

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

プロフィットエンジニアリング研究

5218F017-2 杉浦裕文
指導教員 大野高裕

The Influences of Tribes on the Factors that Encourage the Social Media Engagement

SUGIURA Hirofumi

1. 研究背景と目的

1.1. SNS マーケティング

近年、企業が商品やサービスの認知向上や販売促進、宣伝を目的に利用している SNS マーケティングでは、KPI としてエンゲージメントの総数や、アカウントのフォロワー数で除算したエンゲージメント率が利用されている[1]。エンゲージメントとは実際に投稿を見た人が行う好意的なリアクションの総称のことで、Twitter でいうと「いいね」やリツイートなどが該当する。企業はエンゲージメント獲得のために SNS 上のキャンペーンなど様々な工夫をしており、従来研究から、画像や URL の利用はエンゲージメント向上要因として広く知られている。またその一環として SNS 上でセグメンテーションも盛んに行われている。フォロワーの繋がりを意識したコミュニティを利用したインフルエンサーマーケティングなどがこれに当たる。

1.2. トライブの利用

SNS からは投稿内容と投稿者の属性、両方の情報収集が可能なので、それらを用いた顧客セグメンテーションが積極的に行われている。そのアプローチは 2 つに大別される。1 つ目は投稿者のフォロワー・フォロワーの関係から作成するもので、一般的に「コミュニティ」と呼ばれ、学術的研究ではこれを対象としたものが多い。2 つ目は、投稿のコンテンツから共通の興味関心やライフスタイルを持つ人を抽出する方法であり、これは「トライブ」と呼ばれる。近年、多様化した顧客を捉えるために、トライブに焦点を当てたトライブマーケティングが実務上、活発に行われている[2]。

しかしトライブを利用している企業が増えているものの、トライブの定義は統一されていない。また作成手順も各企業のノウハウや自社サービスを通して行うものが多く、一般化されていない。つまりモデリングや分析も経験則的に進められているのが現状である。

1.3. 本研究の目的

1.1~1.2 で、SNS マーケティングでは KPI としてエンゲージメントを利用していること、およびエンゲージメント獲得のための要因としてトライブというセグメントが近年利用され始めていることを述べた。

しかしトライブの形成と活用は実務主体で進められている概念であるため、トライブの作成手順や、現在利用されている画像などのエンゲージメント向上要因に与える影響への学術的な検討はなされていないのが現状である。

そこで、本研究の目的を以下のように 2 つ設定した。

1 つ目は、企業ごとで異なるため明らかにされていないトライブの作成手順について、実務で利用されている特徴

を損なわないことを意識しつつ、学術的な手法を用いた再現性の高いトライブ作成手順の提案を行うことである。

2 つ目は、作成されたトライブを用いて、所属するトライブの違いが、エンゲージメント向上要因に与える影響をモデル化することである。また両者ともに日本で利用者数の最も多い Twitter を対象とする。

1 つ目の目的について、本研究では、トライブマーケティングを行っている企業調査を基にして、トライブを「共通の興味関心やライフスタイルを持った集団」として定義することし、この定義を基に従来研究や提案の検討を進めた。

2. 従来研究

2.1. セグメンテーションについて

トライブに近い概念であるセグメンテーションに関する研究において、Patrick et al.[3]は「Pain」など痛みに関する言葉が含まれる投稿を抽出し、投稿を地域別に分けることで、地域ごとに政治や病気など痛みの特徴が異なること、ポジティブ・ネガティブの感情が異なることを明らかにした。作成されたセグメントの内容は、結論としてはエンゲージメントに繋げるものではないが、トライブに非常に近いものであるといえる。セグメンテーションの手法は、単語の共起関係により作成されたネットワークグラフから、よく共起している、つまり類似した単語のグループを抽出するものである。本研究でもこの手法を参考にトライブを作成する。

2.2. エンゲージメント獲得のための要因

エンゲージメント研究としては、機械学習を用いたリツイート予測に関するものと、統計手法を用いたリツイート要因の推定に関するものが中心的であり、本研究では統計手法の研究に着目して知見を得た。

Stefan et al. [4]は、政治的な文脈を持つ投稿を対象に、フォロワー数や URL といったリツイート向上要因に加え、ポジティブ・ネガティブ感情がエンゲージメントに影響を与えるのか負の二項回帰を用いて分析し、結果として両感情ともにリツイートに正の影響を与え、特にネガティブ感情が強い影響をもつことを明らかにした。

次に津川[5]は、フォロワー関係から作成されたコミュニティに着目し、投稿の伝播の段階で異なるコミュニティに到達すること、またそのコミュニティの大きさやエッジの数がリツイートに影響するのかを、Stefan et al. [4]と同様に負の二項回帰を用いて分析した。結論として、URL などの他のリツイート向上要因と比較しても、他のコミュニティに到達することはリツイートに強い正の影響を与えることを明らかにした。

このようにリツイート研究の統計的な手法として、負の二項回帰が一般的であるが、変数として投稿者が操作できる範囲でないものや、あくまでコミュニティを利用したもので、トライブといった興味関心ライフスタイルに着目したものは見当たらない。

3. 本研究の提案

本研究では、1.データの取得、2.トライブの作成、3.影響要因モデルの構築という3つの段階に従って検討する。

3.1. データの取得

対象商材として、利用シーンやライフスタイルに特徴が必要なため高関与商材、加えてトライブの特徴だけでは表現しきれないアイテム・カテゴリー・ブランド、それぞれの商材の分類レベルの差も考慮するために、「カメラ」「一眼レフ」「ハーレーダビッドソン」の3つの商材を取り上げる。本研究ではTwitter APIを用いて2019年8月1日～31日の期間に各商材の名称を含む投稿と、そこに紐づく投稿や投稿者の情報を取得した。また投稿してからの経過時間によるエンゲージメントの差をなくすために、全て3日前の投稿を取得するようにした。期間内の投稿数は、「カメラ」は1,062,839投稿、「一眼レフ」は26,229投稿、「ハーレーダビッドソン」は3,035投稿となった。

3.2. トライブの作成

まず作成手法としては前述の通り、作成されたセグメントがキーワードを中心に似た意味の単語群でまとめられているPatrick et al.[3]の研究を参考に行う。また実務で扱うトライブとの違いが生じないかなどを、実際にトライブマーケティングを行っている企業の専門家にヒアリングしながら、トライブの作成フローを下記の通り構築した。

Step1：文字データの加工・修正

文字データの加工について、絵文字や文字化けといったノイズや空白を削除し、アルファベットを全て小文字に統一した。その後、類似・同一投稿、Twitter上の会話であるリプライ、フォロワー0人の投稿の削除をした。

Step2：形態素解析(本研究では名詞のみ抽出)

形態素解析により対象とする品詞を抽出する。本研究では、対象とする商材をどのようなシーンで利用するか、どのような生活の人が利用するかを基準にトライブを作成するため、名詞とサ変名詞のみを対象とした。

Step3：頻出語の把握・修正

Step2で得られた単語を頻出度順に集計し、どのような単語が使われているかを把握する。この段階で「笑」などの意味のないノイズとなる単語や、日本語の場合よく起こる「T+シャツ=Tシャツ」のような複合語を修正していく。

Step4：共起ネットワークグラフの作成

各単語に対してJaccard係数を算出することで単語間の共起関係を計算し、共起ネットワークを作成する。

Step5:Modularityが最大となるように単語をグループ化
コミュニティ分割の精度を表す指標であるModularityが最大になるようにエッジを切り分けることで、単語をグループ化するという流れになる。

これらのステップは、KH Coder[6]というテキスト型データを統計的に分析するためのフリーソフトウェアを利用した。

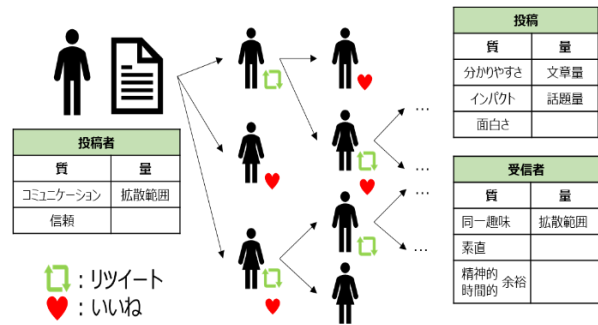


図1. エンゲージメントへの影響要因の概念図

表1. 影響要因の項目検討

投稿者	投稿	受信者	環境
フォロワー数	文字数	フォロワー数	時間
フォロー数	画像	フォロー数	曜日
過去投稿数	ハッシュタグ	過去投稿数	季節
過去いいね数	URL	過去いいね数	居住地
プロフィール充実	動画	プロフィール充実	
公式	コンテンツ	公式	
アカウント作成日		アカウント作成日	

3.3. 影響要因モデルの構築

まず変数の選択として、エンゲージメントへの影響として考えられる要因を網羅的に検討した。具体的には、Twitterの空間を投稿者・受信者・投稿の3つの視点に分けて考え、それぞれの量と質の観点からエンゲージメント向上に寄与する要因を概念的に抽出した。その結果を含んだ伝播の概念図が図1である。一般ユーザーは、投稿者と受信者の両方の立ち位置になり得るので、抽出された要因の中に「素直」などの客観的なものだけでなく、「投稿者」と「受信者に対して」といった対人的な意味を含んでいるものもある。

3つの視点の量と質の観点から抽出した要因について、実際にTwitter APIを利用して取得できる項目を表1に示す。また、抽象度の高い要因は「環境」として表現している。本研究ではトライブの直接的な影響でなく、エンゲージメント向上要因に与える影響に着目しているため、まず投稿者が投稿ごとに改変できる項目かどうかという点を評価した。次に定量的に測定可能な項目か、従来研究で明らかにされている項目と乖離がないかなどの検討を進め、最終的に表1の太文字であるフォロワー数・文字数・画像があるか否か・ハッシュタグがあるか否か・URLがあるか否か・休日か否か、の6項目を選択した。

表2. データセットの例

Eg	LEN	Media	Hash	URL	FW	Holi	tribe
151	1.5452	0	1	1	-0.066	1	1
0	0.9729	1	0	0	-0.14	0	2
304	-0.853	1	0	0	4.5078	1	3
0	-0.717	0	0	0	-0.076	1	4
0	1.0002	0	0	1	-0.167	0	5
12	0.1826	1	1	0	-0.153	0	6
8	1.2454	0	0	0	0.5874	0	7
0	-0.362	0	1	1	-0.17	1	8

※列の名称は後述する説明変数の略称に準ずる

変数の数値データについて、まず文字数とフォロワー数を正規化した。次に作成したトライブに所属する投稿に、そのトライブの番号をラベリングした。また無所属の投稿には抽出されたトライブの個数+1、つまり無所属のグループとなる番号をラベリングした。データセットの例を表2に示す。また最終的な分析に用いたサンプル数は、「カメラ」はデータ数が大きすぎたためランダムサンプリングを行った100,000投稿、「一眼レフ」は16,762投稿、「ハーレーダビッドソン」は1,372投稿となった。

本研究のモデル構造について、トライブをエンゲージメントに直接影響する変数としてではなく、そのトライブのターゲットに向けた投稿にはどの要因を用いることが効果的であるか、に着目している。したがって、目的変数をリツイート数といいね数の合計で表現するエンゲージメント、説明変数をリツイートへの影響要因とし、所属トライブによる階層構造を組み込んだ階層ベイズモデルを構築した。目的変数が負の二項分布に従うと仮定すると、本研究のモデル式は以下の式のように表せる。

$$Eg_i \sim NB(\mu_i, \sigma)$$

$$\mu_i = \beta_i^t X_i$$

$$\beta_i = \{\beta_i^1[tribe[k]], \beta_i^2[tribe[k]], \dots, \beta_i^7[tribe[k]]\}^t$$

$$X_i = \{1, LEN_i, Media_i, Hash_i, URL_i, FW_i, Holi_i\}^t$$

$$\beta^j[tribe[k]] \sim Normal(\beta^j[\text{全体平均}], \sigma_{\beta j})$$

- Eg_i : 投稿*i*のエンゲージメント
- $\beta_i^j[tribe[k]]$: 投稿*i*の所属するトライブ*k*における、*j*番目の説明変数に対応するパラメータ
- LEN : 文字数
- $Media$: 画像・動画の有無
- $Hash$: ハッシュタグの有無
- URL : URLの有無
- FW : 投稿者のフォロワー数
- $Holi$: 投稿日が休日か否か

各パラメータの事後分布の推定にはハミルトニアンモンテカルロ法を用いた。Chain数は4、各サンプリング回数は8,000回とし、各Chainにおいて最初の3,000回を破棄し、残りの5,000個の標本を用いた。

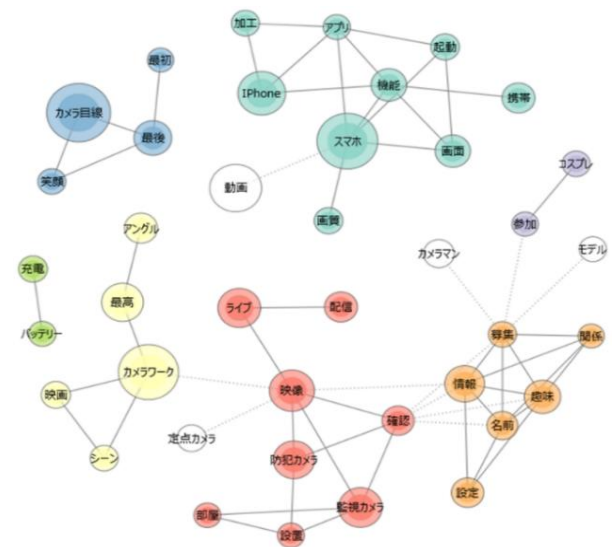


図2. 「カメラ」の共起ネットワーク

表3. 「カメラ」の各トライブの詳細

トライブ	内容	人数
トライブ1	スマホカメラトライブ	9047
トライブ2	画角意識トライブ	6092
トライブ3	コスプレトライブ	1117
トライブ4	固定カメラトライブ	7380
トライブ5	〇〇ファントライブ	5547
トライブ6	友人募集トライブ	2886
トライブ7	充電意識トライブ	1150
トライブ8	無所属	66781

4. 分析結果

4.1. 作成されたトライブ

まず作成されたトライブについて、「カメラ」の共起ネットワークとModularityによるグルーピング結果を図2に、また各トライブを構成する単語から判断したトライブの名称と所属人数を表3に示す。同様にして「一眼レフ」と「ハーレーダビッドソン」のトライブも抽出されたが、販売や入荷といったように明らかに広告である単語から形成されたトライブは除去した。その結果、トライブ無所属のものを含めると「一眼レフ」では8つ、「ハーレーダビッドソン」では9つのトライブが抽出された。

表4. 「カメラ」のパラメータ推定結果

顧客属性 変数	説明変数						
	定数項	文字数	画像	ハッシュタグ	URL	フォロー数	休日
全体平均	0.831*	0.643*	2.233*	-0.003	0.957*	3.704*	-0.094
Tribe1	-0.080	0.178*	0.588*	-0.454*	-0.232	-2.423*	-0.073
Tribe2	0.036	0.132*	-0.083	0.353	-0.137	-0.062	-0.036
Tribe3	0.013	-0.036	-0.056	0.040	0.334*	-1.324	0.051
Tribe4	0.231*	0.130	-0.205	-0.229	-0.050	0.638	-0.099
Tribe5	0.209	-0.026	0.342*	0.016	0.263*	2.796*	0.054
Tribe6	-0.288*	-0.249	-0.350	0.543*	0.069	-1.547	0.117
Tribe7	-0.300*	-0.156	-0.373	-0.274	-0.183	1.224	0.053
Tribe8	0.201*	0.021	0.153	0.037	-0.084	0.702	-0.067

表5. 「一眼レフ」のパラメータ推定結果

顧客属性 変数	説明変数						
	定数項	文字数	画像	ハッシュタグ	URL	フォロー数	休日
全体平均	0.783*	0.448*	1.686*	0.157	0.978*	1.325*	-0.031
Tribe1	-0.171	-0.067	-0.218	0.062	-0.302	-0.110	0.268
Tribe2	-0.075	0.082	0.016	-0.109	-0.302	-0.052	-0.217
Tribe3	0.089	-0.087	-0.135	0.064	0.150	0.191	0.028
Tribe4	-0.133	-0.027	0.658*	-0.014	-0.070	-0.051	0.468*
Tribe5	-0.013	-0.060	-0.028	0.144	-0.106	-0.386	0.059
Tribe6	-0.027	-0.093	-0.038	0.138	0.023	-0.091	0.263
Tribe7	0.011	0.171	-0.266	-0.100	1.071*	0.426	-0.814*
Tribe8	0.311*	0.077	-0.006	-0.185	-0.473	0.099	-0.031

表6. 「ハーレーダビッドソン」のパラメータ推定結果

顧客属性 変数	説明変数						
	定数項	文字数	画像	ハッシュタグ	URL	フォロー数	休日
全体平均	1.216*	0.350*	0.774	0.089	1.072*	0.377	0.371
Tribe1	0.019	0.180	-0.357	0.308	-1.068*	0.189	-0.092
Tribe2	-0.224	0.006	0.490	0.275	0.568	-0.003	-0.311
Tribe3	-0.102	-0.228	-0.180	-0.026	0.937*	0.112	-0.287
Tribe4	0.401	-0.263	0.961	-0.317	-0.059	0.200	-0.427
Tribe5	-0.032	-0.121	-0.967	0.545	0.133	-0.688*	-0.06
Tribe6	0.077	0.009	0.458	-0.264	-0.164	-0.128	0.657
Tribe7	-0.099	0.062	-0.977	-0.402	0.077	0.064	0.219
Tribe8	0.017	0.125	-0.627	0.027	0.277	-0.349	0.413
Tribe9	-0.059	0.218	1.107*	-0.190	-0.690*	0.571*	-0.134

4.2. パラメータの推定結果

表 4~6 は各商材の推定された EAP(事後期待値)を示している。「全体平均」の行はトライブを考慮しないエンゲージメントに対する各説明変数のパラメータであり、それより下の行は、1 行目の「全体平均」のパラメータに対するトライブごとの差分を表現している。HPD90%区間に 0 を含まないパラメータは太字で表記しており、HPD95%区間に 0 を含まないパラメータは*マークをつけている。また収束診断スコアである $\hat{R}$ の値も全てのパラメータにおいて 1.1 以下となったため、定常分布に収束されたと判断できる結果となった。

各説明変数のパラメータはハッシュタグ・休日は棄却され、「ハーレーダビッドソン」の結果のみ画像・フォロワー数も棄却された。また各商材の結果より、商材の分類レベルが小さくなるにつれてトライブによる「全体平均」との差が有意になりづらくなっていることが分かる。トライブによる「全体平均」との差については、正の影響や負の影響など、トライブの特徴によって異なる結果となった。

5. 考察

5.1. 「全体平均」のパラメータ

棄却された説明変数に関して、まず「ハッシュタグ」は 2 つの理由が考えられる。1 つ目は、Twitter は 140 字という文字数制限があり、内容はある程度まとめられているものだと考えられるため、要約の意味を持つハッシュタグはあまり効果がなかったという点である。2 つ目はインスタグラムが好調のため、受信者がハッシュタグそのものに物珍しさを感じなくなった点である。次に「休日か否か」については、Twitter の投稿は前述した文字数制限のため短文のものが多く、隙間時間で見るができる。そのため平日でも移動時間や退社後に見ていると考えられる。

また「ハーレーダビッドソン」のキーワードでは上記に加えて、画像とフォロワー数も有意とならなかった。その理由としては、他の商材の結果と比較した際にサンプルサイズが小さかった点や、ユーザーの年齢や特徴の観点から画像や投稿者という部分で反応しづらい点が挙げられる。

5.2. 商材のレベルと「全体平均」の差の関係

商材のレベルが小さくなるにつれてトライブの影響が見えづらくなった理由として、商材のレベルが小さいほど似た傾向の人が集まりやすくなる可能性が考えられる。本研究でいうと「カメラ」のトライブの場合、スマホ関連のものと防犯カメラ関連のものがあ、これは大きく特徴が違うことが分かる。しかし「ハーレーダビッドソン」のトライブの中には、多数モデル興味のものや電動自転車興味のものがある。これは広義な「モデル」について興味があるという点では共通の部分だといえる。そのため、トライブの分類上は異なっている、根底で共通している部分があるため差が見えづらくなったと考えられる。

5.3. トライブの特徴と「全体平均」の差の関係

トライブごとの全体平均の差について、カメラの結果を一部抜粋して考察する。まず正の影響が出た変数について、トライブ 5 の「○○ファンクラブ」では画像・URL・フォロワー数の差が有意となった。これは自分の推しの写真

などによる視覚効果が強いものや、イベントの紹介などの広告的な側面が上手く作用していることが分かる。フォロワー数もいわゆる強いファンというような人の方がファン同士でも信頼性が高いためということが推察される。

次に負の影響が出た変数について、トライブ 6 の「友人募集クラブ」では文字数・画像・フォロワー数が有意となった。特徴として、このクラブではハッシュタグのパラメータが正で有意になっている。つまり友人募集では「趣味: #カメラ」といった簡易的な説明を複数つけることで、繋がり幅を広げようとする傾向が見える。そのため「文字数」や「画像」といった詳細な説明の補助となるものが負となり、「ハッシュタグ」が強く正になったと推察される。

つまり両方ともトライブに所属する人の特徴を反映して影響の正負が反映されていることが分かる。

6. 結論と今後の課題

本研究の目的である、実務の要素を損なわず再現性のあるトライブの作成手順の提案と、トライブによるエンゲージメント向上要因への影響分析の、2 つとも明らかにすることができた。

この結果は、これからトライブマーケティングを行う企業への参入障壁を下げたこと、トライブに対する広告の投稿指針、商材のレベルでトライブの見えにくさが生じたことから、企業によってどの母集団を対象にトライブの作成を行うべきかなどの示唆が得られた。

今後の課題として、投稿のコンテンツ力や投稿者の性格などの感情的な部分は考慮できていない点やトライブの作成手順が確立されていない点、低関与商材の場合に同様の結果が得られるか不明な点などが挙げられる。

参考文献

- [1]株式会社日経 BP「平成 27 年度経済産業省委託事業商取引適正化・製品安全に係る事業報告書」, http://www.meti.go.jp/policy/economy/consumer/consumer/pdf/sns_report.pdf, (2019 年 9 月 29 日閲覧)
- [2]日経 MJ「キリンレモン、ネット CM はじける、コア層ごとに、動画 50 種類、「クラブ」マーケで若返り、女性つかみ販売量 2 倍。」, (2019 年 3 月 8 日)
- [3]Patrick J Tighe, Ryan C Goldsmith, Michael Gravenstein, H Russell Bernard, Roger B Fillingim「The Painful Tweet: Text, Sentiment, and Community Structure Analyses of Tweets Pertaining to Pain」, JOURNAL OF MEDICAL INTERNET RESEARCH, vol.17, iss.4, e84, pp.1-19 (2015)
- [4]Stefan Stieglitz, Linh Dang-Xuan「Emotions and Information Diffusion in Social Media-Sentiment of Microblogs and Sharing Behavior」, Journal of Management Information Systems, Vol.29, No.4, pp.217-247 (2013)
- [5]津川翔「ソーシャルネットワークのコミュニティ構造がソーシャルメディア上の投稿の拡散規模に与える影響の分析」, The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, pp1-4 (2018)
- [6]KH Coder, <https://kncoder.net/>