

# S08: 情報検索尺度を用いた、コロナ禍における不満の地域性の検出

長谷川康邦, 大野友瑛, 鶴飼颯哉, 夏見光太郎, 齊藤史哲 (千葉工業大学)

## 1. 研究背景と目的

コロナ禍により生活様式が変化したが、これに伴い人々は様々な不満を抱えていると考えられる。  
また、それらの不満にはある程度の地域性があると考えられ、地域ごとに不満を解決するための製品やサービスの創出が求められている。

実際に得られた不満データをもとに解析し、地域ごとの特性を調べ、その結果を新たな製品やサービスの提案へと繋げられるようにしたい。

## 2. 利用データ

Insight Tech様「新型コロナ 不満アンケートデータ」

アンケート実施回数 5回

アンケート実施期間 2020年3月2日~2021年1月10日

有効回答件数 16043件



## 3. 提案手法

本分析手法には情報検索尺度「BM25」を用いる

$$\text{score}(D, Q) = \sum_{i=1}^n \text{IDF}(q_i) \cdot \frac{f(q_i, D) \cdot (k_1 + 1)}{f(q_i, D) + k_1 \cdot \left(1 - b + b \cdot \frac{|D|}{\text{avgdl}}\right)}$$

$$\text{IDF}(q_i) = \log \frac{N - n(q_i) + 0.5}{n(q_i) + 0.5}$$

avgdl: 文書集合の平均単語数  
 $f(q_i, D)$ : TF(ある文書Dにおける単語  $q_i$  の出現頻度)

$k_1, b$ : パラメータ(任意)  
 $k_1 = 1.2$  または  $2.0$   
 $b = 0.75$   
に設定することが多い  
 $N$ : 全文書数  
 $n(q_i)$ :  $q_i$  を含む文書数  
 $|D|$ : 単語の総数

コロナへの不満の文書データに対して、BM25を用いたキーワードの検出を実施

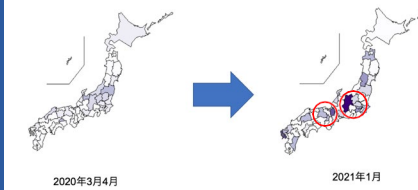


平均BM25を都道府県ごとに算出し、コロプレス図で地域の特徴を可視化

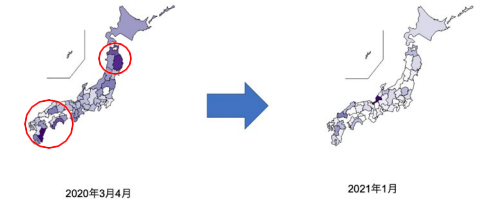
## 4. 結果

各地域における、選定したキーワードの地域別重要度

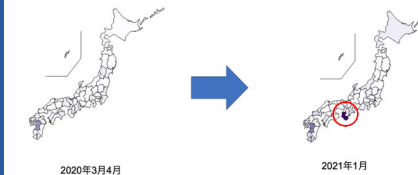
### A. テレワーク



### B. マスク



### C. 不安



### A. テレワーク

→ 首都圏や関西の主要都市で増加

### B. マスク

→ 全国的に減少傾向

### C. 不安

→ 全国的に微増、一部地域で上昇

## 5. 今後の展望

[1] キーワードのさらなる選定と相互作用の検討

→ 保育園, 予防, 自粛など...

[2] 特徴量抽出によるデータの可視化

→ 主成分分析, U-MAPなど...

[3] 地域性以外の着眼点での解析

→ 性別, 職業, 年収など...

### 参考文献

[1] Spärck Jones, K.; Walker, S.; Robertson, S. E. (2000). "A probabilistic model of information retrieval: Development and comparative experiments: Part 2". Information Processing & Management 36 (6): 809–840

[2] 佐藤裕次郎, 山西良典, 西原陽子, 宿泊施設のレビューの時系列分析による季節を表す特徴語の抽出, 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(第21回)

[3] 鳥海 不二夫, 榊 剛史, 吉田 光男, ソーシャルメディアを用いた新型コロナ禍における感情変化の分析, 人工知能学会論文誌/35 巻 (2020) 4 号