

ツイートデータを用いた消費者コミュニティの検出および特徴把握

三宅 伸[†] 大竹 恒平[‡] 生田目 崇[¶]中央大学大学院[†] 東海大学[‡] 中央大学[¶]

1. はじめに

近年、急激な Social Networking Services (SNS) の普及に伴い、人々の交流方法が大きく変化した。ビジネス領域に目を向けると、多くの企業が消費者とのコミュニケーション・チャンネルとして SNS を利用している。SNS を活用したマーケティング戦略は “Social Media Marketing (SMM)” と呼ばれる。これまでに SMM の事例や SMM を対象とした数多くの取組みが紹介されており [1], SNS を介したマーケティング活動やプロモーションが注目されている。ただし、これまでの研究においては、対象とする消費者（顧客）全体に着目したものが多く、消費者間で形成されるコミュニティ単位での SMM に着目したものは少ない。消費者（顧客）の価値観は多様であり、SNS を通じてより効果的にマーケティング活動を行うためには、嗜好が共通するコミュニティ単位で、消費者の特徴を理解することが重要であると考えられる。

2. 本研究の目的

本研究では、あるファッションブランドを対象に、SNS 上に構成されている消費者コミュニティの検出を目的とする。特に、ファッションブランドに関して構成されるコミュニティでは、消費者の嗜好が類似するため、コミュニティ間の差が現れやすいと考えられる。具体的には、コミュニティ内に属する消費者の投稿コンテンツ（ツイート）を利用し、コンテンツの投稿傾向を利用した各コミュニティの特徴を明らかにする。

3. データ概要

日本のファッション市場にて 10 代から 30 代の女性に人気のファッションブランドを対象に、SNS 上の消費者から構成されるコミュニティを算出する。データは Twitter API を利用し、ブランドの公式アカウントをフォローする消費者から、新規にフォローした消費者から遡って 1000 ID を

抽出した。また、それらの ID をフォローする全ユーザの情報を抽出した。対象ツイートは 1000 ID の 2019 年 12 月 1 日から 2019 年 12 月 31 日の期間に投稿されたものである。

4. 分析方法

4.1 ネットワーク分析とコミュニティ検出

本研究では消費者ネットワークを、アカウント間のフォロー関係を利用し可視化する。ブランドのアカウントをフォローする 1000 人の消費者をネットワークのノードとし、彼らをフォローするユーザの共起から、式 (1) の Dice 係数によりアカウント間の類似度を算出する。

$$Dice(A, B) = \frac{2|A \cap B|}{|A| + |B|} \quad (1)$$

ネットワークの可視化には、類似度が 0.03 以上のノードのみ利用し、力学モデルの “Fruchterman Reingold model” を利用する。また、算出されたネットワークに対し、式 (2) の Modularity [2] を用いたコミュニティ検出を行う。

$$Modularity = \frac{1}{2W} \sum_{i', j'} \left(W_{i', j'} - \frac{w_{i'} w_{j'}}{2W} \right) \delta(C_{i'}, C_{j'}) \quad (2)$$

δ はそれぞれのノードが同じコミュニティに属するか否かをクロネッカーデルタで表したものであり、 W は全てのノードにおける類似度の合計値である。また、 $w_{i'}$ と $w_{j'}$ は、各ノード間の類似度を表す。そして Modularity が最大になるようコミュニティへの分割を行う。

4.2 トピック分類モデル

各コミュニティに属する消費者の各投稿コンテンツを用い、トピックモデルを利用して、トピック別の確率分布を推定する。LDA モデル [3] では、単語 w によって表現された文書の集合と、トピック数 K によって、各文書 b のトピック z_n の確率分布 $P(z_n | b) (n = 1, \dots, K)$ と、各トピック z_n における単語 w の確率分布 $P(w | z_n) (w \in V)$ を推定する。本研究ではトピック数 K を 6 とし、期間内のツイートに含まれる単語をモデルに代入する。

5. 分析結果

分析の結果、37 のコミュニティが検出された。図 1 にコミュニティ検出の結果を表す。

Detection and characterize consumer communities on Twitter

[†] Shin Miyake, Chuo University[‡] Otake Kohei, Tokai University[¶] Takashi Namatame, Chuo University



図 1. コミュニティ検出結果

図 2 の中心部分にコミュニティの集中が見られた。これはコミュニティ内に属するノード数の多い、大規模なコミュニティにより構成されていることがわかった。また、このコミュニティの集中の周りに中規模のコミュニティが広がり、ネットワークの外側に、小規模のコミュニティが散らばっていた。

次に、各コミュニティの特徴を明らかにするため、コミュニティ内に属する消費者が実際にツイートを利用し、トピック分類を行う。以下の表 1 に各トピックでの代表単語をまとめる。

表 1. 各トピックの代表単語(上位 10 個)

トピック 1	トピック 2	トピック 3	トピック 4	トピック 5	トピック 6
日	フォロー	人	時	フォロー	フォロー
France	名	こと	こと	名	応募
ans	プレゼント	好き	情報	プレゼント	日
私	投稿	日	お願い	応募	プレゼント
月	キャンペーン	年	my	12	クリスマス
年	クリスマス	剣	度	抽選	方法
発売	応募	私	時間	キャンペーン	12
人	12	住所	絶対	円	品
限定	毎日	100	私	方法	ツイート

トピック 1 は“日”、“発売”や“限定”など単語から新商品などに関する投稿と推察される。トピック 2, 5, 6 ではどれも“フォロー”や“プレゼント”などの単語から、応募キャンペーンに関する内容であると推察される。トピック 3 は、“好き”、“住所”などの単語から Twitter を介した商品販売に関する内容であると推察できる。トピック 4 は“情報”や“時間”、“絶対”など情報のトレンドに関する内容であること推察できる。この LDA モデルをもとに、各消費者の投稿コンテンツを分類し、各消費者のトピックに対する投稿傾向を算出した。ここでは、中規模コミュニティの各トピックへ

の投稿傾向を表す。図 3 は、左のノードを中規模のコミュニティとし(番号はコミュニティ分類の際の分類番号とする)、右のノードを各トピックとした 2 部グラフであり、相互のノード間のエッジを消費者のコンテンツの投稿傾向、重みを比率で表している。

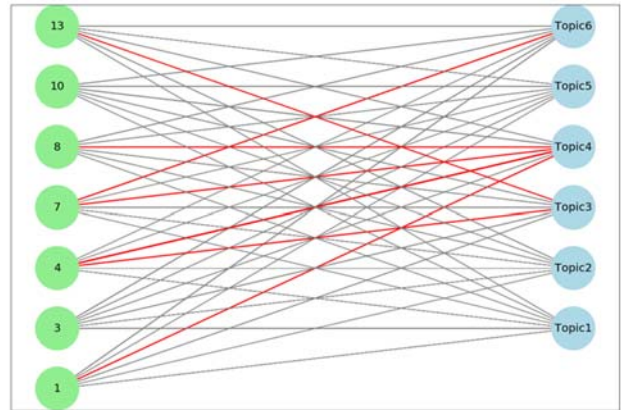


図 3. 中規模コミュニティの投稿傾向

太い線が、0.2 以上の高い投稿傾向を表す。コミュニティ 1, 4, 7, 8 は、トピック 4 に関する投稿傾向が最も多く、さらにコミュニティ 4 はトピック 3, コミュニティ 7 はトピック 6 に関する投稿も多く行っていることがわかった。以上の結果から、規模によって投稿傾向のあるトピックが異なることがわかった。

6. 結論

本研究では、SNS 上の消費者が成すコミュニティを、フォロー関係を利用して分割し、コミュニティ別でツイートの投稿傾向を分析した。その結果、フォロー関係だけではなく投稿している内容にも違いがあることが分かった。多数のトピックを投稿している、逆に少数のトピックのみを投稿しているコミュニティがあるなど、コミュニティによって投稿しているトピックに違いがあることが明らかとなった。さらに、コミュニティ内の繋がりには、消費者個人の嗜好的特徴も影響することが明らかとなった。

参考文献

- [1] 大西 浩志, “ソーシャルメディアとマーケティング研究,” *Japan Marketing Journal*, Vol. 34, No. 3, pp. 58-68, 2015.
- [2] G. Agarwal and D. Kempe, “Modularity-maximizing graph communities via mathematical programming,” *The European Physical Journal*, Vol. B. 66, pp. 409-418, 2008.
- [3] D. M. Blei, A. Y. Ng, and M. I. Jordan. “Latent Dirichlet allocation,” *Journal of Machine Learning Research*, Vol. 3, pp. 993-1022, 2003.