

TETDMの社会実践に向けて

滋賀県立大学 工学部
砂山渡

発表内容

- 近未来チャレンジTETDM
- 問題解決とデータ分析のプロセス
- TETDMを用いた社会実践



近未来チャレンジTETDM

- ⑤ 複数のテキストマイニング技術を柔軟に組み合わせて使える統合環境を構築
- ⑤ 電子テキストを扱う多くのユーザの創造的活動を支援するツールの提供

2010年度から5年以内に達成する課題として
人工知能学会全国大会「近未来チャレンジ」の
プロジェクトとして発足、**2015年度の大会で卒業認定!**



テトリーヌ by えむたこ

TETDMプロジェクトメンバー

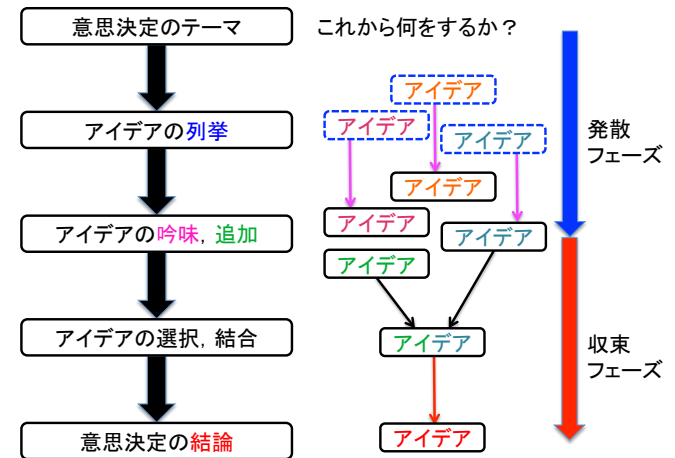
- ⑤ 砂山渡(滋賀県立大学 工学部)
- ⑤ 高間康史(首都大学東京 システムデザイン学部)
- ⑤ 西原陽子(立命館大学 情報理工学部)
- ⑤ 徳永秀和(香川高等専門学校)
- ⑤ 串間宗夫(宮崎大学 医学部附属病院医療情報部)
- ⑤ 阿部秀尚(文教大学 情報学部)
- ⑤ 梶並知記(神奈川工科大学 情報学部)
- ⑤ 松下光範(関西大学 総合情報学部)
- ⑤ Danushka Bollegala(ダヌシカ ボレガラ)(リバプール大学)
- ⑤ 佐賀亮介(大阪府立大学 大学院工学研究科)
- ⑤ 河原吉伸(大阪大学 産業科学研究所)
- ⑤ 川本佳代(広島市立大学 大学院情報科学研究科)
- ⑤ 西村和則(広島工業大学 工学部)

TETDM(Ver.2.00)公開中

⇒ <http://tetdm.jp/> よりダウンロード可能

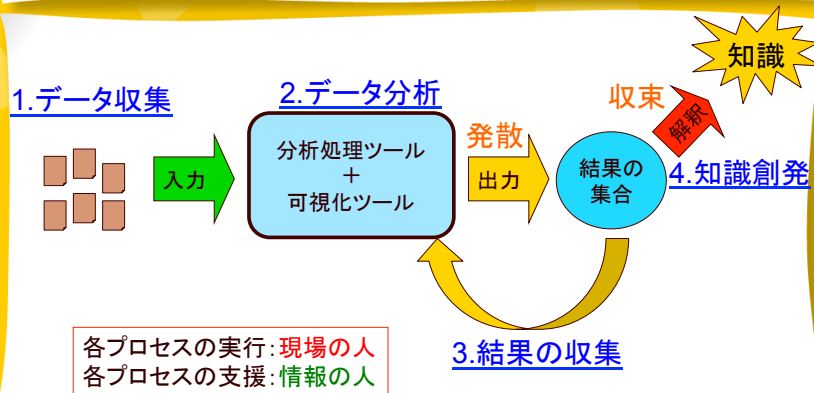
半年に1度のペースでバージョンアップ予定

問題解決(意思決定)のプロセス



TETDMで、このプロセス全体を支援する

データ分析のプロセス



1. データ収集

- ⇒ 問題解決(目的の達成)に必要なデータを収集
 - 数値データ
 - テキストデータ
 - 画像データ
 - 音声データ
- ⇒ 必要なデータが何か分からない場合もある
 - 目的に関係がありそうなデータを取りあえず集める

社会問題に関係しそうなデータ

- 何が,何を: RFIDタグ(IC), IoT ものの情報
- どこで: GPS 位置情報
- どのように: モーションキャプチャ 動きの情報
- 関連データの紐付け: ex) マイナンバー
- 変化: 時系列データ

入力データと分析の目的

- ⑤ **ことばによる出力の解釈が可能なデータ**
 - 主にテキストデータ
 - 出力を**ことば**で解釈して知識発見につなげる
- ⑥ **ことばによる出力の解釈が困難なデータ**
 - 主に動画像によるデータ
 - 出力の内容を**感覚**で理解して知識獲得につなげる
ex) プロ選手や専門家の動作
(考えるより先に感覚での動作が必要なもの)

2. データ分析

- ⑦ **データ分析ツールにデータを入力**
 - 分析処理ツール + 可視化ツール
(いわゆるデータマイニングツール)
- ⑧ **ツールは定められた処理は実行し結果を出力**
 - 有効な可能性が高いもの
 - 入力されたデータを整理したもの
 - 思考や判断の材料になるもの

すぐに使える知識を出力するわけではない!

分析ツールTETDM テキストマイニングのための統合環境 <http://tetdm.jp>

The screenshot shows the TETDM interface with the following data:

項目	値	スコア
主語一貫性 (文)	63%	(60/94) -17
主語一貫性 (単語)	61%	(71/115) -19
主語含有率	89%	(84/94) -10
不連続表出数(よう/など/いろいろ/という)	23	-23
長文の数(100字以上)	6	-30

3. 結果の収集

思考の発散フェーズ

- ⑤ 汎用性が高い/本質につながる知識を得るためには多くの結果が必要
- ⑤ 多くの結果を得るためには多くの**試行錯誤**が必要
- ⑤ 試行錯誤の要素
 - 処理ツール(処理方法)の選択
 - 可視化ツール(結果の表示方法)の選択
 - 情報の絞り込み(部分データの作成)方法の選択

シルエットの向こう側にいる有名人は誰か？
というクイズがあったときに
多くの「特徴的な」ヒントが得られるほど
正解にたどり着きやすい

15

結果の収集における着目点

- ずれからの知識発見

- ある基準データ(たとえば平均値)からのずれに意味を見出す
 - 平均気温
 - 平均睡眠時間
 - 確率の期待値
 - データの傾向(平均)を知ってそこからのずれを探る



16

4. 知識創発

思考の収束フェーズ

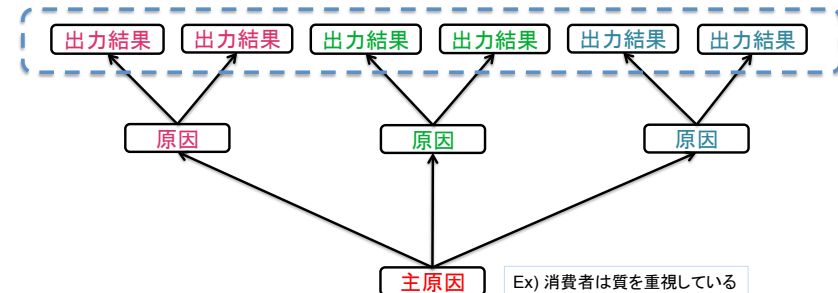
- 1) 収集した分析結果をまとめた**主要な知識(大局的な因果関係)**を探る
- 2) 発見した因果関係をもとに**シナリオ(行動の組み合わせ)**を構築する

17

主要な知識の探索

Ex) 10000円以上の商品が良く売れている

結果の全体像がわかると、何が起きているのかを理解しやすい



すべての結果を引き起こした**根本的な原因**がわかると
明確な方針を立てることが可能

18

知識創発の例



Ex) 消費者は質を重視している Ex) 10000円以上の商品が良く売れている

1) 大局的な因果関係を捉える

Ex) 質が良い商品に対しては、
値段にこだわらない購買行動が見られる

2) 大局的な因果関係にもとづいて 新たなシナリオ(行動の組合せ)を構築する

Ex) コストをかけてでも、質の良い製品を作ることが大事!

19

価値ある創発に繋げるには

⇒ 価値(効果)の高いシナリオ =
発見した因果関係の本質度 × (経験, 背景知識 + 発想力)

- 分析結果への、人間の知識の追加
「入力としては与えられなかった
データに関する知識」の追加
「データとして表せない経験、知識」の追加
- 分析結果の、人間の知識による選択と統合
「既存知識の組み合わせ発想力」

人間の知恵と積極的な思考が必要!

20

TETDMを用いた社会実践

⇒ 研究者と現場の人との連携

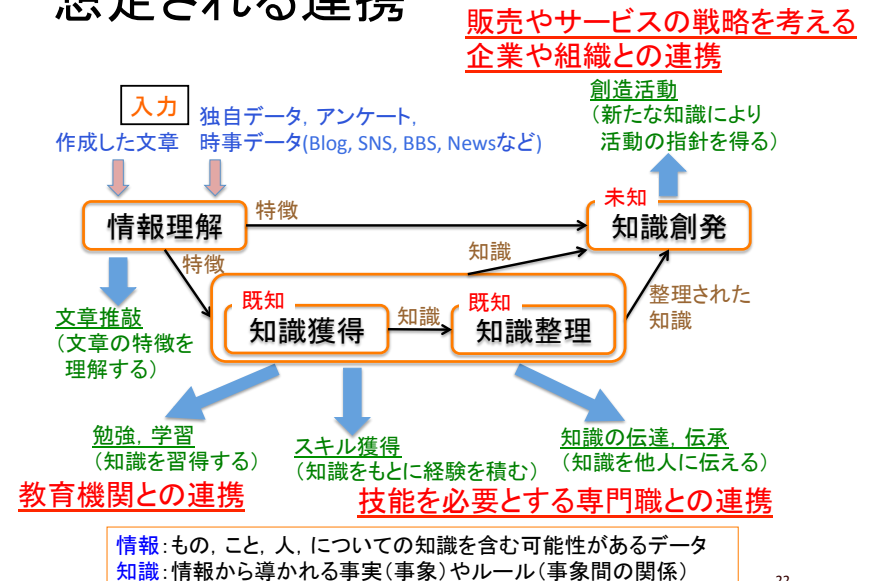
- 研究者の役割:

- シーズ(既存ツール)の提供
- ツール開発
- 分析支援

- 現場の人の役割:

- ニーズ(社会問題)の提示
- データの提供
- 分析の実践

想定される連携



22

社会連携に向けた取り組み

- ⑤ 人工知能学会 インタラクティブ情報アクセスと可視化マイニング研究会(SIG-AM)(2012年度から)
- ⑤ 人工知能学会: AIツールセミナー(2013,2015)
- ⑤ 広島市立大学: 市民向け公開講座(2014,2015)
- ⑤ 滋賀県立大学: 地域連携事業の推進(2016-)
- ⑤ オープンキャンパス, 技術交流会等への出展

教育機関との連携

- ⑤ 文章の作成と評価の支援
 - 論理的で可読性が高い作文の支援
 - 客観的指標の提示による文章評価支援
- ⑤ 小学生のプログラミング必修化
 - 結果がわかりやすいアプリケーションの構築支援
 - プログラム作成についての情報交換
 - 学習意欲の向上を含む学習環境の提供

まずは各大学内での活用実績を積み上げる

技術職の方との連携

- ⑤ 客観的な視点の確認
 - 共通認識の確認
 - 共通認識と自分の認識との位置関係やずれの確認
- ⑤ スキル獲得の支援
 - 暗黙知や形式知の抽出と伝達
 - 形式知データベースの作成

まずは医用, 看護関連での連携に着手する

企画, 開発職の方との連携

- ⑤ データ分析に基づくシナリオ生成支援
 - データ分析環境の提供
 - データ分析スキルの獲得支援
- ⑤ データ分析スキルの共有
 - データ分析プロセスの共有
 - データ分析結果の共有

埋もれているデータの活用や
社会問題の解決策を模索する企業や自治体との連携に着手

社会連携に向けた環境づくり

⇒ データ分析環境の整備

- ツール開発支援&アップロード環境整備
- 利用者支援: 操作アシストキャラの開発
- 興味ある活用が可能なツールの提供

⇒ データ分析スキルの獲得

- TETDMの利用履歴の共有
- TETDMの操作履歴の共有と再現

今後の展望

- ⇒ 具体的な社会実践に取り組むことで
世の中のデータの積極的活用と
データ分析が可能な人材の育成により
住みやすい社会の実現に貢献していきたい

