


情報可視化

高間 康史
首都大学東京 システムデザイン学部

1

自己紹介

- 高間 康史 (たかま やすふみ)
- 首都大学東京 教授
- 1999 東京大学工学系研究科 博士課程修了, 博士 (工学)
- 1999-2002 東京工業大学 大学院総合理工学研究科 助手
- 2002-2005 都立科学技術大学 助教授
- 2005-2014 首都大学東京 システムデザイン学部 准教授
- 2014- 現職



2

研究テーマ

- 情報可視化, 知的インタフェース
- Webインテリジェンス, AI
- データマイニング, レコメンデーション

(著書: 監修, 翻訳含む)



3

情報環境の進歩がもたらすもの


- (従来) 向上した計算機能力の活用
- より高速に, 大規模に, 小型に, . . .
- 人間が情報活用のボトルネックに
- 解決策: 人間と計算機の「役割分担」の見直し
- 人工知能: 知的処理の代行
- 可視化: ボトルネックの解消



4

可視化 (Visualization) とは

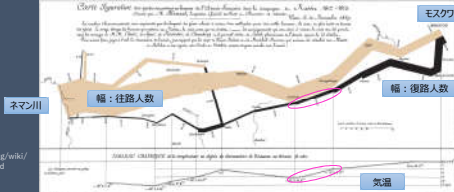
- 見えないものを見るように
- 見えにくいものを見やすく
- 人間の知覚能力の活用
- 計算機能力の向上による対象範囲の拡大
- 処理対象データの大規模化
- グラフィクス能力の向上
- インタラクティブ性の向上



5

可視化の例1: インフォグラフィック

- 1869 by Charles Joseph Minard (フランス)
- ナポレオンのモスクワ遠征 (1812-13) の悲惨さを可視化
 - 軍団の進行方向, 通過場所, 兵数減少 (42万→1万), 気温

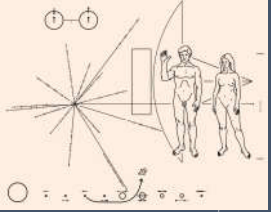


https://en.wikipedia.org/wiki/Charles_Joseph_Minard

6

可視化の例2：インフォグラフィック

- 1972 バイオニア探査機に搭載された金属板
- 地球外生命体に理解してもらおう事を意図



- 中性水素の超微細遷移 (長さ, 時間の単位)
- 人間の身長 (探査機との相対的サイズ)
- 太陽の相対的位置
- 探査機の太陽系における経路

Courtesy NASA/JPL-Caltech

可視化とは…

- データをコンピュータグラフィックスで描画したものである
 - 絵にただけでは「可視化」ではない
 - テキストだって「可視化」と呼べる場合もある
- アート・デザインのセンスが必須である
 - 必ずしも必要ではない
 - エクセルのグラフも立派な「可視化」
 - 「目新しさ」と「慣れ親しみ」

8

可視化の技術的展開

- データ解析 + コンピュータグラフィックス (CG)
- 科学・工学分野での利用
 - 気象学, 流体力学, 化学反応, 物理現象
 - 振動解析, 自動車設計, CAD
 - 医療分野: 器官の視覚化, CT画像処理
- より広い分野への展開
 - ビジネス分野: データマイニング, 統計解析, アンケート解析
 - 情報検索



計算機の機能向上により適用範囲拡大


9

可視化対象の拡大

- Scientific Visualization
 - 流体力学
 - 地理データ (Geographic data)
- Information Visualization (情報可視化, InfoVis)
 - ビジネス応用 (データマイニング)
 - 情報検索システム

両者の違いは?

対象の拡大



10

科学的可視化 vs. 情報可視化

- Scientific Visualization
 - 本質的な構造 (2,3次元) を持つデータが対象
 - データが本来持つ構造をわかりやすく見せることが目的

How to visualize?
- Information Visualization
 - 抽象的データ
 - 多次元データ
 - 描画だけでなく, 構造を与えることも必要

How to visualize?

可視化技術自体は共通性あり

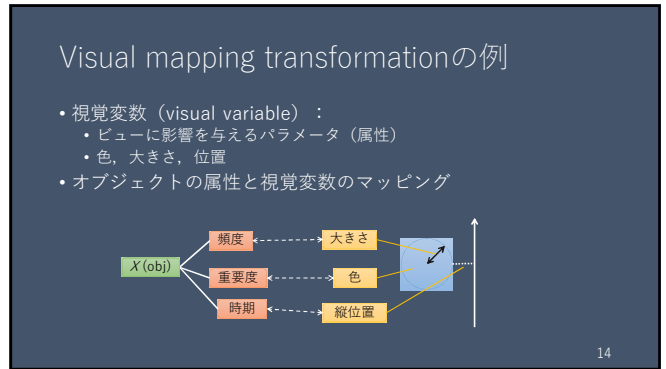
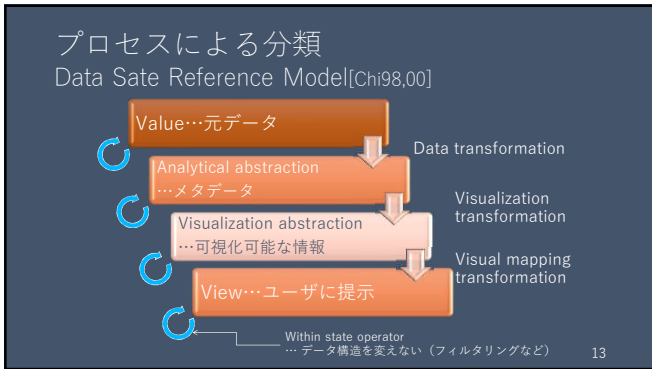
構造に関する自由度が高い

11

可視化問題の定式化

- 対象: $D = \{o_1, \dots, o_n\}$
 - データオブジェクト: $o = \{a_1, \dots, a_k\}$
 - オブジェクト間関係: $R = \{r_{ij}\}$ ← 時間変化を伴う場合も
- データオブジェクトの可視化
 - 属性情報の提示… cf. チャーノフの顔グラフ
- 全体構造 (Overview) の可視化
 - オブジェクト間の関係 (R) … cf. ネットワーク / オブジェクトの分布 (D) … cf. 散布図
 - 属性間の関係 (相関) … cf. 気温と兵士数の関係

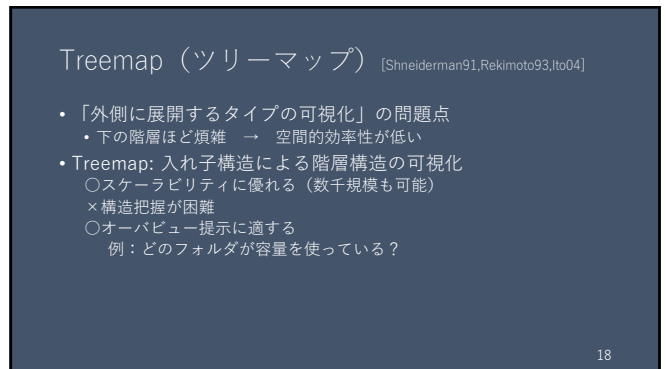
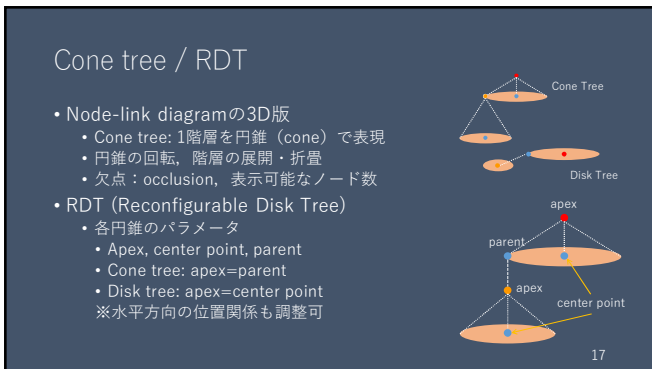
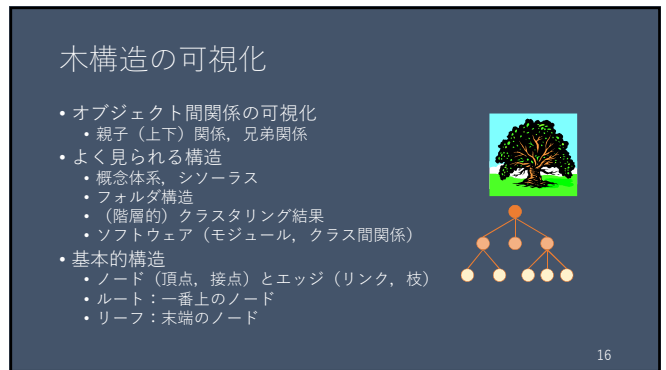
12



Webサイト可視化の例

データ・変換	例1	例2	例3	例4
Value	Webページ集合			
Data Transformation	リンク解析		キーワード解析	
Analytical Abstraction	ネットワーク構造		キーワード集合	
Visualization Transformation	幅優先探索		ソート	
Visualization Abstraction	木構造	木構造	ネットワーク構造	キーワードリスト
View (Visual Mapping Transformation)	ツリーマップ	コーンツリー	パネモデル	タグクラウド

15



Treemapの基本 ~Slice & Dice~

19

ネットワーク構造の可視化

- ネットワーク構造の例
 - Web: ハイパーリンク構造
 - ソフトウェア (モジュール, クラス間関係)
 - ソーシャルネットワーク (人間関係)
 - 遺伝子ネットワーク (遺伝子間の相互作用)
 - 神経回路網
 - 通信網
- 可視化対象として近年有名
 - 複雑ネットワーク: 現実世界の多様なネットワークの可視化・分析
 - ネットワーク可視化ツール: Cytoscape

20

ネットワーク可視化の課題

- 可視化対象データの形式
 - D : データ集合, A : 属性集合
 - 距離・類似度行列 ($|D| \times |D|$)
 - データ・属性行列 ($|D| \times |A|$)
 - 各データ ($|A|$ 次元ベクトル) からデータ間類似度 (距離) 計算
 - ユークリッド距離, コサイン類似度
- 本質的に多次元構造 ($\sim |D|$ 次元)
 - 2,3次元空間に写像する必要
- 一般的な可視化手法
 - ノードとエッジによる表現

21

ネットワーク描画の条件

- オブジェクト間の関係が正しく表現されていること
 - 類似データのグループ発見 (コミュニティ抽出)
- リンクが交差して描画されないこと
 - 接続関係の見やすさ
- ネットワーク構造が把握しやすいこと
 - 特徴的構造の発見: スター構造, サイクル
- オブジェクト同士が密集しすぎないこと
 - 大規模ネットワークの可視化, テキストラベルの配置

22

力学的モデルの利用

- 動的なノード配置方法
 - ノード位置を逐次計算
- 大規模ネットワークに適用可能
- インタラクティブな可視化
- バネモデル: 単純な力学的モデル
 - リンク=バネ: 自然長=ノード間の関連度・描画距離
 - 力学的エネルギー小=ひずみの少ない配置
 - $EE = \sum_{e_{ij} \in E} \frac{1}{2} k_{ij} (d_{ij} - l_{ij})^2$
- エネルギーが減少する方向にノードを逐次移動

23

可視化の対象タスク

- プレゼンテーション
 - インフォグラフィクス
 - データの提示
 - 支援目的: Storytelling (データ閲覧・解釈の誘導)
- 探索的データ分析
 - EDA (Exploratory Data Analysis)
 - データから知見・仮説を得る
 - 支援目的: 多様な観点からの分析・探索空間の拡大
- モニタリング
 - 日々発生するデータの効率的な確認
 - 対象データ: オンラインニュース, 2チャンネル, Twitter, etc.

24

EDAと可視化

- ビジュアルアナリティクス(Visual analytics)[Thomas05]
ビジュアルデータマイニング [Mozzafari08, 白山06]
 - 統計モデル, 可視化, 人間の直感を組み合わせることで大規模・複雑なデータセットから洞察を獲得
 - 人間をデータ探索プロセスに組み込む→知覚能力の活用
 - Viewを介して人間とデータがインタラクション
- 可視化 = 探索的プロセス [Chen09,10]
 - 満足するviewを見つける作業
 - 膨大な探索空間をインタラクティブに

25

Visual Analyticsのプロセス

分析者が行うべき作業

1. ビューから興味ある部分, 特徴的部分に**気づく**
 - 特異値, 変動点など
2. 気づいた点の背後にある理由を考察
 - **解釈, センスメイキング (sense making)**
3. 気づきを積み上げ, **知見**を得る
 - **洞察 (insight)**, 包括的理解, アクションへ

26

EDA支援のアプローチ

- インタラクションの導入
 - 効率から試行錯誤へ
 - 操作の結果を即座に反映
 - 「後戻り」を可能に→失敗の許容
- 比較作業の支援
- 協調作業の支援
 - 多様な観点・知識の導入→探索空間の拡大
 - 三人寄れば文殊の知恵

27

可視化とインタラクション

- Visual Analytics Mantra by Keim
 - Analyze First – Show the Important
 - Zoom, Filter and Analyze Further
 - Details on Demand.

(まず分析, 重要な部分を確認後, ズーム・フィルタリングして分析, 必要に応じて詳細を確認)

- Zoom, Filter, Details: ビューへのインタラクション

28

インタラクションの分類

～ユーザ要求の観点から [Y07]

- Select (選択): 関心ある対象にマーク … 他の操作に先立って実行
- Explore (探索): 現在と異なるものを見る
 - アイテムの追加/削除, 表示範囲の変更 (panning)
- Reconfigure (再構成): 異なる配置で見る
 - データ・属性の順序変更 (ソート), アイテム位置の変更
- Encode (符号化): 異なる表現で見る
 - Viewの変更, 色・サイズ・形状割当の変更
- Abstract/Elaborate (抽象化・精緻化): 詳細を見る
 - Details-on-demand, drill-down, zooming
- Filter (フィルタ): 条件を付けて見る
 - Dynamic queryなど
- Connect (接続): 関連アイテムを見る
 - 関連アイテムの検索・提示, 関係の強調

29

情報可視化システム設計の「原則」

- Visual-information-seeking mantra [Shneiderman96]
 - Overview first, zoom and filter, then details on demand (オーバービューを見てズーム/フィルタリングした後, 必要に応じて詳細を確認)
 - Overview first, zoom and filter, then details on demand …
- Focus+Context
 - Focus … 詳細情報 (detail) へのアクセス
 - Context … 全体像 (Overview), 情報間の関連

30

Overviewを見せることの意義

- Overviewそのものに意味
 - 全体傾向の把握：
 - データの存在する範囲は？ / 典型的なデータは？
 - / グループは(いくつ)あるか？
 - データ分布の把握：
 - 正規分布？ / 一様分布？ / 属性間に相関関係は？
 - 統計的分析を適用する前に重要
- 補足的情報：オブジェクト絞り込みの手がかり
 - 興味あるオブジェクトの発見 → 詳細データの確認
 - 外れ値の発見

31

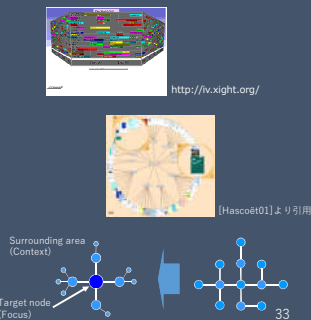
OverviewとDetailの両立

- 同一画面内で両立
 - Distortion: 配置を「歪める」ことで多数オブジェクトを描画
 - 中心オブジェクトは詳細に、周辺オブジェクトは大まかに
- 別々の画面でそれぞれ提示
 - 協調的マルチビュー (Cooperative Multiple View)

32

Distortionの諸技術

- Perspective wall [Mackinlay91]
 - 遠近感の利用
- Fisheye view [Furnas86]
 - 魚眼：中心部分を拡大
- Magnifying glasses [Hascoët01]
 - ルーペのメタファ
- Hyperbolic tree [Lamping95]
 - 双曲空間の配置を平面に写像



33

協調的マルチビュー (Coordinated Multiple View, CMV)

- 複数のビューを組み合わせた情報提示
 - ビュー同士はデータを介して紐付け
 - 連動・協調 (coordination)：ビュー間の関連づけ
- 代表的マルチビューシステム
 - Overview & Detail views
 - Focus & Context views
 - Different views: 違いを強調
 - Master / Slave views: 一方が他方をコントロール
 - Small-multiples: ビュー (アイコンなど) を多数並べる

Dual Views

34

代表的連動

- ブラッシング (Brushing)
 - オブジェクトを選択 → 対応オブジェクトを強調表示



- ナビゲショナル・スレービング
(Navigational slaving)

- ナビゲーション操作の連動



35

CMV導入のガイドライン [Baldonado00]

- 適切に利用 → 認知的負荷軽減
- 不適切 → 学習コスト, 余分なディスプレイサイズ
- 多様性の原則：属性, モデル, ユーザ特性, 情報の粒度, ジャンルなどに関して多様性がある場合に利用
- 複雑さの原則：異なるビューによって関連・相違が明らかになる場合に利用
- 分層の原則：データを扱いやすい単位にまとめたり, 次元間の相互作用に気づきやすくする様に利用
- 節約の原則：利用は最小限にとどめ, 学習コストやコンテキストスイッチに要するコストに配慮する

36

意味的インタラクション (Semantic Interaction)^[Ender12]

- Visual Analyticsにおけるインタラクション
 - Sense-making: EDAを通じて複雑なデータセットを理解
 - データ分析の専門家ではなく、ドメインの専門家を支援するためのインタラクション
- Sense-makingの2フェーズ
 - 採餌 (Foraging) : 興味ある/適切な情報をフィルタリングし収集
 - 統合 (Synthesis) : 収集情報に基づく仮説の構築・検証、情報間の関係理解
- Semantic interaction
 - 統計モデル (採餌能力) と分析者 (空間的統合能力) の連携

37

既存の採餌ツール

- 統計モデルでデータを処理→可視化
- インタラクション：モデルのパラメータを直接操作
 - キーワードの重み調整, 文書/キーワードの削除
 - 視覚的コントロール (e.g. スライドバー) によるパラメータの調整
- モデルやパラメータに対する知識が必要
 - パラメータ変更が結果に与える影響に関する知識

38

既存の統合ツール

- 思考プロセスを空間的レイアウトとして外在化
 - 空間内でデータに対する仮説・洞察を構築可能に
 - 視覚的メタファ
- インタラクションはメタファの外側
 - ツールバーを用いて統計モデルのパラメータを直接操作

39

従来インタラクションとの違い

[従来]
計算モデルに対する知識が必要

人間 → 視覚的メタファ → 生成 → 計算機
 フィードバック ← コントロールパネル → フィードバック

[Sem. Int.]
操作の意味を解釈、計算モデルへのフィードバック

人間 → 視覚的メタファ → 生成 → 計算機
 フィードバック ← フィードバック

40

Semantic Interactionの特徴

- 空間的可視化内で発生するインタラクション
 - オブジェクトの移動, 位置固定
 - 複数オブジェクトを重ねる (まとめる)
- 視覚的メタファの利用
 - 位置が近い=類似
 - 場所に意味
- インタラクションの意味を解釈し、モデルのパラメータに反映
 - 複数オブジェクトをまとめる → 両方に共通する属性の重要度↑
 - オブジェクト位置の変更 → 距離学習
- 背後にあるモデル・パラメータの複雑さからユーザを保護
 - モデル更新結果は視覚的に反映

41

おわりに

- 本日は話した内容
 - 可視化とは
 - 代表的可視化技術：ツリーマップ, Cone Tree, ばねモデル
 - 探索的データ分析支援：インタラクション, CMV, Semantic interaction
- その他の話題
 - その他の可視化技術：統計グラフ, 多次元データの可視化, 接続行列, タグクラウド, etc.
 - モニタリング支援, ストーリーテリング
 - 可視化システムの実際：いくつかの開発事例の紹介
 - 可視化システムの評価

42