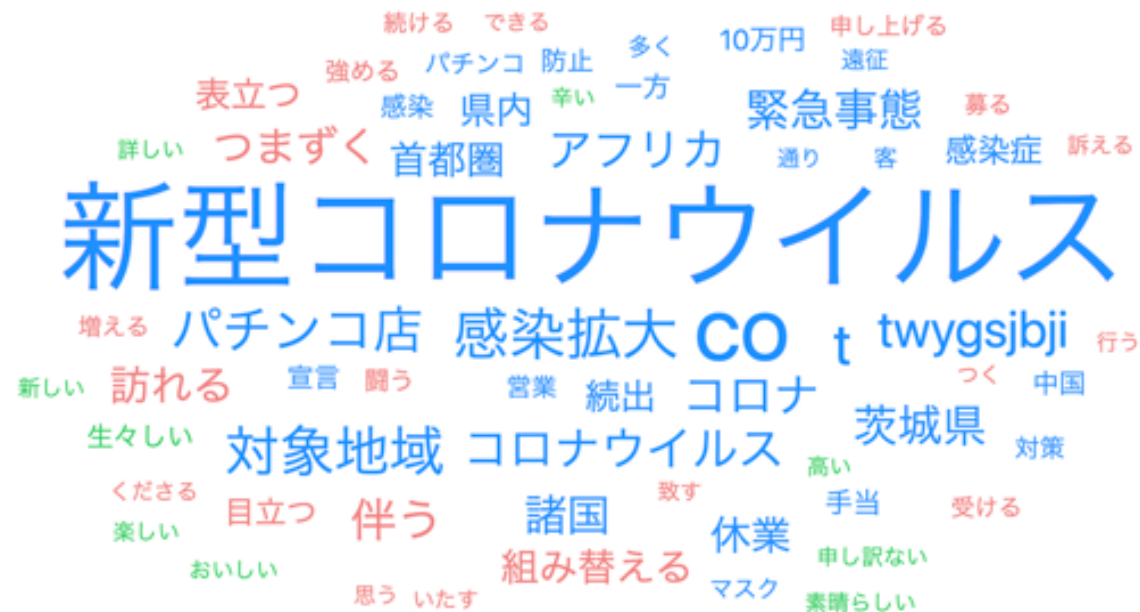


テキストマイニング超入門



講師紹介

- 氏名

- 岡崎 凌（おかざき りょう）



JDLA
Deep Learning for
GENERAL
2019 #3

- 学歴

- 大阪大学 基礎工学部卒業
- 大阪大学大学院 工学研究科修士課程修了

- 職歴

- 産業科学研究所
半導体の画像解析AIシステム導入
- 和から株式会社
セミナー・企業向け研修
- オンスク.jp オンライン授業
- 澤田経営道場 統計学講師
- データ分析コンサルティング

- 趣味

- サッカー
- ピアノ
- 数学（数学オリンピック銀賞）



- スキル

- 統計学
- 多変量解析学
- 機械学習
- データ分析
- プログラミング言語(Python, R, C, C++, Fortran, Java)
- AI実装（Deep Learning）

本日の内容

1. テキストマイニングとは？
2. 様々な活用事例
3. 実行の手順
4. ツールによる実演
5. 習得するための手順

本日の内容

1. テキストマイニングとは？
2. 様々な活用事例
3. 実行の手順
4. ツールによる実演
5. 習得するための手順

テキストマイニングとは？

テキストマイニングの定義

the discovery by computer of new, previously unknown information,
by automatically extracting information from different written resources.

コンピュータを活用して、テキストデータから自動的に情報を
抽出し、**今まで分からなかった、新しい情報を発見すること。**

Marti Hearst

SIMS, UC Berkeley

<http://people.ischool.berkeley.edu/~hearst/text-mining.html>

テキストマイニングとは？

データ分析には2種類ある

定量分析：
数/数値に基づく分析

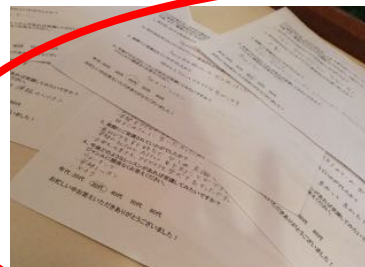


質的変数 (数)
例：性別



量的変数 (数値)
例：年齢

定性分析：
文章等に基づく分析



文章



音声

テキストマイニング



写真／動画

テキストデータの難しいところ

1. 言葉の定義

「今日ご飯楽しかったです」



デートの場合



苦手な上司との飲み会の場合

テキストデータの難しいところ

2. 日本語特有の難しさ

(英語) I want to learn Text Mining right now.

→区切られており簡単に単語ごとに集計できる

(日本語) テキストマイニングをこれから勉強します。

→文節に区切らないと集計できない

本日の内容

1. テキストマイニングとは？

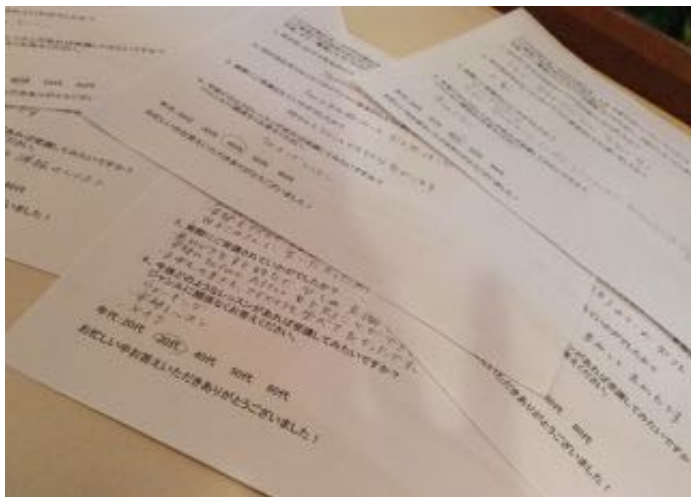
2. 様々な活用事例

3. 実行の手順

4. ツールによる実演

5. 習得するための手順

応用事例1:アンケート分析



セミナーの感想

良い点

- ・説明がわかりやすかった
- ・手を動かしながら学ぶので理解しやすい
- ・実際のデータを分析できるので楽しい
- ・「〇〇」のセミナーも参加したい

...

改善点

- ・時々早口になることがある
- ・解答をより充実させてほしい

何が評価されているのか、何が課題なのか発見する

応用事例2: 口コミ分析



口コミによる評価

良い点

- ・料理がとても美味しかった
- ・店内がとても清潔だった
- ・接客が素晴らしかった

...

改善点

- ・案内地図がわかりにくい
- ・PayPayが使えない

料理(商品)の良い点、悪い点は何か？
売れる店(商品)と、売れない店(商品)の違いは？
なりすましの口コミと、そうでないものの違いは？

応用事例3:SNS



SNS・ニュースアプリによるトレンドの分析
(例：コロナウイルス)

- ・ 感染拡大
- ・ 現金給付
- ...



今何が注目されているのか？
なぜ注目されているのか？
いいね・閲覧数が多い記事と少ない記事の違いは？

- [illegible]

Category	Percentage
中立 (Neutral)	78.3 %
ネガティブ (Negative)	13.6 %
ポジティブ (Positive)	8.1 %

喜び 😊

好き 😊

悲しみ 😞

恐れ 😨

怒り 😡

Copyright © 2020 Wakara Corp. All Rights Reserved.

本日の内容

1. テキストマイニングとは？
2. 様々な活用事例
3. 実行の手順
4. ツールによる実演
5. 習得するための手順

テキストマイニングの手順

データ取得

- アンケート・口コミ収集、Webスクレイピング

データ前処理

- 異常値の処理、ノイズデータの削除

形態素解析

- 今日/は/とても/楽しか/つ/た/。

集計・可視化

- 抽出後リスト、KWIC、対応分析、共起ネットワーク

本日の内容

1. テキストマイニングとは？
2. 様々な活用事例
3. 実行の手順
4. ツールによる実演
5. 習得するための手順

本日使用するツール

- KH Coder

本日の内容

1. テキストマイニングとは？
2. 様々な活用事例
3. 実行の手順
4. ツールによる実演
5. 習得するための手順