

テキストマイニングの実践 — 4日目 —

2020/7/31

ビジネス科学研究科
経営システム科学専攻

スケジュール

- 1日目: 7/1(水)
 - 説明 — テキストマイニングの手順
 - 実習 — データをよく知る (Excel)
- 2日目: 7/10(金)
 - 説明 — テキストマイニング ツールの使い方 (KHCoder)
- 3日目: 7/17(金)
 - 説明 — データ分析の実践 (KHCoder)
 - 実習 — データ分析の実践 (KHCoder)
- 体育の日: 7/24(金)
- 4日目: 7/31(金)
 - Text Mining Studio 利用体験
 - 実習 — データ分析の実践 (KHCoder)
- 5日目: 8/7(金)
 - 発表 — データ分析の実践 (KHCoder)

本日の内容

- 前半 18:20 ~ 20:10
 - Text Mining Studio 紹介
- 後半 20:20 ~ 21:00
 - 実習 グループワーク (議論や資料作成)

コースコンテンツ

第4回テキストマイニング資料

公開期間: 2020-07-31 18:00:00 ~ 2020-07-31 21:00:00



第4回テキストマイニング資料

practice-4.pdf - 2020-07-31 07:57:06

Text Mining Studio 資料

URL: <https://primedrive.jp/access?key=70896d3c-3e3f-4000-a000-000000000000>

パスワード:

2020-07-17 07:38 - 原田 智彦- 1.1版 (編集 | 履歴 | 閲覧確認)

予告: 発表会(8/7)の予定

- Q&A (Max ~18:35)
- 発表会 (各グループ 説明10分, 質疑5分)
 - 前半 18:35 ~ 19:35
 - グループ1 ~ グループ4
 - 後半 19:45 ~ 21:00
 - グループ5 ~ グループ9

予告: 発表内容

1. 利用データ

例) 「covid_2020.xlsx」であれば, 日/月/シーズン別の投稿件数など

2. テーマ設定

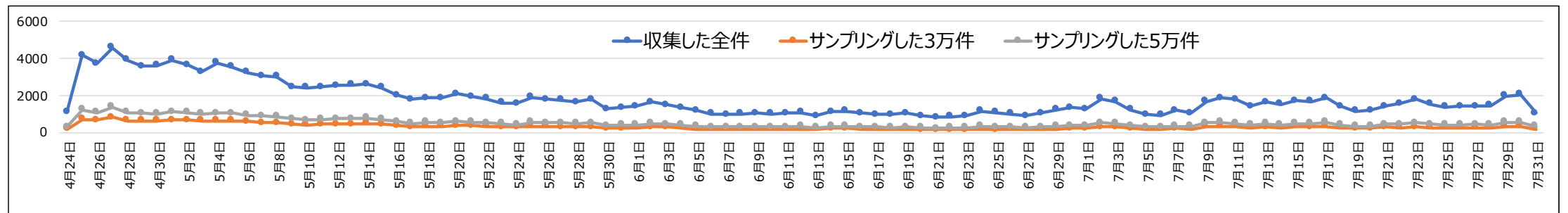
例) 「rakuten_2019.xlsx」であれば, 好評価のエリアに倣って, 低評価のエリアを改善する

3. 分析結果 (プロットおよび考察)

- テーマや仮説にもとづくストーリーで, 分析を進めるのがベター
- 支持する**プロット**とユーザーの**生の声(原文)を使って**主張する

covid_2020-0731.xlsx

- ハッシュタグ「#新型コロナ」を付けて Tweet されたデータ
 - 2020/4/24~7/30 に Tweet されたデータを Search API (1%) で収集
 - データ件数の日別分布を保持し, 30,000 or 50,000 件をサンプリング



緊急事態宣言中(~5/24)

緊急事態宣言後(~6/1)

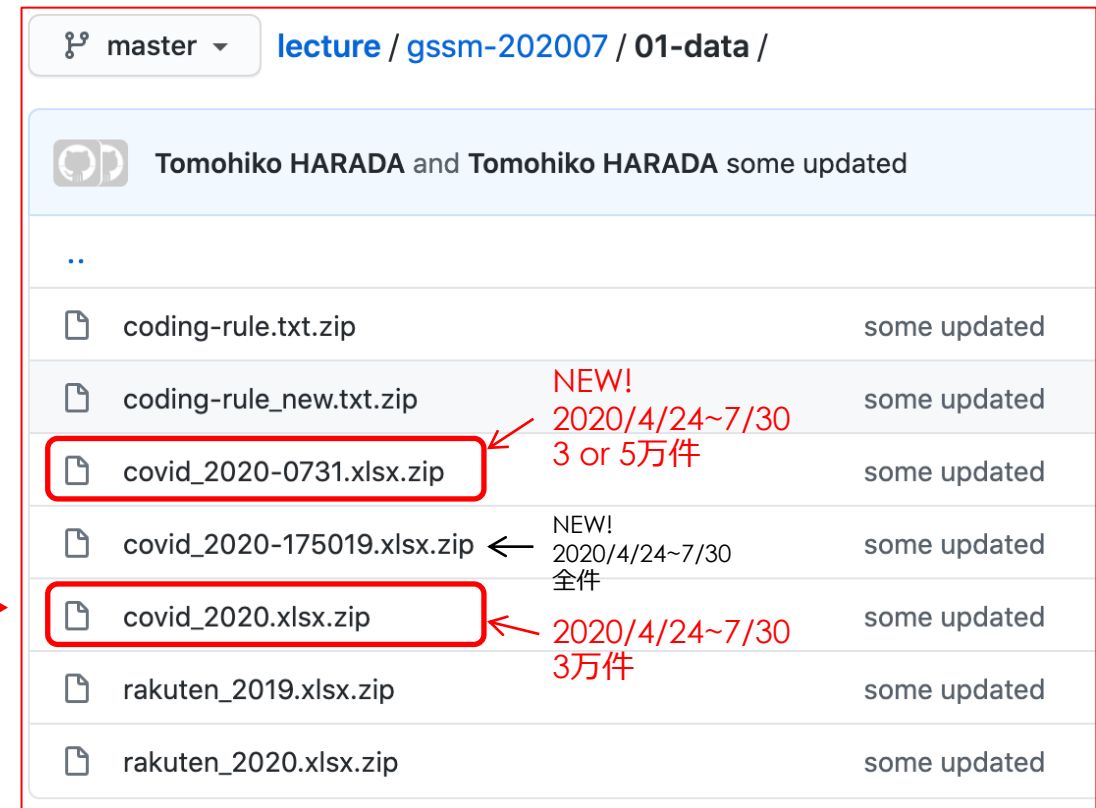
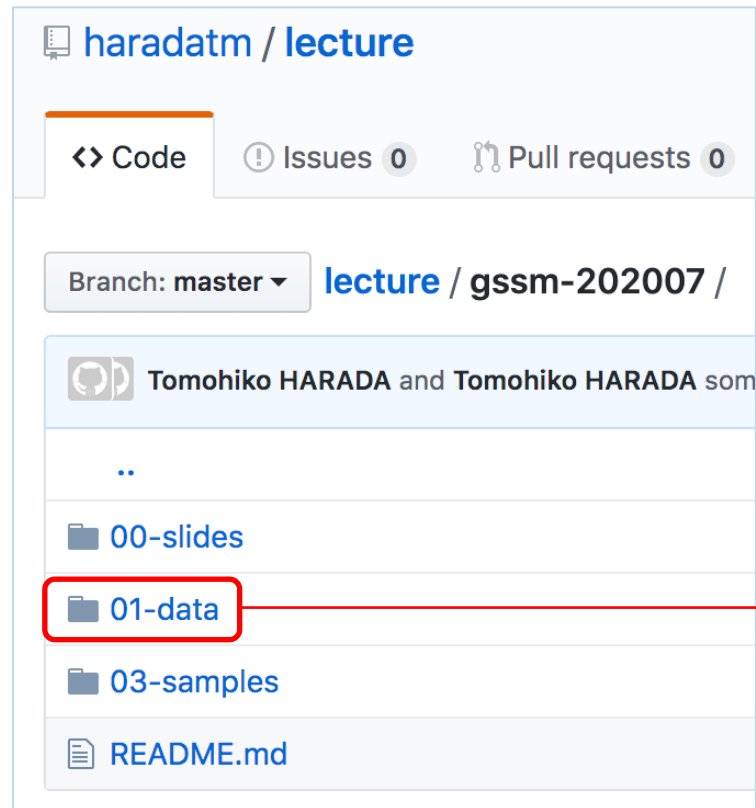
東京アラート発令中(~6/10)

東京アラート発令後

Tweet情報	6項目	ID, 投稿日, 投稿時刻, お気に入り数, retweet数, 言語
Tweet本文	1項目	コメント
ユーザー情報	4項目	ユーザーID, フォロワー数, フォロー数, ユーザーのこれまでの記事投稿数
その他の属性	1項目	シーズン {緊急事態宣言中, 東京アラート発令中, 東京アラート解除後}

covid_2020-0731.xlsx

- <https://github.com/haradatm/lecture/tree/master/gssm-202007>



課題 (4日目)

【注意】

TMS は、**全学のリモートデスクトップ** には**インストールできません!**
各自の Windows PC にインストールしてください

- 講義で利用した「rakuten_2019.xlsx」を使って、**TMS**で「ことばネットワーク」を作成し、KHcoder (共起ネットワーク) で見えない気づきがあれば述べてください
 - Windows PCがない方の救済: 受講したTMSセミナーを通して気づいた KHcoder との違い(メリデメ等)を述べてください
- グループワークについて、以下を記載して提出してください。
 - 所属する グループ名、グループで取り上げる 分析テーマ
 - メンバ全員の名前 および 発表者名
- 形式: PPT(PDF), 提出先: manaba, 期限: 8/14 23:55

残りの時間はグループワークです

グループ1	201940108
	202040051
	202040055
	202040074
	202040077
グループ2	201945014
	202040057
	202040063
	202040080
	MSI
グループ3	202040052
	202040062
	202040070
	202040072
	202040170

グループ4	201947529
	202040058
	202040060
	202040068
	202040069
グループ5	201940015
	201947523
	202040059
	202040066
	202040073
グループ6	201840109
	201940129
	202040064
	202040078
	202040079

グループ7	202040071
	202040076
	202040409
	202040413
	201540111
グループ8	202020027
	202020051
	202040067
	202040075
グループ9	202040053
	202040054
	202040056
	202040061

参考書

(KH Coder)

- [1] 樋口耕一. 社会調査のための計量テキスト分析ー内容分析の継承と発展を目指して【第2版】 KH Coder オフィシャルブック. ナカニシヤ出版, 2020.
- [2] 樋口耕一. テキスト型データの計量的分析ー2つのアプローチの峻別と統合ー. 理論と方法, 数理社会学会, 2004, 19(1): 101-115.
- [3] 牛澤賢二. やってみよう テキストマイニングー自由回答アンケートの分析に挑戦!. 朝倉書店, 2019

(Windows環境によるデータ収集方法の参考に)

- [4] テキストマイニングソフトを利用した新未来洞察手法の研究. 第10分科会, (財)市場創造研究会. http://www.shijo-sozo.org/news/第10分科会_1.pdf

参考書

(R を使った参考書)

- [5] 金明哲. "テキストデータの統計科学入門." 岩波書店, 2009.
- [6] 石田基広. "RMeCab によるテキスト解析. R によるテキストマイニング入門." 森北出版, 2008, 51-82.

(他のツールを使った参考書)

- [7] 那須川哲哉. "テキストマイニングを使う技術/作る技術: 基礎技術と適用事例から導く本質と活用法." 東京電機大学出版局, 2006.
- [8] 上田隆穂, 黒岩祥太, 戸谷圭子. "テキストマイニングによるマーケティング調査." 講談社, 2005.

(統計解析を中心とした参考書)

- [9] 前田忠彦; 山崎誠. 言語研究のための統計入門. くろしお出版株式会社, 東京, 2013.