

## スマートフォンのレビューテキストマイニングによる消費者ニーズ分析 A Study on Customer Needs Analysis by Text Mining of the Review about Smartphone

加納 康大<sup>‡</sup>, 大槻 明  
Kodai Kano, Akira Otsuki

### 1. はじめに

本研究は計量マーケティングに関する実証研究である。具体的に、Amazonのスマートフォンに関する口コミレビューを取得し、売り上げランキングの上位と下位のレビューを星評価別に分類したうえ行列モデルを作成し、テキストマイニング及び共起ネットワーク分析のアプローチから分析を行った。さらに、レビューを価格帯ごと、星評価別に分類した行列モデルを作成したうえで同様のアプローチで分析を行い、これら二種類の分析結果から、売上や価格帯の観点から消費者ニーズの変化や共通点を明らかにし、商品開発やマーケティング戦略に役立つ知見の提案に繋げる。

### 2. 提案コンセプト

#### 2.1 分析対象データ及び分析モデル

分析対象をスマートフォン(以下「スマホ」として、2021/10/25~26の期間に、Amazonのスマホランキングを対象にレビューデータをスクレイピングすることで分析対象データを取得した。スマホを分析対象にした理由は次の通りである。

- 日常生活において、スマートフォンは欠かせない存在となりつつある。
- 今後さらに激化するだろうスマホ市場において、消費者ニーズの把握は重要度が高いと考えられる。

そして、本研究では、表1, 2に示す2通りの行列モデルで分析を行う。1つ目は、Amazon売り上げランキングの「上位」、「下位」と星評価の「上位(評価4, 5)」「下位(評価1, 2)」の4象限である。2つ目はAmazonサイト上での値段分類を参考に価格帯毎に分類するイメージで、「10001~50000(高い)」「~10000(安い)」と星評価の「上位(評価5)」「下位(評価1)」の4象限である。表1と表2で対象とする星評価の数が違う(例:表1の上位は評価4, 5だが、表2の上位は評価5のみ)が、理由は分析モデル1で十分なデータを得ることができなかったためである。

表1, 2は、取得したレビュー件数を上述の分析モデルに振り分けて件数を表示したものであるが、スクレイピングした10849件のうち、重複をレビューを除くと8393件となり、この中から表1, 2に該当するレビューを抽出した結果、分析対象データは4744件となった。

表1. 分析モデル1

|        | ランキング上位 | ランキング下位 |
|--------|---------|---------|
| 評価4, 5 | 481件    | 328件    |
| 評価1, 2 | 141件    | 114件    |

表2. 分析モデル2

|     | 10001~50000 (高い) | ~10000 (安い) |
|-----|------------------|-------------|
| 評価5 | 2952件            | 679件        |
| 評価1 | 800件             | 313件        |

### 2.2 TF-IDFを用いた特徴語の抽出

TF (Term Frequency) は、単語が1つの文書中にどれだけ出現するかを表すものであり、IDF (Inverted Document Frequency) は、全文書の中でその単語が出現する文書がどれだけあるかを表すものである。IDFは式1にて求められる。Nは文書数、dfは当該単語を含む文書の数、logは2を底とした対数を求める関数であることを意味する。

$$IDF = \log \frac{N}{df} \quad (1)$$

そして、これらを掛け合わせたものがSeki[1]の論文でも示されているTF-IDFである。例えば、似たような文書がいくつかあった場合に、TF-IDFを用いることによって、それらの文書の違いが何なのかを大まかに把握することが可能となる。本研究では、表1, 2の各象限ごとに特徴語を抽出することで、象限間の共通性や相違性を明らかにする。

### 2.3 共起ネットワーク分析

大槻ら[2]が提案した共起グラフネットワーク分析は、単語の共起頻度表を作成し、それをグラフデータに変換したうえで共起ネットワークマップ(以下「共起マップ」として可視化する分析手法である。この共起マップでは、次数の高さに比例して単語が大きく表示され、共起頻度の高さに比例してエッジが太く表示されるため、重要な単語及び共起語を見える化することが可能である。つまり、共起グラフネットワーク分析を用いて前節で抽出した各象限ごとの単語のつながり関係を追跡することで、各象限ごとの特徴語の詳細な解釈が可能になると考えられる。

### 3. 分析結果

#### 3.1 分析モデル1の分析結果

##### 3.1.1 TF-IDF分析による特徴語の抽出結果

分析モデル1のTF-IDF分析の結果を表3に示す。表3から下記①~④の特徴語が抽出できたため、次節で共起ネットワーク分析を用いてさらに詳細な解釈を行う。

1 次の単語は4象限全てに登場した特徴語

「購入」、「画面」

2 次の単語は評価5, 4のみに共通して登場した特徴語

「バッテリー」

3 次の単語は評価2, 1のみに共通して登場した特徴語

「起動」

4 次の単語はランキング下位でかつ評価5, 4のみに登場した特徴語

「カメラ」

表3. TF-IDF分析の結果(分析モデル1)

| 注1) | 上位/評価5,4 | 下位/評価5,4 | 上位/評価2,1 | 下位/評価2,1 |
|-----|----------|----------|----------|----------|
| 1   | バッテリー    | 画面       | 購入       | 画面       |

|   |     |       |    |    |
|---|-----|-------|----|----|
| 2 | 購入  | 使用    | 起動 | 購入 |
| 3 | 画面  | バッテリー | 画面 | 起動 |
| 4 | スマホ | 購入    | 問題 | 認証 |
| 5 | 問題  | カメラ   | 端末 | 指紋 |

注1) TF-IDFスコアの順位, 表4も同様

### 3.1.2共起ネットワーク分析結果

図1~4に, 分析モデル1の4象限全ての共起ネットワークマップを示す。

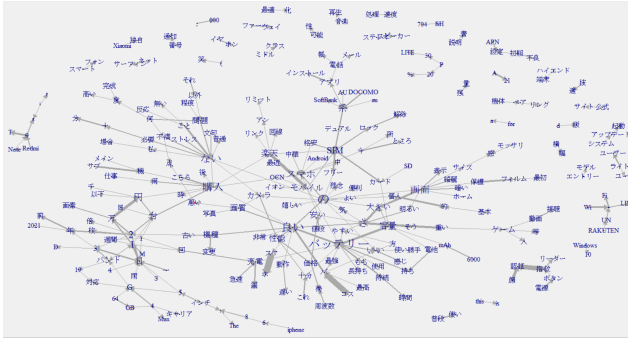


図1.「ランキング上位」—「評価5,4」

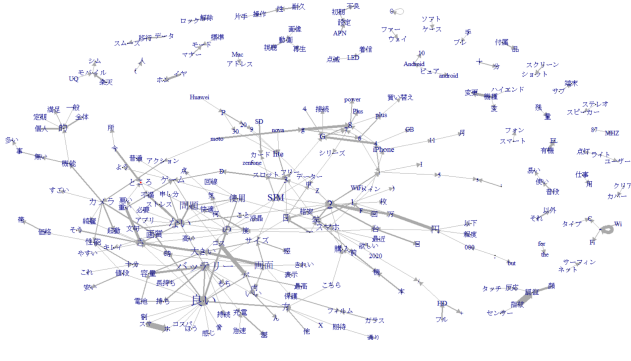


図2.「ランキング下位」—「評価5,4」

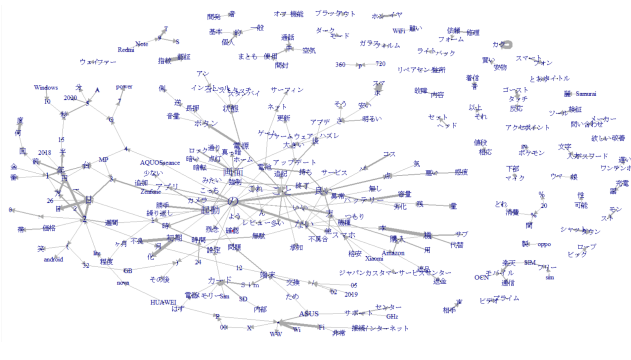


図3.「ランキング上位」—「評価2,1」

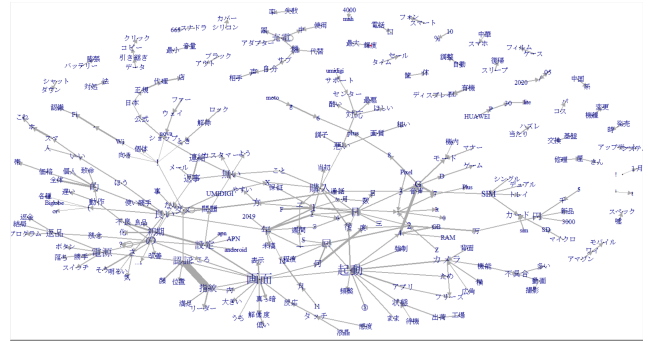


図4.「ランキング下位」— 評価2,1」

次に, 3.1.1節の①~④の特徴語について, 共起ネットワークマップを用いて解釈を行った結果について述べる。

#### 1 4象限全てに登場した特徴語「購入」, 「画面」

まず, 「購入」については, 次に示す特徴が見られた。

- ・ 上位/評価5, 4:メイン→機→購入, サブ→機→購入, 仕事→用→購入
- ・ 下位/評価5, 4:本→機→購入, 購入→2→台→目
- ・ 上位/評価2, 1:本→機→購入, 代替→機→購入, サブ→機→購入
- ・ 下位/評価2, 1:対応→悪い→購入

つまり, 本機としてだけでなく, 複数台所持を目的としたサブ機や, 仕事用で購入されていることが分かる。これらからを踏まえ, 今後の提案として, これら価格帯の商品で他端末とのデータのやり取りのスムーズ化を測り, 利便性をあげるためのサービス(メモ帳や保存内容の同期化など)を充実していくことが望ましいと考えられる。

次に, 「画面」の共起ネットワークを分析した結果, 次の特徴が明らかとなった。

- ・ 上位/評価5, 4:画面→暗い, 画面→明るい→さ, 画面→大きい
- ・ 下位/評価5, 4:画面→大きい, 画面→サイズ→大きい
- ・ 上位/評価2, 1:画面→真っ暗, 画面→暗い, 画面→大きい
- ・ 下位/評価2, 1:画面→真っ暗, 画面→大きい

特に, 評価2, 1で「暗い」や「真っ暗」といった共起語が見られたため, これらの単語について元のレビューデータを確認したところ, 直射日光下での画面の視認性に関する問題であった。ゆえに, 今後の改善点としては, さらに画面の明るさ上限の引き上げや, 周囲の明るさを感知し明るさを調節するセンサーを低価格帯にも実装することで消費者ニーズにあったスマホの提案ができるのではないかと考えられる。

#### 2 評価5, 4のみに共通して登場した特徴語「バッテリー」

特徴語「バッテリー」に関して, 評価5, 4を比較すると, 次に示す通り, 性能についてのポジティブな共起語が見られた。

- ・ 上位/評価5, 4:バッテリー→容量, バッテリー→持続→時間, バッテリー→良い, バッテリー→十分
- ・ 下位/評価5, 4:バッテリー→容量, バッテリー→長持ち, バッテリー→良い, バッテリー→十分

また共起語「十分」から, 現状の性能に満足した評価がなされていることがわかる。これらを踏まえると, バッテリー性能で今回抽出した商品性能と同等の性能を搭載することで消費者からは一定の評価を得ることが期待できると考えられる。

3 評価2, 1のみに共通して登場した特徴語「起動」

特徴語「起動」に関して、評価2, 1を比較すると、次に示す通り、性能についてのネガティブな共起語が見られた。

- 上位/評価2, 1: 起動→問題, 勝手→起動→画面
- 下位/評価2, 1: 起動→アプリ→フリーズ, 起動→カメラ→フリーズ, 強制→起動

これらを踏まえると、スマホのアプリ起動処理が上手くいっていないと考えられる。ゆえに、今後の改善点としてはメモリやCPUの性能向上により、処理の安定化を図るべきと考える。

4 ランキング下位でかつ評価5, 4のみに登場した特徴語「カメラ」

評価5, 4のランキング上位・下位を比較すると、下記に示す通り共に性能についてポジティブな共起語が見られた。

- 上位/評価5, 4: カメラ→性能→良い, カメラ→画質→良い
- 下位/評価5, 4: カメラ→機能→的→満足, カメラ→綺麗, カメラ→性能, カメラ→画質→良い

つまり、カメラ性能への評価は上位でも下位でも違いが無く、スマホのカメラ性能は頭打ちの状況を迎えているのではないかと考えられる。2021年に発売されたiPhone13 (価格111,672円)のカメラ画素数は1,200万画素、一方2021年に発売されたRedmi 9T (価格15,700円)のカメラ画素数は4,800万画素である。単純な画素数で見れば価格差が大きいにもかかわらず後者の方が性能が高い。また、現代では撮った写真はアプリで加工処理が容易に行える。加工を駆使することで誰でも簡単に、きれいな写真の作成が可能である。以上を踏まえると、カメラ性能はスマホ市場のニーズとしてこれ以上の性能向上は望まれている可能性が考えられる。

3.2 分析モデル2の分析結果

3.2.1TF-IDF分析による特徴語の抽出結果

分析モデル2のTF-IDF分析の結果を表3に示す。表4から下記⑤～⑧の特徴語が抽出できたため、次節で共起ネットワーク分析を用いてさらに詳細な解釈を行う。

- 5 次の単語は4象限全てに登場した特徴語  
「購入」
- 6 次の単語は「評価1」のみに共通して登場した特徴語  
「画面」「充電」
- 7 次の単語は「金額が高い」かつ「評価5」のみに共通して登場した特徴語  
「バッテリー」

表4. TF-IDF分析の結果(分析モデル2)

| ) | 高い/評価5 | 安い/評価5 | 高い/評価1 | 安い/評価1 |
|---|--------|--------|--------|--------|
| 1 | 購入     | 購入     | 購入     | 購入     |
| 2 | 問題     | 使用     | 画面     | 充電     |
| 3 | スマホ    | 問題     | 返品     | 画面     |
| 4 | 使用     | スマホ    | 起動     | 使用     |
| 5 | バッテリー  | 機種     | 充電     | アプリ    |

3.2.2共起ネットワーク分析結果

図5～8に、分析モデル2の4象限全ての共起ネットワークマップを示す。

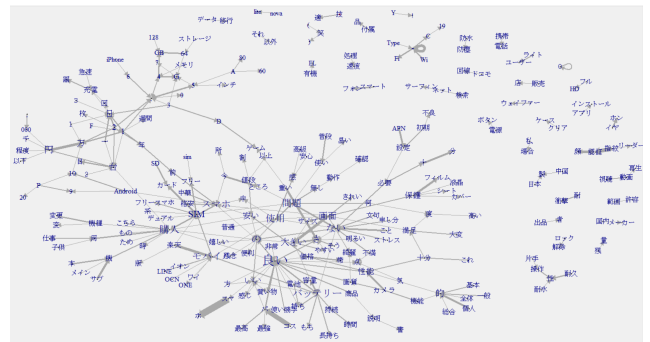


図5. 「金額が高い」ー「評価5」

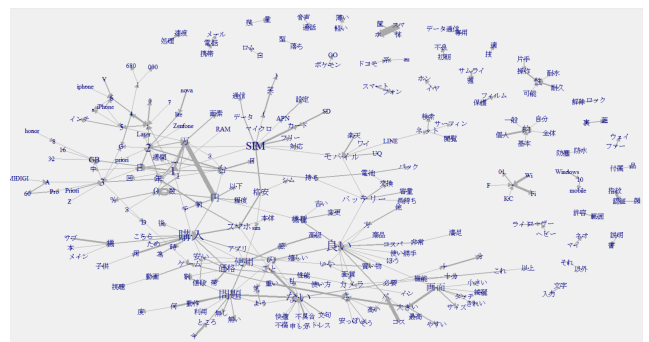


図6. 「金額が低い」ー「評価5」

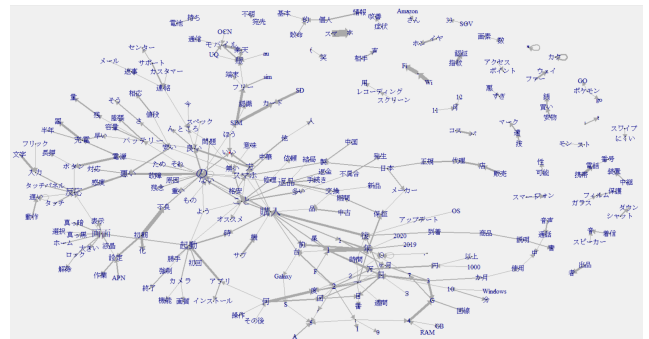


図7. 「金額が高い」ー「評価1」

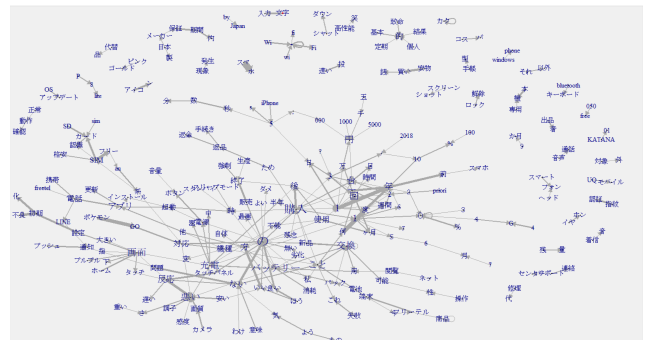


図8. 「金額が低い」ー「評価1」

次に、3.2.1節の⑤～⑦の特徴語について、共起ネットワークマップを用いて解釈を行った結果について述べる。

#### 5 4象限全てに登場した特徴語「購入」

「購入」に関する共起ネットワークを下記に示す。

- 高価格/評価5:本→機→購入, メイン→機→購入, サブ→機→購入, 子供→購入, 仕事→用→購入
- 低価格/評価5:本→機→購入, メイン→機→購入, サブ→機→購入, 子供→用→購入
- 高価格/評価1:サブ→機→購入, 交換→品→購入
- 低価格/評価1:購入→残念

評価の高い口コミからはサブ機としてだけでなく、本機としても購入されている。また、評価の高い口コミでは、子供に与えるスマホとして購入されていることもわかった。しかし現状子供の利用に合わせたコンテンツを含む製品が少ないことを踏まえると、今後の提案としてこれらコンテンツの拡大を提案する。PRITIMES[3]のアンケートからも、保護者の不安点として長時間利用が多くなることがあげられている。よって、顔や指紋の認証内容に合わせて利用時間の制限の付与、閲覧制限を自由に加えられるフィルタリングや、誤購入等を制限する本人チェック機能の向上などがあげられる。

#### 6 評価1のみに共通して登場した特徴語「画面」, 「充電」

「充電」は次の⑦のバッテリーの共起語として登場していたため、ここでは「画面」に関する共起ネットワークを下記に示す。

- 高価格/評価5:大変→満足→画面
- 低価格/評価5:画面→タッチ
- 高価格/評価1:画面→反応→悪い, 勝手→起動→画面
- 低価格/評価1:画面→反応→悪い, 画面→タッチ→反応→ない

評価の低い、または低価格の商品レビューから、画面のタッチセンサーの反応に関するネガティブな評価が見うけられた。特に低価格帯でより多くの書き込みがなされていたため、今後の改善点として、タッチセンサーの改善、認識の向上に加え、低価格化によるコストパフォーマンスの向上が望まれると考えられる。

#### 7 「金額が高い」のみに登場した特徴語「バッテリー」

「バッテリー」は下記に示すように「充電」とセットで登場していた。具体的に、「高価格/評価5」を除く3象限においてネガティブな状況が明らかとなった。

- 高価格/評価5:バッテリー→良い, バッテリー→長持ち
- 低価格/評価5:バッテリー→交換, 半年→充電
- 高価格/評価1:バッテリー→膨張, バッテリー→容量
- 低価格/評価1:バッテリー→交換, 半年→充電→バッテリー, バッテリー→消耗

評価の低い、または低価格の商品レビューから、バッテリーの耐久性に関するネガティブな評価がなされている。また口コミからはバッテリーの膨らみによってケースに入らない事例も確認した。今後の改善点としては、充電の繰り返し使用に対する耐久性の向上、また劣化による膨張の抑制があげられる。

#### 参考文献

- [1] Seki Yohei: Sentence Extraction by tf/idf and Position Weighting from Newspaper Articles, *Proceedings of the Third NTCIR Workshop 2002*.

- [2] 大槻明, 町田悠貴, 川村雅義: 「暇」ツイートのテキストマイニングによる潜在的ニーズの発見法に関する一考察, *情報通信学会誌*, Vol.36, No.2, pp.111-126, 2018.

- [3] PRITIMESアンケート:  
<https://prtimes.jp/main/html/rd/p/000000056.000096733.html> (2022/6/13閲覧済み)