

## データ分析の基本と TETDMの活用法

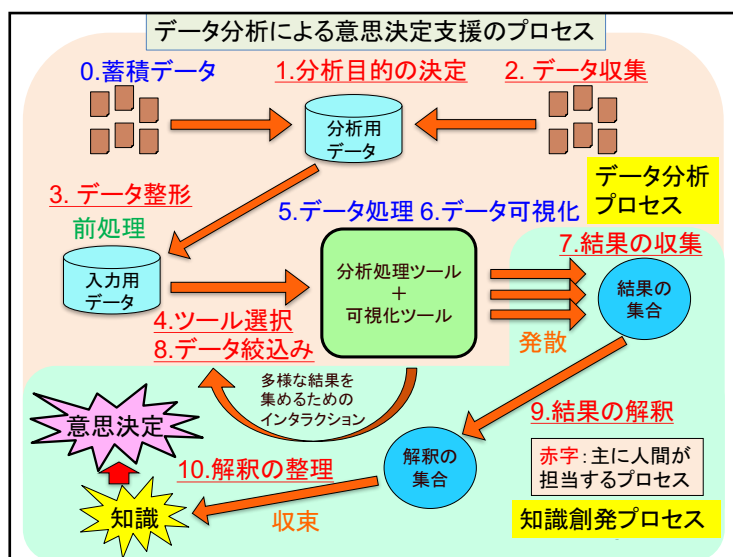
滋賀県立大学 工学部  
砂山渡

## データ分析とは

- ➡ 人間の活動における意思決定の内容および根拠を探る操作
- ➡ データ分析による意思決定プロセスはコンピュータによる客観的な情報の提示と人間による試行錯誤や判断との組合せ

人間とコンピュータの協働作業により実現される

2



## 有効なデータ分析のためには

- ➡ 正しいデータ分析のプロセスを理解して人間が分析に関わる
- ➡ データの意味がわかる, 直接データに携わっている現場の人が分析する
- ➡ ツールの詳細な動作原理よりもツールの出力の意味を理解する

4

5

[illegible]

**セグメント評価(セグメントにフォーカス)**

セグメント情報まとめ 単語抽出(文章評価用)  
レポート評価(結果+意見)  
コミュニケーション要約

**セグメント間, 文間, 単語間の関連度評価**

単語間関連度  
段落順序評価(トップダウン)  
テキスト分類(再帰的クラスターリング)  
テキスト間類似度(独自性)  
段落間木構造(類似度, トップダウン)  
段落間類似度表示  
段落間ネットワーク  
(類似度順, ばねモデル)

**文評価(文にフォーカス)**

文情報まとめ 文章要約(展望台)  
意見文抽出 主題関連文評価(光と影)  
類似文抽出 失礼単語抽出 単語冗長文抽出  
長文抽出 主語抽出 主題関連語評価(川下り)

**単語評価(単語にフォーカス)**

単語情報まとめ 単語抽出 TFIDF  
単語頻度リスト 文内キーワード抽出(Yahoo)  
専門用語抽出(FLR) 専門用語抽出(C-Value)  
専門用語ハイライト 文内キーワード抽出(頻度)

**形態素解析**

形態素解析(lgo) なし  
形態素解析サンプル(lgo)  
辞書(オンライン) 辞書再構築  
タイピング 英文音読  
プログラムチェック サンプル1  
データ送信テスト サンプル2  
連動データ確認(フォーカス)

**処理ツール一覧**

**Ver.4.0時点**  
\* 茶色は可視化ツール

**情報アクセス**

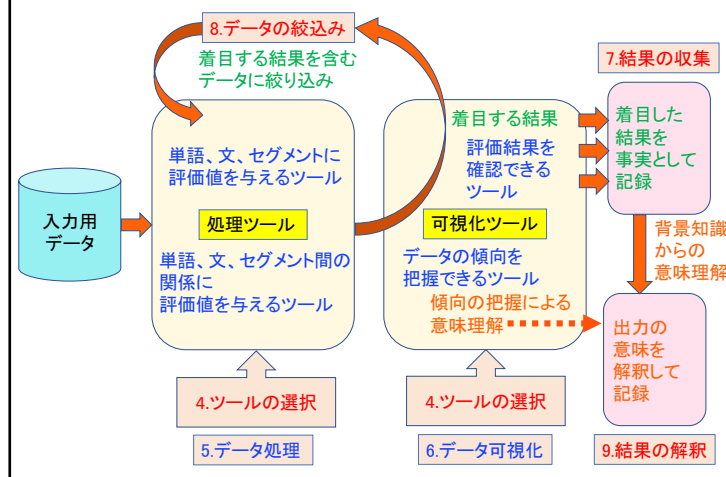
Webブラウザ  
URLアクセス  
2ちゃんアクセス

**文書整理**

テキストエディタ 単語置換  
セットセグメント テキスト2分割  
段落並び替え

- キーワード表示(展望台)
- キーワード集合比較
- 主題関連語表示(川形式)
- 単語間関係(関連度)
- 文内キーワード表示(タグクラウド)
- 専門用語表示
- 段落間ネットワーク(評価値順)
- テキスト分類表示(地図形式)
- 段落並び替え

## プロセス4.から9.の詳細



## 8. データの絞込み

- 着目する情報によるデータの絞込みが可能
  - 着目する単語を含むセグメント
  - 着目する文を含むセグメント
  - 指定の評価値や属性値をもつセグメント
- 着目している情報を明示するウィンドウを用意 (Ver.4.0から)

データ (セグメント) の「絞込み」条件と件数

表形式表示	で選択した	単語を含む	3	乙姫
表形式表示	で選択した	単語を含む	2	玉手箱
→表形式表示	で選択した	単語を含む	1	寿命

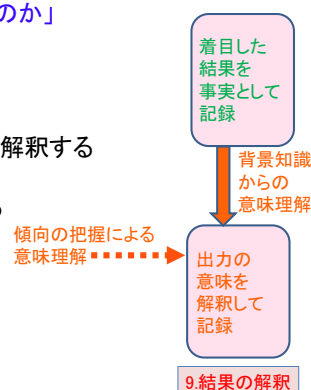
10

## 9. 結果の解釈

着目した結果が「なぜ出力されたのか」  
出力の意味(根拠)を捉える

1. 可視化ツールを利用して  
出力データの傾向\*2をもとに解釈する
2. データについての  
背景知識\*1をもとに解釈する

\*1データ分析の専門家が  
持ち合わせているとは限らない  
\*2傾向を把握するためにも  
データについての背景知識が  
必要となることもある



## 結果と解釈の登録インターフェース

「結果」と「解釈」をその場で入力できる

結果と解釈の登録 1 / 60

パネル 1: テキスト評価アプリケーション + テキスト表示 (HTML) (urashima.txt2)  
 パネル 2: テキスト評価 (分析結果まとめ) + テキスト表示 (HTML) (urashima.txt2)  
 パネル 3: テキストエディタ + テキスト表示 (urashima.txt2)

<結果> (必須)  
 <解釈> (必須)

[結果] 5 4 3 2 1 [登録] 終了

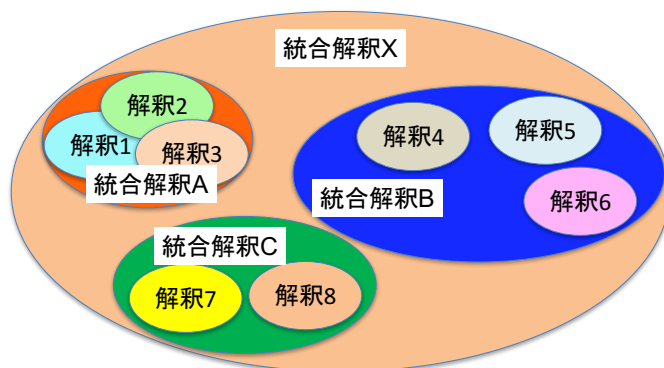
[解釈登録手順]

1. どのパネルの結果について登録するかを選択する (複数選択可能)
2. 「！」追加ボタンで表示される「！」を、気になる結果の上に置く (任意)
3. 気になる結果の内容 (事実) を具体的に、<結果>のところに書く
4. 結果の意味するところ、結果から言えることを、<解釈>のところに書く
5. 解釈の[重要度]を選択、5(特に重要)→4(重要)→3(まあ重要)→2(あまり重要でない)→1(重要でない)
6. [登録]ボタンを押して、解釈を登録する (結果、解釈、重要度、画面が保存されます)

\* <結果>と<解釈>は、自分以外の人にも伝わるように具体的に書くのがポイントです  
 \* 登録した解釈は、メニュー知識階級の「一覧」、「詳細」ボタンから確認、操作できます

## 10. 解釈の整理

集めた解釈を再帰的に1つに統合して  
データからわかることをまとめ、汎用的な知識につなげる(帰納推論)



## 知識創発インタフェース

集めた「解釈」を再帰的に統合できる

## まとめ

- ⑤ データ分析による意思決定の  
基本プロセスについて説明した
- ⑤ データ分析による意思決定プロセスを  
支援するツールとして  
TETDMの活用方法を説明した

15

## AIツール入門講座のご案内

- ⑤ TETDM(テキストマイニングツール)
  - 日時: 2019年3月11日(月) 10:00-17:00
  - 場所: 早稲田大学(西早稲田キャンパス52号館301号室)
  - 対象者: これからデータ分析を行ってみたい方, データ分析の必要がある方, データ分析の手順を理解したい方, データ分析ツールの導入を検討している方, 自由記述アンケートや業務日誌の簡単な分析を行いたい方, 大学事務職員の方, 広い意味でデータ分析に関わる学生の方
  - 参加費: 一般会員5000円, 学生会員2000円, 一般非会員8000円, 学生非会員3000円
  - 通常のツール講座  
の半額です!
  - \* 一般参加者1名につき, 同伴学生参加者1名無料

16