D-028

# Twitter からの土産情報の抽出・分析と活用法の検討
# Consideration of Extraction, Analysis and Utilization of Information about Souvenirs on Twitter

真辺 諒† 長尾 哲志† 安藤 一秋‡
Ryo Manabe Noriyuki Nagao Kazuaki Ando

## 1. はじめに

旅行に行った際，多くの人が土産を購入するが，土産の選定は悩ましい問題である．土産の選定時に，現地でしか購入できない商品や人気のある定番商品，現在注目されている商品およびそれらの口コミ情報等がわかれば，土産の選定が比較的容易になると考える．

長尾らの研究[1]では，現地でしか購入できない土産に焦点を当て，土産情報を QA サイトや口コミサイト，ブログから自動収集して，ユーザに提示するシステムの構築を目指している．しかし，長尾らが対象とする土産情報は，比較的静的な情報であるため，リアルタイム性や話題性に関する情報が不足する．

そこで，動的な情報を補うため，Twitter に注目する．Twitter は即時性や簡便性の高さから多くの人が利用し，日々大量の情報を発信している．現在，月間アクティブユーザ数は 3 億 1000 万人，一日に投稿されるツイート数は 5 億件を超える．国内ユーザ数は 3500 万人となっており，およそ日本人の 4 人に 1 人が利用していることになる．このように大量に発信される情報の中には，土産に関する情報も存在している可能性がある．

本稿では，土産に関するリアルタイム性や話題性に関する情報を収集するためのリソースとして，Twitter の有用性を確認することを目的とする．特に，ツイート内に土産に関するどのような情報が含まれているのか，また，有用な情報がどの程度存在するのか等について調査・分析する．

## 2. 都道府県別のツイート分析

### 2.1 分析目的と方針

土産は，地域別に特徴があると考えられる．そこで，土産に関するツイートを都道府県別に分析する．土産名，店舗名，評判，土産画像等を土産に関する有用な情報と仮定し，ツイート中にこれらの情報がどの程度含まれているかを調査する．そして，都道府県毎の特徴やツイートに含まれる情報について考察する．

分析方法について述べる．まず，Twitter から土産に関するツイートを収集し，都道府県別に分類する．そして，土産名，店舗名，評判，土産画像などの情報，楽天市場やまとめサイトなどの他サイトの宣伝ツイートや土産に関係しないノイズツイート等のノイズがどの程度含まれているのかを人手（1 名）で調べる．また，画像を含むツイートと含まないツイートに分けて分析し，画像の有無による情報の違いを比較する．

†香川大学 大学院工学研究科 Kagawa University Graduate School of Engineering
‡香川大学 工学部 Kagawa University Factory of Engineering

### 2.2 分析データ

Twitter Rest API を利用し，「土産」または「みやげ」と都道府県名を含むツイートを収集する．本分析では，2015 年 12 月 1 日から 31 日の一ヶ月間に収集したツイート 124,082 件を利用する．本分析には，北海道，京都，徳島，香川，愛媛，高知の 6 県のいずれかに分類されたツイートを利用する．

6 県のツイート件数を表 1 に示す．なお，北海道と京都に分類されたツイートは数が多いため，画像を含むツイートと含まないツイートをそれぞれ 200 件ずつ抜粋して分析対象とする．徳島，香川，愛媛，高知に関しては収集データの全てを対象とする．

表 1．6 県のツイート件数

| 都道府県名 | データ数 |
|---|---|
| 北海道 | 4,619件 |
| 京都 | 2,488件 |
| 徳島 | 137件 |
| 香川 | 369件 |
| 愛媛 | 147件 |
| 高知 | 168件 |

### 2.3 分析結果と考察

画像有りツイートの分析結果を表 2 に，画像無しツイートの分析結果を表 3 に示す．表 2 と表 3 を比較すると，どの都道府県もノイズツイートの数が画像有りツイートに比べて画像無しツイートの方が多く，画像を含むツイートは有用な情報を含む可能性が高いことが分かった．

宣伝ツイートからは，商品名や店舗名等の土産情報が獲得できると判断したため，ノイズツイートと区別して整理したが，ツイート内には宣伝ツイートも多数含まれることが分かった．そこで，画像有りツイートの内，特に宣伝ツイート数が多い北海道のツイートに対して分析を行う．

宣伝ツイートとノイズツイートをフィルタリングした北海道のツイート 50 件のうち，土産名を含むツイート数は 22 件（44.0%），店舗名を含むツイート数は 2 件（4.0%）となった．いずれも件数が大幅に減少したことから宣伝ツイートには，土産名や店舗名が含まれている可能性が高いことが分かった．しかし，宣伝ツイート 145 件の内 109 件が「六花亭のバターサンド」に関するほぼ同一のツイートであった．これらのツイートに関してはフィルタリングする必要がある．

次に，表 2 を基に都道府県のツイートの特徴を考察する．観光地として有名な北海道，京都に比べて四国四県は，土産に関するツイート数が少なく，また，画像有りツイート数も数十件程度しか存在しなかった．しかし，愛媛は有用な情報の割合が高い．また，ツイート数の多い北海道は，

表 2. 画像有りツイートの分析結果

| | 土産名 | 店舗名 | 評判 | 土産画像 | 宣伝 | ノイズ |
|---|---|---|---|---|---|---|
| 北海道（200件） | 163件(81.5%) | 113件(56.5%) | 12件(6.0%) | 192件(96.0%) | 145件(72.5%) | 5件(2.5%) |
| 京都（200件） | 116件(58.0%) | 9件(4.5%) | 39件(19.5%) | 160件(80.0%) | 46件(23.0%) | 15件(7.5%) |
| 徳島（18件） | 8件(44.4%) | 0件(0.0%) | 4件(22.2%) | 16件(88.9%) | 1件(5.6%) | 1件(5.6%) |
| 香川（20件） | 10件(50.0%) | 0件(0.0%) | 2件(10.0%) | 16件(80.0%) | 2件(10.0%) | 3件(15.0%) |
| 愛媛（27件） | 17件(63.0%) | 3件(11.1%) | 9件(33.3%) | 24件(88.9%) | 1件(3.7%) | 2件(7.4%) |
| 高知（50件） | 28件(56.0%) | 2件(4.0%) | 1件(2.0%) | 45件(90.0%) | 20件(40.0%) | 5件(10.0%) |

表 3. 画像無しツイートの分析結果

| | 土産名 | 店舗名 | 評判 | 宣伝 | ノイズ |
|---|---|---|---|---|---|
| 北海道（200件） | 81件(40.5%) | 1件(0.5%) | 17件(8.5%) | 72件(36.0%) | 89件(44.5%) |
| 京都（200件） | 60件(30.0%) | 7件(3.5%) | 14件(7.0%) | 41件(7.0%) | 104件(52%) |
| 徳島（119件） | 31件(26.1%) | 5件(4.2%) | 8件(6.7%) | 16件(13.4%) | 83件(69.7%) |
| 香川（349件） | 39件(11.2%) | 1件(0.3%) | 11件(3.2%) | 259件(74.2%) | 60件(17.2%) |
| 愛媛（120件） | 40件(33.3%) | 2件(1.7%) | 5件(4.2%) | 35件(29.2%) | 69件(57.5%) |
| 高知（118件） | 27件(22.9%) | 3件(2.5%) | 7件(5.9%) | 22件(18.6%) | 77件(65.3%) |

土産名と店舗名が多く含まれていたが，評判情報は他県と比べて低い．

観光地として有名な北海道，京都と四国四県を比較したが，土産情報として大きな特徴は見られなかった．また，どの県においても土産名は高頻度で出現するが，店舗名，評判などの情報は少なかった．そこで，次節では，商品名を含むツイートにおいて，どのような特徴が見られるのか分析する．

## 3. 商品名に基づくツイート分析

### 3.1 分析目的と方針

商品名を用いてツイートを収集し，有用な情報がどの程度含まれているか，商品名に対してどのような情報が併記されやすいか，また，取得したデータに含まれる有用な情報の割合について分析・考察する．本分析では，宣伝ツイートをフィルタリングし，一般ユーザのツイートに限定して，人手（1 名）で分析する．

### 3.1 分析データ

北海道土産の「ドゥーブルフロマージュ」をキーワードに設定し，2016 年 6 月 10 日から 17 日の一週間で収集したツイート 190 件から宣伝ツイートをフィルタリングした 111 件のツイートを分析対象とする．

### 3.2 分析結果と考察

111 件のデータを分析した結果，ツイートから商品画像と評判に関する情報が多く得られた．商品画像は 29 件（26.1%）のツイートに含まれていた．評判は 50 件（45.0%）のツイートに含まれており，商品名とその評判が併記されやすい可能性が高いと考えられる．商品画像と評判の両方を含むツイートは 16 件（14.4%）存在した．

また，「ルタオのドゥーブルフロマージュ」という「店舗名＋商品名」といった表記が多く見られた．111 件中 25 件（22.5%）にこの表記が用いられており，店舗名の自動抽出に応用できると考えられる．

## 4. おわりに

本稿では，土産情報を収集するためのリソースとして，Twitter の有用性を確認するための分析を行った．

都道府県別の分析の結果，土産名は高頻度で出現するが，店舗名，評判などの情報があまり含まれないことが分かった．また，観光地として有名であるか否かで収集できるツイート数に大きな差があるため，あまりツイートされない都道府県の情報に関しては活用が難しいと考えられる．しかし，テレビ等のメディアで特定の土産が紹介されるなど，世間から注目された場合，その土産に関するツイートが急上昇する「バースト現象」が起こる可能性がある．バースト現象は，リアルタイム性や話題性の目安になる．

一方，商品名による分析では，ユーザによる評判が併記される可能性が高いことが分かった．以上より，土産に関するリアルタイム性や話題性に関する情報を収集するためのリソースとして，Twitter は有用である可能性がある．

今後の課題として，様々な都道府県の多様な商品について分析を継続する．また，バースト現象[2]に関しても分析を行い，その有用性について調査する．イベント抽出技術[3]を参考にしながら，ツイートから土産情報を抽出する手法を検討する．

### 参考文献

[1] 長尾 哲志，安藤 一秋，“オンラインショップで購入できない土産を提示するシステムの検討”，FIT 2015 論文集，pp.71-72，(2015).

[2] 佐々木 謙太朗，田村 一樹，吉川 大弘，古橋 武，“Twitter における話題語の抽出と周期に基づく分類”，言語処理学会 第 19 回年次大会 発表論文集，pp.806-809，(2013).

[3] 山田 渉，菊池 悠，落合 圭一，鳥居 大祐，稲村 浩，太田 賢，“マイクロブログを用いたイベント情報抽出技術”，情報処理学会論文集，Vol.57，No.1，pp123-132，(2016).