

投稿内容に着目した LDA によるトライブ抽出法の提案 ～ Twitter を用いたトライブマーケティング～

プロフィットエンジニアリング研究

5220F020-3 寺畑勇希
指導教員 大野高裕

Tribe Extraction Method by LDA Focusing on Posted Content ～Tribe Marketing Using Twitter～

TERAHATA Yuki

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重要な課題である。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、生活必需品や耐久消費財が普及し、満ち足りたモノ主体の消費から、精神的、感覚的な欲求や趣味嗜好のニーズを満たそうとするコト消費志向が増大したことによって、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。これに対し企業は、従来、性別など人口動態変数や居住地域など地理的変数といったデモグラフィック情報に基づくセグメンテーションにより対応してきた。しかし、コト消費志向が増大するにつれて、一人の人間が場面ごとに複数の異なる顔(価値観や趣味嗜好)を持つようになったため、従来のセグメンテーションでは対応しきれなくなっている。そのため、個人が場面ごとにどのような志向やニーズを持つかを反映したセグメンテーションが求められる。Twitter など SNS から発信される情報は、こうした場面ごとの個人のニーズや価値観、感情などが含まれているため、これらのデータ分析を通して、新たなセグメンテーションの可能性が期待できる。

以上の背景から、SNS による投稿内容を用いて、多様化する”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の一側面を一つの集団(トライブ)として抽出する意義が指摘できる。こうしたトライブに着目することで、各トライブに対して、的確なアプローチを行うこと(トライブマーケティング)が可能になるからである。

過去のトライブマーケティングの成功事例として、キリンレモン(選定トライブ:アーティスト好き)やウィダーinゼリー(選定トライブ:部活動熱心層)が挙げられる。キリンレモンに関するトライブマーケティングにおいて、対象トライブをアーティスト好きに選定したのは、アイドルや歌手などのアーティストに最も興味を示す 20 代に向けて販促したかったからである。この背景には、”幼少期にキリンレモンを飲んでいて 20 代に再び手にとってもらえるブランドにしたい”という企業側の思惑が隠れている。しかし、このとき、20 代はアイドルなどのアーティストに興味を抱きやすいだろうという経験則で、企業側はトライブ選定を行っている。本来ならば、現在のキリンレモン購買者の中には、どのようなトライブが存在し、各トライブはどのような特徴を持っているのかという現状把握を基に、対象トライブを選定した方が、トライブマーケティング

の成功確率が上がると考えられる。そのため、現状として、トライブ推定に確立された方法論が提示されていないことが課題として挙げられる。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 に着目し、Twitter データに基づくトライブ推定方法の提案を行い、また、実務に寄り添ったとき、よりエンゲージメント向上に繋がるアプローチすべきトライブの選定方法を確立しなければならない。

なお、マクドナルド元 CMO である足立は、「消費者の心を動かしたり、買いたいと思わせたりするクチコミが増えるほど売上に貢献しやすい」と述べているため、Twitter API から取得できるフォロワー数などのツイートの形式的な特徴だけでなく、ツイート内容面での評価も必要となる。また、守口ら[1]は、「受動的なツイートでは効果が少なく、ポジティブなクチ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。よって、本研究ではツイート文章のポジティブ度合いをツイート内容面での評価として用いる。あわせて、商品特性によっても、消費者の利用背景や利用目的が大きく異なるため、トライブ分類結果の傾向やエンゲージメントに繋がるトライブの特定方法に違いが生じると考えられ、これらも考慮する必要がある。

そこで、本研究の目的は、①Twitter の投稿内容からトライブを抽出する一般的な方法と②ツイートの形式的特徴と内容面評価を反映したよりエンゲージメントに繋がるトライブ特定方法を、商品特性ごとに提案することである。本研究を通して、学術的に裏付けがありかつ実用的なトライブマーケティングの方法論を提示できると考える。

2. 従来研究

2.1. Twitter に関する従来研究

Twitter に関する従来研究では、エンゲージメントの予測やフォロワーフォロワー関係に基づくコミュニティ分析を扱った研究が大半を占めているが、消費者の興味関心に着目した研究として、Zuo et al [2]が挙げられる。この研究では、New York Times 記事について言及しているツイートを収集し、記事の URL から、ユーザーの関心を分類後、ユーザーの属性や位置情報を抽出することで、どのようなユーザーがどういった話題に言及するかをネットワーク図で可視化し、傾向を分析している。この研究は、URL からユーザーの興味関心を抽出するという試みが行われているが、最終的にユーザーの属性や位置情報をもとにグループの特徴を捉えているので、従来のセグメンテーションとさほど変わらない分類になってしまっている。

2.2. トライブマーケティングに関する従来研究

これに対し、杉浦[3]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、①トライブ分類の正確性、②対象商品の選定、③トライブマーケティングに対する実務的貢献という 3 つの観点で課題があると考えられる。

まず①に関して、トライブ分類を、共起ネットワーク上の頻出語が各ツイート内に含まれるか否かでトライブへの所属を決定しているため、ツイート文全体を考慮した分類とは言い難い。②に関して、カメラなど的高関与商品のみを分析対象としているため、低関与商品への適用が課題となる。③に関して、この研究ではトライブ分類の有効性が示されているが、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を行うことで、どのトライブにアプローチすべきかという具体的な提案手法を提示することができると考えられる。

3. 研究内容

3.1. 研究の前提

3.1.1. コミュニティとトライブの違い

Twitter 関連の従来研究におけるコミュニティ分類で扱われる“コミュニティ”はフォローフォロワー関係で区切られた集団である一方、“トライブ”は、こうした関係ではなく、価値観や趣味嗜好で区切られた集団である。

3.1.2. トライブの前提

3.1.1 の通り、トライブはコミュニティと比較して、“ゆるい繋がり”を持った集団であるため、一人の人間が複数のトライブに所属することが多いと考えられる。よって、一人の人間が場面ごとに持つ異なる複数の顔を把握する必要性から、一人の人間が複数トライブに所属する場合、複数種類の人間として置き換えて各分析を行う。

3.2. 研究方法

研究手順に関しては、以下の 5 ステップで行う。

Step1. ツイートデータ収集

Twitter API を用いた Python コードを作成し、特定商品に関する投稿を収集する。

Step2. 各ツイートの Positive/Negative 度(PN 値)算出

形態素解析によって分割した各ツイートの各単語の PN 値(Positive/Negative 度合い)の平均値をツイート全体の PN 値とする。(高村ら[4]が作成した単語感情極性対応表に基づき算出する。)

Step3. トライブ抽出

精度の高いトライブを抽出するため以下 2 段階で行う。

① ストップワードの抽出&除外

ストップワード辞書に掲載されている言葉と複数トピックに出現する言葉をストップワードとして除外する。

② LDA 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にどのようなトピックが、どの程度含まれているかを推定することで、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算出し、かつ、各ツイートをトピックごと

の成分に情報圧縮できる Latent Dirichlet Allocation (LDA)を用いたトピック分析を採用する。これにより推定されたトピックをトライブとする。

Step4. トライブ分類の有効性分析(回帰分析)

① 各トライブのエンゲージメントに影響を与える要因

図 1 の通り、目的変数をエンゲージメント (EG = いいね数+RT 数)、説明変数をフォロワー数(x_1)、フォロワー数(x_2)、文字数(x_3)、ハッシュタグの有無(x_4)、URL の有無(x_5)、ツイートタイム(x_6)、PN 値(x_7)とし、(1)の回帰式をもとに、各トライブについて重回帰分析を行う。

$$EG = a_1 x_1 + a_2 x_2 + a_3 x_3 + a_4 x_4 + \dots + a_7 x_7 \quad (1)$$

この分析を通して、エンゲージメント向上におけるトライブ分類の有効性を示す。

② PN 値に繋がるトライブ内容(トピック)

図 2 の通り、目的変数を 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、(2)の回帰式をもとに、重回帰分析を行う。

$$PN = P_1 C_1 + P_2 C_2 + P_3 C_3 + P_4 C_4 + \dots + P_{10} C_{10} \quad (2)$$

C_i : 各ツイートにおけるトピック i の構成割合

この分析を通して、ツイート内容面での評価を行う。

①と②を通して、どのトライブにアプローチすべきかを判断するために必要な、ツイートの形式面・内容面という両面の特徴を捉えることができると考える。

Step5. トライブ特徴の詳細把握

トライブごとに再度 LDA によるトピック分析を行うことによって、各トライブの詳細な特徴を把握する。さらに、目的変数を 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、重回帰分析を行うことで、各トライブ内で PN 値向上により繋がるトピックを特定できる。

3.3. 対象商品

Twitter を用いたトライブマーケティングを行う際、対象商品によって、トライブの抽出パターンやトライブ分類の有効性の程度が異なってくると考える。よって、本研究では、図 3 の通り、①商品レベル/ブランドレベルと②高関与商品/低関与商品という 2 軸で商品分類を行う。

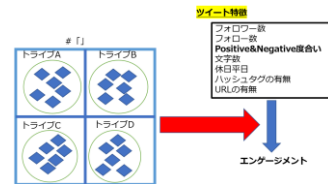


図 1. 研究概要図①



図 2. 研究概要図②

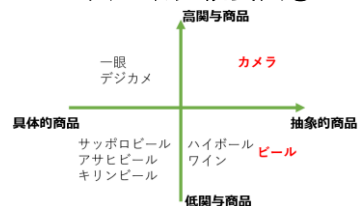


図 3. 対象商品分類図

4. 分析結果

4.1. 使用データ

本稿では、カメラとビールに関するツイートから推定した結果を述べる。Twitter API を用いてキーワードを、#カメラもしくはビールに選定し、#カメラに関しては、データ数 3602(2021/10/22~11/05)を収集し、ビールに関しては、データ数 196537(2021/10/22~11/06)を収集した。

4.2. トライブ分類(Step3)

LDA によるトピック分析の際のトピック数決定に関しては、①低ければ低いほど良いとされる Perplexity という指標と②各トピック(トライブ)の特徴が的確に把握できるようなワード分類になっているかといった 2 側面から決定した。

4.2.1. トライブ分類結果(#カメラ)

2 週間分のデータ(データ数: 3602)を用いて、LDA によるトピック分析を行った結果、表 1 のようなトライブが抽出された。主に、趣味嗜好/性能/利用歴でトライブが推定されているということが明らかになった。

4.2.2. トライブ分類結果(ビール)

2 週間分のデータ(データ数: 196537)を用いて、LDA によるトピック分析を行った結果、表 2 のようなトライブが抽出された。主に、購買シーン/購買時重視ポイントでトライブが推定されていることが明らかになった。

4.3. 各トライブ内での回帰分析結果(Step4)

Step4 における①と②に示した回帰分析を実行した。

4.3.1. 各トライブ内での回帰分析結果(#カメラ)

表 3 と表 4 に示す通り、各トライブでエンゲージメント向上に及ぼすツイートの特徴要因が大きく異なることが明らかになった。表 3 の通り、ツイートの特徴がネガティブ(PN 値が低い)であるほどエンゲージメント向上するのが、Tribe4 であり、ポジティブ(PN 値が高い)であるほどエンゲージメント向上するのが、Tribe9 である。一方、表 4 の通り、PN 値向上に最も影響するトライブ内容は Tribe9 であり、PN 値低下に最も影響するトライブ内容は Tribe4 である。以上の分析から、Tribe4 に関して、Tribe4 の内容量が増加すると PN 値が低下し、PN 値が低下するとエンゲージメントが向上してしまうことから、ツイートがネガティブであればあるほど、エンゲージメントを集めてしまうという炎上の危険性を抱えていることが読み取れる。逆に Tribe9 に関して、Tribe9 の内容量が増加すると PN 値が向上し、PN 値が向上するとエンゲージメントも向上することから、Tribe9 内容量とポジティブ度合いは相関関係にあり、ツイートがポジティブであればあるほどエンゲージメントをより集めるため、Tribe9 はアプローチすべきトライブであるといえる。

4.3.2. 各トライブ内での回帰分析結果(ビール)

表 5 と表 6 に示す通り、各トライブでエンゲージメント向上に及ぼすツイートの特徴要因が異なることが明らかになった。表 5 の通り、ネガティブ(PN 値が低い)であるほどエンゲージメントが向上するのが、Tribe4 である。一方、表 6 では、PN 値向上に最も影響するトライブ内容は Tribe4 である。以上より、Tribe4 に関して、Tribe4 の

内容量が増加すると PN 値が向上し、PN 値が向上するとエンゲージメントが低下してしまうことから、ツイートがポジティブであればあるほど、エンゲージメントが集まらないというマーケティング効率の悪いトライブであることが読み取れる。

表 1. トライブ分類結果(#カメラ)

トライブ1	花好き
トライブ2	性能重視
トライブ3	旅行好き(女性多め)
トライブ4	ペット好き
トライブ5	公園好き
トライブ6	一眼初心者
トライブ7	風景好き
トライブ8	モデル撮影好き
トライブ9	オンラインショップ
トライブ10	プロカメラマン

表 2. トライブ分類結果(ビール)

トライブ1	通販好きトライブ
トライブ2	ビールと合う手料理重視トライブ
トライブ3	ビールと合う外食トライブ
トライブ4	共感トライブ
トライブ5	質重視トライブ
トライブ6	飲むタイミング重視トライブ
トライブ7	缶ビール好きトライブ
トライブ8	帰宅途中ビールをよく買うトライブ
トライブ9	会話重視トライブ
トライブ10	多様なお酒好きトライブ

表 3. 回帰分析結果①(#カメラ)

トライブ	フォロワー数	フォロー数	文字数	ハッシュタグの有無	URLの有無	休日平日	PN
1	0.587***	0.255*	-0.005	0.433	1.384*	-0.195	-0.410
2	0.713***	0.251*	0.021***	0.652*	2.160	-0.057	-0.676
3	0.234**	0.238**	0.003	-1.856***	0.985*	0.043	0.699**
4	0.443***	1.087***	0.005	1.343*	2.743***	-0.722***	-1.015*
5	-0.346*	1.114***	0.019***	0.678	0.742	0.248	0.195
6	0.707***	-0.382*	0.011***	0.748	1.194**	-0.479**	-0.922*
7	0.745***	1.130***	-0.001	0.613	-0.130	0.0252	-0.350
8	0.228*	0.531***	0.007	1.673***	2.448***	0.0880	0.718
9	-2.012**	0.845**	0.052***	3.043	0.321	-0.037	0.930*
10	0.585***	-0.069	0.015*	-0.395	2.869***	-0.250	-0.789

表 4. 回帰分析②(#カメラ)

Topic1	Topic2	Topic3	Topic4	Topic5	Topic6	Topic7	Topic8	Topic9	Topic10
-0.772***	0.193	-0.914***	-1.888***	-0.972***	-0.899***	-0.168	-0.215	1.340***	-0.453

表 5. 回帰分析①(ビール)

トライブ	フォロワー数	フォロー数	文字数	ハッシュタグの有無	URLの有無	休日平日	PN
1	2.270***	-0.054**	0.011***	-0.286***	1.286***	0.077*	-0.002
2	1.270***	0.063***	0.012***	0.064	1.858***	-0.067**	-0.180**
3	1.612***	0.100***	0.013***	0.158**	1.543***	-0.015	-0.006
4	0.379***	0.063*	0.010***	-0.535***	1.717***	0.0222	-1.780***
5	1.082***	0.002	0.012***	0.246***	1.615***	-0.075*	0.311***
6	1.156***	0.122***	0.011***	0.382***	1.498***	-0.061	0.255***
7	1.306***	0.097***	0.013***	-0.034	1.072***	0.170***	0.132
8	0.745***	2.101***	0.011***	0.328***	1.181***	-0.388***	0.174
9	1.614***	0.074***	0.010***	0.150*	1.301***	0.0360	0.178*
10	0.137***	0.669***	0.009***	0.254***	1.477***	0.087*	0.135

表 6. 回帰分析②(ビール)

Topic1	Topic2	Topic3	Topic4	Topic5	Topic6	Topic7	Topic8	Topic9	Topic10
0.583**	1.955***	-2.568***	4.346***	-2.157***	-1.853***	-1.455***	-2.078***	-0.211***	-1.549***

表 7. Tribe4 詳細把握結果②(#カメラ)

Topic1	Topic2	Topic3	Topic4	Topic5
-3.073***	0.419	0.592	-1.831***	1.262***

表 8. Tribe4 詳細把握結果①(#カメラ)

Topic1	野良猫撮影
Topic2	子猫撮影(カメラ初心者)
Topic3	被写体重視
Topic4	暮らし撮影
Topic5	実家周辺撮影

表 9. Tribe4 詳細把握結果①(ビール)

Topic1	他の酒との比較
Topic2	キャンペーン
Topic3	飲酒楽しみ
Topic4	家族とお酒
Topic5	疲れた後のお酒
Topic6	コロナ禍での飲酒

表 10. Tribe4 詳細把握結果②(ビール)

Topic1	Topic2	Topic3	Topic4	Topic5	Topic6
-4.6772 ***	2.67352 ***	0.629193	3.48155 ***	-4.363 ***	-0.25237

4.4. トライブ特徴の詳細把握(Step5)

4.4.1. トライブ特徴の詳細把握(#カメラ)

抽出された 10 個のトライブそれぞれについて LDA によるトピック分析を行った。本予稿では、Tribe4 に関する詳細な分析結果のみ述べる。表 8 に示す通り、Tribe4(ペット好きトライブ)は、日常の暮らしや具体的な被写体によって、さらに 5 つのトピックに分類できた。表 7 の通り、Tribe4 において、最も PN 値低下に影響を与えるのは、Topic1 である。4.3.1 で述べた通り、Tribe4 は炎上の危険性を持ったトライブであり、Topic1 の内容が最大の炎上要因となることが読み取れる。

4.4.2. トライブ特徴の詳細把握(ビール)

抽出された 10 個のトライブそれぞれについて LDA によるトピック分析を行った。本予稿では、Tribe4 に関する詳細な分析結果のみ述べる。表 9 に示す通り、Tribe4(共感トライブ)は、他のお酒との比較や飲酒の感想によって、さらに 6 つのトピックに分類することができる。表 10 に示す通り、Tribe4 において、最も PN 値向上に影響を与えるのは Topic4 である。4.3.2 で述べた通り、Tribe4 は炎上の危険性を持ったトライブであり、Topic1 の内容が最大の炎上要因となるということがわかる。

5. 考察

4 章にて 3 章で提案した分析方法が①Twitter の投稿内容からトライブ抽出する方法と②エンゲージメントにより繋がるトライブの特定方法として有効であることが示された。この結果を踏まえ、①トライブ分類、②トライブ分類有効性分析(回帰分析)、③トライブ詳細把握という 3 点に関して、商品特性が与える影響を考察する。

5.1. トライブ分類の特徴

低関与商品であるビールは、「商品の購買シーンや使用シーン/類似商品との比較」によってトライブ分類されている一方、高関与商品であるカメラは、「趣味嗜好/商品の性能/利用歴」によってトライブ分類されている。以上より、高関与商品であればあるほど、趣味など利用シーンが多様化するため、それがトライブに表れていると考えられる。一方、低関与商品であればあるほど、他商品との比較

や他商品との関わり方に焦点が合いやすく、それがトライブとして表れている。

5.2. トライブ分類の有効性分析(回帰分析)

トライブごとにエンゲージメントに影響を与える要因を Step4 で分析したが、高関与商品の方が低関与商品よりトライブごとの違いが大きかった。これは、5.1 で述べた通り、高関与商品であればあるほど趣味などの利用シーンが多様化し、利用シーンによってその商材の価値や利用意義が異なるため、トライブごとのエンゲージメントに影響を与える要因がより異なってくると考えられる。一方、低関与商品においても、トライブごとにエンゲージメントに影響を与える要因は異なり、他商品との比較や関わりに関するトライブが多いため、別商品とのコラボレーションなどマーケティング戦略に活用できる点が多い。

5.3. トライブ特徴の詳細把握

抽出された各トライブ内で、LDA によるトピック分析を行った結果、低関与商品であるビールに関して、購買/利用シーンの深堀り/比較対象商品の選定といった観点で精緻化することができ、高関与商品に関しては、消費者の生活/趣味嗜好の具体化といった観点で精緻化を行うことができた。すなわち、消費者インサイトの深堀りという観点では高関与商品が勝り、他商品との比較という観点では低関与商品の方が勝ると考えられる。

6. 結論と今後の課題

結論として、3 章で提案した分析方法が、本研究の研究目的である①Twitter の投稿内容からトライブ抽出する方法と②エンゲージメントに繋がるトライブの特定方法として有効であることが示され、商品特性が上記の①と②にもたらす影響を明らかにすることができた。このような方法を用いたトライブ抽出と特定方法を通じて、企業側がトライブマーケティングを行う際のトライブの特定やアプローチ方法の選択に有効な手法の提案と知見が示されたと考えられる。

今後の課題として、①各ワードから算出された PN 値をもとに、ツイート全体の PN 値を算出しているため、文章から正確な PN 値を算出したとまではいかない、②ツイート発信者の性格/ツイートの意図までを把握することができない、③提案した方法論が実際のトライブマーケティング戦略としてどれほど通用するかといった検証が困難であれ難しいという 3 点が挙げられる。

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] Zuo et al.: “The Geography and Politics of News-Sharing Communities in Twitter”, An Exploration of Social Identity (2012)
- [3] 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 創造理工学研究科 経営デザイン専攻 修士論文予稿集 PP.81-84(2019)
- [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)