

# Twitterの特徴を考慮したLDAによるトライブ推定と重みづけ提案

早稲田大学 \*寺畑勇希, 東京大学 川中孝章, 早稲田大学 枝川義邦, 早稲田大学 大野高裕

## Tribe estimation and proposal of tribe weighting process by LDA considering the characteristics of Twitter

Waseda Univ. \*Yuki TERAHATA, The University of Tokyo Takaaki KAWANAKA, Waseda Univ. Yoshikuni EDAGAWA, Waseda Univ. Takahiro OHNO

### 1. 研究の背景と目的

SNSを通して、自社製品のPRを行う企業にとって、エンゲージメント率の向上は必要不可欠である。近年のツイート当たりの平均エンゲージメント率は、0.47%(2017年)→0.45%(2018年)→0.38%(2019年)と低下傾向にある。なぜなら、SNSの発展に伴い、消費者の異質性を考慮する必要性が増しているからだと考えられる。そのため、企業側がSNSで消費者にアプローチする際に、“特定の興味関心や趣味嗜好”に対して共通の価値観を持った消費者を一つの集団(トライブ)として捉える必要があると考える。それによって、各トライブに的確なアプローチを行うこと(トライブマーケティング)が期待されるからである。

そこで、本研究では、①投稿の気軽さと②拡散力という2つの観点から、SNSの中でもTwitterにおけるトライブマーケティングを題材として扱う。Twitterでのトライブマーケティングを行う際に、企業側は経験則でトライブ抽出を行っているため、学術的なトライブ抽出方法を定義しなければならず、①各ツイート文全体からテキストベースでトライブを抽出する方法と②各ツイート文がどのトライブに所属するかを文章全体から判断する方法を確立し、さらに、上記のトライブ分類の有効性を学術的に裏付ける必要がある。

また、Twitterマーケティングに関して、守口ら[1]は、「受動的なツイートでは効果が少なく、ポジティブな口コミを少しずつ増やすことが効果的である」と述べているため、トライブを用いて、多様化する消費者の趣味/価値観を正確に捉え、ポジティブな口コミを増やすことに繋げていかなければならないと考えられる。よって、本研究の目的は、①消費者の価値観をより

正確に捉えるトライブ分類方法とトライブがツイートの特徴とエンゲージメントの関係性に与える影響を示し、トライブ分類の学術的な有効性を明らかにすると同時に、②消費者のエンゲージメント&ポジティブ投稿の量とトライブとの関係性について分析する。

### 2. 従来研究

杉浦[2]はTwitterにおけるトライブマーケティングに関して、各ツイートのテキスト分析と頻出語を基にした共起ネットワーク図から、トライブを抽出し、トライブごとにエンゲージメントに影響を与える要因が異なることを明らかにした。しかし、どのトライブがポジティブな投稿を生みやすいかなどコンテンツに関する評価や考慮がない。また、トライブ分類に関して、共起ネットワーク上の頻出語が各ツイート内にあるかないかでトライブへの所属を決定しているため、ツイート分全体を考慮しての分類とは言い難い。

### 3. 研究内容

#### 3.1 研究方法

トライブ抽出後に、「どのような価値観や趣味を持っている消費者が自社の商品を使っているのか」が明確でなければならず、企業側がそれをマーケティングに活かすことを念頭に置くと、消費者の趣味や価値観と商品の利用シーンが密接な関係にあるような商品が理想的である。以上をもとに、「カメラ」を研究対象商品とする。

研究手順に関しては、以下の4ステップで行う。

(1)ツイートデータ収集

Twitter APIを用いたPythonコードを作成し、カメラに関する投稿を収集する。

(2)各ツイートのPositive Negative 度合い(PN値)算出  
形態素解析によって切り分けた各ツイートの各単語のPN値(Positive/Negative 度合い)の平均値をそのツイートのPN値とする。(高村ら[4]が作成した単語感情極性対応表における各単語のPN値を用いる)

(3)トライブ抽出

精度の高いトライブ抽出を以下3段階で行う。

①ストップワードの抽出&除外

①-1 DF(Document Frequency)を算出

予め設定した閾値よりもDF値(ある単語が出現する文書数)が大きい単語をストップワード(処理対象外とする単語)とする

①-2 単語間の類似度を算出

DFの上位単語について、意味的に類似している単語をストップワードにする(word2vecの利用)

②LDAを用いたトピック推定(トライブ分類)

従来研究より、それぞれのツイートがどんなトピックを、どの程度の割合で持っているか推定することで、文書を分類することが求められるため、各ツイートにおけるトピック構成比率を求めることができ、各ツイートをトピックごとの成分に情報圧縮できるLDAを用いたトピック分析を行う。また、求められたトピックをトライブとする。

③類似トピック結合

LDAを適用した結果から各トピックに分類された単語集合を入力し、各トピックの単語集合に対して、TF-IDFコサイン類似度を利用したクラスタリングを行う。

(4)各研究目的に対する分析

① 研究目的①

図1の通り、目的変数をエンゲージメント(いいね数+RT数)、説明変数をフォロワー数、文字数、PN値、ハッシュタグの有無、URLの有無、ツイートタイムとし、各トライブで重回帰分析を行う。

② 研究目的②

図2の通り、目的変数をエンゲージメントまたはPN値、説明変数を各ツイートにおける各トピックの構成比率とし、重回帰分析を行う。

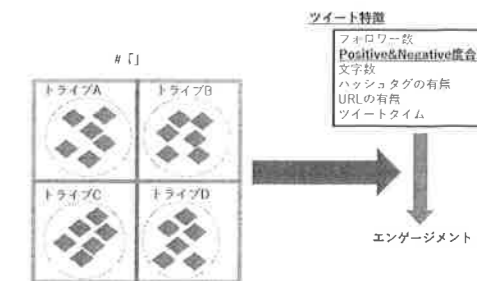


図1.研究概要図①

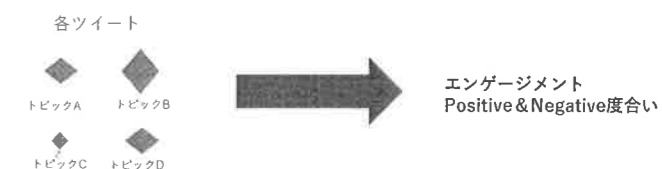


図2.研究概要図②

### 3.2 まとめ

ツイート文全体を考慮しながら消費者の価値観を捉えることのできるトライブ分類の方法を確立することで、従来研究より洗練されたトライブ抽出方法を提案する。トライブごとに「ツイートの特徴とエンゲージメントとの関係性」が異なることを示し、トライブ分類の学術的な有効性を明らかにすることで、企業側がトライブマーケティングを行う際の後押しとなる。また、消費者のエンゲージメント&ポジティブ投稿をより多く集めるトライブの特徴を定量的に示し、注力すべきトライブの決定方法を確立することで、企業側がどのトライブにアプローチするか決める際の判断基準となる。

### 参考文献

- [1]守口剛、上田雅夫、奥瀬喜之、鶴見裕之「消費者行動の実証研究」中央経済社(2018)
- [2]杉浦裕文「SNSのエンゲージメント向上要因に受信者のトライブが与える影響」早稲田大学創造理工学研究所 経営デザイン専攻 修士論文予稿集 pp.81-84(2019)
- [3] Zuo et al., "The Geography and Politics of News-Sharing Communities in Twitter" ,An Exploration of Social Identity (2012)
- [4] 高村大也, 乾孝司, 奥村学 「スピンモデルによる単語の感情極性抽出」, 情報処理学会論文誌ジャーナル, Vol.47 No.02 pp. 627--637(2006).