

# SNSを活用した大規模アンケート調査の代替可能性について

早稲田大学：海野智揮, 佐々木邦明

## 研究の背景・目的

MMで長年用いられているアンケート調査の課題解決策としてSNS等のログデータの代用が注目されている。そこで本研究では、MMにおける仮説検証を行うアンケート調査の代替に向けて、SNSから取得したデータの活用可能性を探る

### アンケートの主な課題

- 回答の**不正確性**
- 時系列性の**欠如**
- 回答者配慮による**制限** 村瀬ら(2021)

### SNSデータの主な課題

- 欲しい情報(回答)を得られるとは**限らない**
- **急なデータ提供の終了**がある

## MMでのSNS活用

モビリティマネジメントにおいてSNS等のログデータを活用することには以下のようなメリットがあると考え、検証を行った。

### 態度変化の計測

実施したMMについての市民の意見や感情を**即時かつ時系列的**に取得し、態度や行動の**きっかけ**を見つけることができる。

### 行動変化の計測

### コスト面の優位性

同程度の情報量を集めるにあたってアンケートより**安い**場合が多い

## ネットワーク形成の容易さ

### 特徴

ユーザの記憶や感情に依存せず条件を設けることで機械的にユーザ同士のつながりを取得することができる

### 活用例

ネットワークの可視化、分析を蓄積することで移動データなど様々なデータとの組み合わせが期待できる

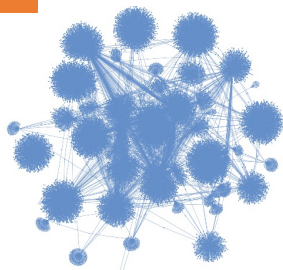


表1：対象ユーザの媒介中心性

ユーザネーム	媒介中心性
sagami_railway	0.1951
kishino_yamaki0718	0.1562
hidemi826	0.1492
sagami_kanako.one_heart_train	0.1348

表2：対象ユーザの次数中心性

ユーザネーム	次数中心性
run.nobu	0.0733
tae.oko.chobi	0.0703
bgzp_ra	0.0682
himalbum	0.0678

東急電鉄、相模鉄道のInstagram公式アカウントのフォロワーからそれぞれランダムに50アカウント抽出し、さらに各アカウントのフォロワー・フォローアカウントを最大990件取得した。そして、そのすべてのアカウントでネットワークを形成し同時に媒介中心性・次数中心性を求め、アカウントの属性を調べた。その結果、ユーザの統一性は見られず、また、アカウントを大量にフォローしているのみで活動していないようなアカウントが上位に来ることがわかった。

図1：Instagram対象ユーザのネットワーク

## 時系列的な追跡が可能

### 特徴

ユーザがSNSに投稿した瞬間からデータの取得ができる

### 活用例

時系列的なMMに関わる話題の取得と、話題に対するユーザの感情の動きの取得によるMM施策の評価

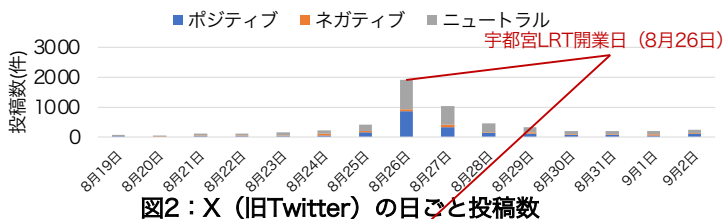


図2：X (旧Twitter) の日ごと投稿数

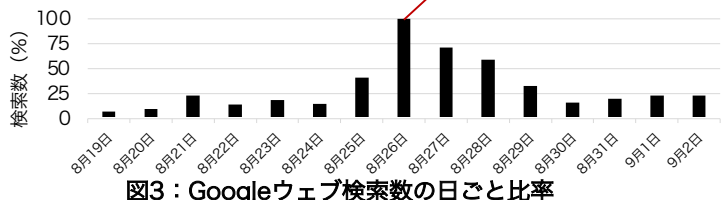


図3：Googleウェブ検索数の日ごと比率

“宇都宮ライトレール”のキーワードでX、Googleトレンドそれぞれで検索をかけて開業前後1週分の結果をまとめた。また、Xについては投稿のテキストを取得し機械学習手法であるBERTを使用して各投稿の感情を分類した。また、感情分類した投稿のうちの頻出単語を表5に示す。

表3：ポジティブ・ネガティブ分類投稿の頻出単語

ポジティブ単語	出現回数	ネガティブ単語	出現回数
路線	4139	広告	1421
年	4000	電車	1345
新規	3698	遅れ	1208
車両	3610	現金	1183
いい	3521	イラスト	1157

検索結果については、X上の投稿数とGoogleウェブ検索の傾向は一致しており、Xのデータは世間の注目を反映しているといえることがわかった。テキスト分析については、ポジティブ投稿には特徴的な単語がみられなかった。

一方で、ネガティブ投稿には**広告・イラスト**、**現金**といったキーワードが見られ、**広告イラストの内容が問題視されていたこと**、**大量の現金払い乗客による電車遅延**といったイベントを表現できているといえ、**SNS上のテキスト分析はユーザのMM施策の評価とその要因の分析に活用できると考えられる。**

## コスト面の優位性

### 特徴

- 少量データは無料取得可能
- SNSごとに提供機能が異なり、欲しいデータを取得できない場合もある

### 活用例

少数ユーザの投稿を時系列的に取得してMM前後の態度変容を確認する

表4：調査手法別費用

データ取得件数	紙・郵送	電子アンケート	X (投稿)
1000件	661,000円	116,000円	free
2000件	1,183,000円	116,000円	100.00USD/month
100万~無制限	-	-	2,000,000円~/月

## まとめ

### ネットワーク形成の容易さ

SNSデータでは代表性の担保は難しいものの、ユーザネットワークの形成と時系列的な追跡を組み合わせることでユーザを分類することで非集計的にユーザの態度・行動の変化について検証が可能だと考えられる。

### 時系列的な追跡が可能

イベントの発見と、それに対する感情判定ができたことから新規施策の評価に活用できると考えられる。

### コスト面の優位性

アンケートと比べて安いコストでデータ取得が可能であり、特に安価で行いたい調査の際はアンケートの代替になりうる。