLES

Visualisation orientée pixel

- Pipeline de visualisation

Données brutes → Cubes de données → Forme visuelle → Image finale

**Transformation des données** → **Mapping visuel 2D** → **Création vue**

**Interaction utilisateur**

Visualisation orientée pixel : mapping visuel

- Principe
  - Associer un pixel à une donnée sur l'image, en respectant leur proximité originelle
  - Plusieurs courbes de remplissage d’espace envisageables

Exemple de mapping visuel:

1. Parker, *
2. Diaw, *
3. Batum, *
4. *, Spurs, *

Diagramme de mapping visuel:

- **Piste 1**: Parker, *, Diaw, *, Batum, *, Spurs, *
- **Piste 2**: Parker, *, Diaw, *, Batum, *, Spurs, *
- **Piste 3**: Parker, *, Diaw,*, Batum, *, Spurs, *
- **Piste 4**: Parker, *, Diaw, *, Batum, *, Spurs, *
- **Piste 5**: Parker, *, Diaw, *, Batum, *, Spurs, *
DONNEES MULTI-DIMENSIONNELLES

Visualisation orientée pixel : mapping visuel

• Courbes fractales

Peano-Hilbert

Morton (ordre Z)

Niveaux du semi-treillis non respectés : exploration difficile

• Courbes en scanline et spirale

• Exploration
  - Observation : rotation, zoom de type fisheye etc…
  - Action : recalcul de la vue suivant de nouvelles dimensions
DONNEES MULTI-DIMENSIONNELLES

Visualisation orientée pixel

- Pipeline de visualisation

Données brutes → Cubes de données → Forme visuelle → Image finale

Transformation des données → Mapping visuel 2D → Création vue

Interaction utilisateur

Visualisation orientée pixel : mapping visuel et requêtage

- Associer un pixel à une donnée
- **Flexibilité d’interrogation** : pas de filtrage, mais degré de pertinence des données
- **Visualisation** : couleur et proximité des données en fonction de leur pertinence
- Plusieurs courbes de remplissage d’espace (spirale...)
DONNEES MULTI-DIMENSIONNELLES

Visualisation orientée pixel : visualisation multi-dimensionnelle

- Une fenêtre pour la visualisation des résultats de la requête (facteur de pertinence)
- Autres fenêtres pour projection sur différentes dimensions

Visualisation 2D – association de 2 dimensions aux axes + remplissage par courbe fractale
DONNEES MULTI-DIMENSIONNELLES

Visualisation orientée pixel : interaction

1. L’utilisateur définit initialement ses requêtes
2. Possibilité de les modifier interactivement dans l’interface au vu des résultats

Limitations – Filtrage, projection gérées, mais pas de zoom sur les données brutes
Main scientific challenges in current researches

• Real-Time visualization of huge datasets
• Virtual or augmented reality for enhanced interaction (immersive displays)
• Multimodal visualization: interaction in a seamless way (reduced cognitive load)
• Collaborative visualization
• HCI: design and ergonomic guidelines for information visualization

☞ « Highest priority task » [Zudilova and al. 2009, chapter 1]
BIBLIOGRAPHIE

Ressources en ligne : concepts

Semiologie graphique  www.sciences-po.fr/cartographie/semio/
Treemap  www.sc.umd.edu/hcil/treemap/

Ressources en ligne : outils

- 30 freeware pour la visualisation d'information (post en anglais et reprise en français)
  - www.computerworld.com/s/article/9214755/Chart_and_image_gallery_30_free_tools_for_data_visualization_and_analysis

- Toolkits opérationnels pour découvrir la visualisation interactive
  - InfoVis  http://philogb.github.com/jit/ (freeware)
  - VTK (Kitware Visualization Toolkit)  http://www.vtk.org/ (visualisation 3D)
  - Prefuse Information Vizualisation Toolkit  http://prefuse.org/ (freeware: graphs, trees)
Ouvrages de référence : concepts et techniques


Ouvrages de référence : recherche & veille technologique


Travaux cités


BIBLIOGRAPHIE

Travaux cités (2)


Travaux cités en ligne