

# COVID-19関連記事が株価に与える影響力

## 1. 研究動機・目的

私たちは普段の生活の中で多くの情報をメディアから受け取っている 例：テレビ、新聞、ネットニュース

**行動が変化**

行動を変化させる力を**影響力**と定義

**影響力を定量化できないか？**

そこで、メディアが発信する情報を**COVID-19関連記事**、行動の変化を**株価の変動**として、2つのデータの関係性を分析

**株価の変動を予測 = 影響力の定量化**

## 2. 研究方法

### 1. 使用したデータ

・記事データ

NHK NEWS WEB 「コロナ関連記事全記録」見出し

期間：2020年1月25日～2020年5月31日

・株価データ

東証1部2186銘柄の始値

期間：2020年1月25日～2020年5月31日

### 2. 株価の変動の定義

株価の変動を定量的に判断

**ボリンジャーバンド**  
を用いて変動を定義

上例図の緑線・・・±σ  
青線・・・始値

始値 < -σ : 下落  
-σ < 始値 < σ : 変動なし

σ < 始値 : 上昇

下例図で、赤線：下落 緑線：変動なし 青線：上昇



### 3. 記事データの数値化

記事データを**形態素解析**、**分かち書き**

名詞のみを抽出

$$TFIDF_{w_i, d_j} = TF_{w_i, d_j} \times IDF_{w_i}$$

$TF_{w_i, d_j}$  = 文書  $d_j$  中の単語  $w_i$  の出現頻度

$$IDF_{w_i} = \log \frac{1 + \text{全文書数}(I)}{\text{単語 } w_i \text{ が登場する文書数}}$$

<https://www.takapy.work/entry/2019/01/14/141423>

出現日数 ≥ 100 の各名詞に対する **tf-idf 値** の変動を上式を用いて取得

### 4. 影響力の算出

株価の変動と tf-idf 値の変動を比較することで2つのデータの近似度を **I 近似度**、**R 近似度** として取得し、影響力を算出する。

## 3. データの解析・結果 I

始値と tf-idf 値を **正規化**

毎日に2データの差の絶対値の総和を図のように算出し、その値を **I 近似度** とする

各業種の企業ごとにこの操作を行いそれぞれの企業に対して **I 近似度** が小さい順に 1 ~ 3 位の単語を決定

業種ごとに各単語の 1 ~ 3 位になった回数を集計

結果(抜粋)

サービス業

単語	強化	安倍首相	東京
回数(1位)	73	48	34

情報・通信業

単語	強化	東京	対策
回数(1位)	81	55	21



## 4. データの解析・結果 II

前述の方法により毎日に株価の変動を **上昇(1)**、**変動なし(2)**、**下落(3)** の3つに分類

同日の tf-idf 値を、平均 + σ を基準として **小(0)**、**大(1)** の2つに分類

株価の変動と tf-idf 値の大小を組み合わせ、上図のような座標を取得し、相関係数を算出。その値を **R 近似度** とする

※ R 近似度が小さいほど上昇の影響が強く、大きいほど下落の影響が強い

結果(抜粋)

業種	下落 1 位	上昇 1 位	業種	下落 1 位	上昇 1 位	業種	下落 1 位	上昇 1 位
その他製品	休校	新型コロナ	化学	中止	新型コロナ	石油・石炭製品	休校	新型コロナ
その他金融業	休校	再開	医薬品	中止	新型コロナ	空運業	休校	新型コロナ
ガラス・土石製品	休校	再開	卸売業	休校	新型コロナ	精密機器	休校	新型コロナ
ゴム製品	休校	再開	小売業	休校	新型コロナ	繊維製品	休校	新型コロナ
サービス業	中止	新型コロナ	建設業	休校	新型コロナ	証券・商品先物取引業	休校	新型コロナ
ハルブ・紙	中止	新型コロナ	情報・通信業	休校	新型コロナ	輸送用機器	休校	新型コロナ
不動産業	中止	新型コロナ	機械	休校	新型コロナ	金属製品	休校	新型コロナ
保険業	中止	新型コロナ	水産・農林業	休校	新型コロナ	鉄鋼	休校	新型コロナ
倉庫・運輸関連業	中止	新型コロナ	海運業	休校	新型コロナ	鋳業	休校	新型コロナ

それぞれに影響が強い単語のグラフ

**下落**



**上昇**



## 5. 考察・今後の展望

・影響力の大きい単語として得られた結果は両近似度とともに **業種間の差異** はほとんど見られなかった。

・今後の展望としては、**N-gram** を用いてテキストを分割し、周辺の単語も考慮に入れた分析を行う。最終的に、他の期間や手法によって得られた影響力と比較し、**影響力の変化** や **最適な手法** を見つける。

### 参考文献

[1] 五島圭一、高橋大志、寺野隆雄(2015)「ニュースのテキスト情報から株価を予測する」『2015年度人工知能学会全国大会(第29回)』

[2] NHK NEWS WEB 特設サイト 新型コロナウイルス「コロナ関連記事全記録」  
<https://www3.nhk.or.jp/news/special/coronavirus/chronology/>