

Q&A コミュニティにおける投稿内容の経年的変化

吉見憲†

成蹊大学†

1. はじめに

Yahoo!知恵袋に代表される Q&A コミュニティはソーシャルメディアに比べて投稿量は少ないものの、質問や回答がカテゴリ化されているため、ユーザーの投稿意図に対してより精度の高い情報を取得することが可能である。本研究では、国立情報学研究所がヤフー株式会社から提供を受けて研究者に提供している「Yahoo! 知恵袋データ (第3版)」のデータセットを用いて、同一カテゴリにおける投稿内容の経年的な変化について検討する。Q&A コミュニティにおける投稿内容が商品やサービスの評価に関する先行指標として用いることができれば、マーケティングや商品開発等に活用することが期待できる。

2. 先行研究

ブログ等の Web コミュニティからの観光情報の抽出と分析は以前より行われてきた[1]。スマートフォンやソーシャルメディアの普及によりその応用範囲は広がっているものの、精度の高い分析を行うためにはノイズへの対応が必要となる[2][3]。

一方で、Yahoo!知恵袋の投稿データを分析対象とした先行研究では、投稿意図のはっきりとした質問文を用いたことにより、計量テキスト分析の手法から有用な知見を得ることに成功している[4]。ただし、当該の研究では3年分の投稿データを一括して扱っているため、時間経過が与える影響等については詳細に扱うことができていない。そこで、本研究では年度ごとにデータを区分することにより、経年的な変化について分析する。

3. 分析内容

3.1. 分析に用いるデータ

本研究では、国立情報学研究所が提供する「Yahoo!知恵袋データ(第3版)」の2019年度提供版と2020年度提供版における質問文を利用した[5]。Yahoo!知恵袋は日本最大の Q&A コミュニティであり、多くのユーザーから質問と回答が寄せられている。当該データには、期間中(2014年4月1日-2018年3月31日)に投稿され解決

した質問の10%がランダムサンプリングされたものが収録されている。

分析に当たっては、提供されたデータから「>地域, 旅行, お出かけ>国内>観光地, 行楽地」カテゴリの投稿のみを抽出した。そのため、明確に観光, 行楽目的の投稿であると解釈できる。分析対象となった投稿は21,546件となった。各年度の月別投稿数は図1の通りである。8月頃に投稿数が増え、冬場にかけて投稿数が減っている点は共通しているが、他の年度に比べて2014年度前半のみ著しく投稿数が多くなっている。

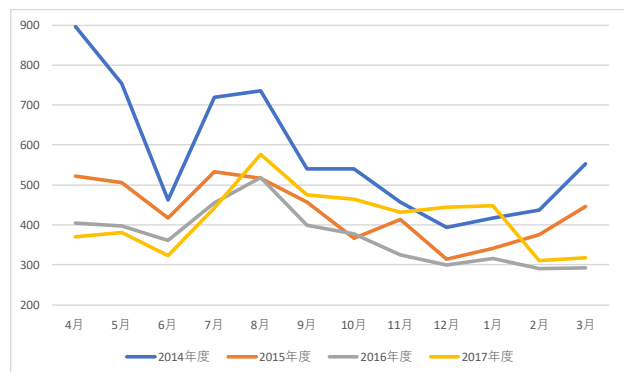


図1 各年度の月別投稿数

3.2. 分析手法

分析に際して、フリーのテキストマイニングソフトウェアであるKH Coder (<https://khcoder.net/>)を使用した。形態素解析は付属のChaSen(茶筌)を用いている。

分析の手順として、まず経年的な全体の傾向をコレスポンデンス分析で可視化し、先行研究の結果と比較した。次に、登場頻度が高い地名の年度別の登場数についてカイ二乗検定から有意差が生じているか検討した。最後にこれらの結果を踏まえて、国内旅行の動向とどの程度一致しているか考察した。

なお、分析対象の単語については出現数500回以上という制約を設けている。ここでの出現数は投稿中の出現の有無のみをカウントしているため、1投稿に複数回登場しているものであっても1回となる。

4. 分析結果

4.1. コレスポンデンス分析

コレスポンデンス分析の結果は図1の通りとなった。横軸を示す成分1は50.6%、縦軸を示す成

