

知能情報実験 III（データマイニング班）
Twitter 上のテキスト文を対象とした「コロナで何が困っ
ているのか」を見つける

グループの学籍番号 205759A, 205720E, 205763J, 195719J

提出日：2022 年 6 月 9 日

目次

1	はじめに	0
1.1	実験の目的と達成目標	0
1.2	テーマ**とは	0
2	実験方法	0
2.1	実験目的	0
2.2	データセット構築	1
2.3	モデル選定	1
2.4	パラメータ調整	1
3	実験結果	1
4	考察	1
5	意図していた実験計画との違い	1
6	まとめ	1

概要

概要本文書は知能情報実験 III（データマイニング班）におけるレポートのテンプレートとして用意したものである。一般的な実験レポートに関する補足と共に、データマイニング班における実験レポートに求められる内容を確認するために用意した。ここに書いてある事柄は全てを必須とするわけではなく、適宜取捨選択や追加編集してもらって構わないが、実験報告書としての位置づけを忘れずに利用すること。

1 はじめに

1.1 実験の目的と達成目標

知能情報実験 III は、情報工学分野のより専門的な知識を理解・習得することを目的として、半年間でシステムの開発やデータ解析等に取り組む実施される。その中の一つデータマイニング班においては機械学習外観ならびにその応用を通し、対象問題への理解、特徴量抽出等の前処理、バージョン管理やデバッグ・テスト等を含む仕様が定まっていない状況下における開発方法、コード解説や実験再現のためのドキュメント作成等の習得を目指す。

1.2 テーマ**とは

本グループでは Twitter 上のテキスト文を対象として「コロナで困っていること」を見つけることを対象問題として設定した。SNS を使ってコロナで発生している問題を可視化し、件数が多い項目を整理することで、その後の改善策を見出して今後に応用できる。それによって現在発生している問題に対処しつつ今後同様の問題が発生する際により効果的な対策ができるようになることが期待できる。

2 実験方法

2.1 実験目的

今回の実験の目的として、Twitter 上のテキスト文を対象として「コロナで何が困っているのか」という事柄を視覚的にわかるようにした。具体的な内容として、まず、Twitter 上のテキスト文の中から「コロナ」という単語が含まれた文章を抽出し、ネガティブかポジティブかの判定を行った。次に、その結果の中からネガティブな文章を取り出し、クラスタリングを行った。最後に、クラスタリングされた事柄の中で発言している人数が多い順に順位付けを行うことで優先的に対処しなければならない事柄を視覚的にわかるようにした。

2.2 データセット構築

Python の snsrape ライブラリ・モジュールを使用し、データセットを csv 形式のファイルで用意した。次に、tweet の文章だけを csv に出力した。この csv ファイルと、日本語評価極性辞書を利用した Python 用 Sentiment Analysis ライブラリ oseti を使ってネガティブとポジティブに分けた。その後、ネガティブに判定された tweet だけを抽出し、csv ファイルに出力した。

2.3 モデル選定

2.4 パラメータ調整

3 実験結果

4 考察

5 意図していた実験計画との違い

6 まとめ

参考文献

- [1] レポート作成の手引き レポートの基本的形式に関するガイド, <https://www.kanazawa-u.ac.jp/wp-content/uploads/2015/01/tebiki2.pdf>, 2020/07/02.