

# SNSにおける化粧品のクチコミ分析

千葉商科大学 商経学部 商学科  
佐藤 優月, 赤木 茅, 江草 遼平, 橋本 隆子

人工知能学会 第24回 SIG-BI研究会  
大濱信泉記念館 / 2024年3月11日(月) セッション4

本研究は、「千葉商科大学・数理データサイエンス教育プログラム」における「特別講義(データサイエンス)」の一環であり、千葉商科大学 基盤教育機構による助成を受けている

# 目次

- 1 研究背景
- 2 先行研究
- 3 研究目的
- 4 データの説明
- 5 分析手法
- 6 結果
- 7 考察と展望
- 8 参考文献

# 1 研究背景

- **クチコミの拡大[1]**

- ・ SNSによってクチコミを共有する傾向が拡大した
- ・ クチコミの企業業績への影響が増大
- ・ クチコミをマーケティング活動に効果的に取り入れる手法が模索されている

- **化粧品のクチコミ**

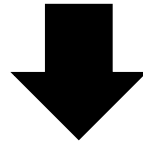
- ・ 種類が非常に多く、それぞれの特徴や効果を比較・検討することが難しい
- ・ クチコミの発信や、クチコミによる情報収集が活発に行われている分野の一つ

- **@cosme[2]**

- ・ 日本最大の化粧品・美容の総合サイト
- ・ 月間訪問者1900万人
- ・ 累計クチコミ数1940万件

## 2 先行研究

- **@cosmeの特定のマスカラに対する消費者のクチコミの分析[3]**
  - ・ 単語の重みを計算し, 特徴語として100語を抽出
  - ・ そこから6個のコミュニティが抽出され, 持続性や価格の評価、化粧品の落としやすさ等の異なるトピックや特徴を持っていることを明らかにした
- **レディースファッションに関するクチコミの季節や年による傾向の検証[4]**
  - ・ 楽天データセットからレディースファッションに関するデータを抽出し, 分析結果からレディースファッションに関するクチコミには季節的な特徴があることや, 年別の「おしゃれ」という語句と類似度が相対的に高い語句などを明らかにした
- **化粧品クチコミの自動的分類, レーティングの予測の試み[5]**
  - ・ 楽天データセットから美容・コスメ・香水に関するデータを抽出し, クチコミを3つのトピックに分類, 各トピックで重みが大きい単語を用いてクチコミのレーティングを予測する機械学習モデルを構築した
  - ・ 5段階のレーティングの予測正答率はやや低かったが, 2段階に集約することで正答率が向上することを明らかにした



- **@cosmeや楽天市場といったサイト上のクチコミを分析している**
  - @cosme, 楽天市場
    - SNSと比べて利用者数や日常的な利用時間は少ないと推察される
  - X (旧Twitter)
    - 日本国内で利用者の母数が特に大きいSNSの1つ
    - 化粧品に限らない発信や日常的なユーザー間の交流がある
    - より幅広いユーザーからの多様なクチコミの取得が期待される
- **マーケティング的な売上との関連性を分析したものではない**

# 3 研究目的

## Xにおける化粧品のクチコミと、化粧品の売上の関連を分析する

- **仮説：**

- クチコミの評価と売上には関係がある
  - 消費者は購入する化粧品を選定する際に、他の消費者の評価を参考にする
  - クチコミの評価が高い化粧品は購買意欲を高め、評価が低い化粧品は購買意欲を下げると思われる

- **手法：**

- BERTによってベクトル化された@cosmeの総合ランキング上位20位の化粧品に対するX上のクチコミをクラスタリングし、クチコミの特徴を分析、また、その売上との関連を分析する

- **結果：**

- クチコミの特徴は明らかになったが、売上との関連は見られなかった

## 4 データの説明

### ① X上のクチコミ

収集期間：2024/1/19~2024/1/25, 2024/2/5~2024/2/11

収集対象：@cosmeの総合ランキング上位20位の化粧品

Twitter APIにより収集：2589件



明確に広告やスパム等であると思われるものを手動で除外



分析の対象：1561件

## 広告・スパムについて

Amazon

ORBIS(オルビス) エッセンスインヘアミルク 洗い流さないトリートメント ヘアミルク 美容液 本体 140g 1.本体



場所: amazon.co.jp

クチコミに関する情報が含まれていない広告ツイート

セザンヌの超細芯アイブロウって知ってる？👁️  
👩 眉尻の1本1本まで描ける0.9mmのアイブロウペンシルなんだよ😱 しかも水・汗・皮脂に強いから落ちにくいし、無香料とか無鉱物油とか敏感肌にも優しいんだよ😊 7色展開で自分の髪色やメイクに合わせられるから便利だよ〜💕

セザンヌの超細芯アイブロウって知ってる？🍀  
🌈 眉尻の1本1本まで描ける0.9mmのアイブロウペンシルなんだよ😊 水・汗・皮脂に強いから落ちにくいし、無香料とか無鉱物油とか敏感肌にも優しいんだよ😊 7色展開で自分の髪色やメイクに合わせられるから便利だよ

内容が類似していることからスパムだと考えられるツイート



## ② @cosme SHOPPINGにおける売上データ

収集対象とした20個の化粧品を2024/2/26の売上データに基づき  
30位以内の売上上位群，30位以下の売上下位群に分類

ランク/ 売上群	化粧品名	ランク/ 売上群	化粧品名
1/下位	エッセンスインヘアミルク	11/下位	超細芯アイブロウ
2/上位	ヴォワールコレクチュールn	12/上位	サンシェルター マルチ プロテクション トーンアップCC
3/下位	ハンオールブロウカラ	13/上位	ライトリフレクティングセッティングパウダー プレスト N
4/下位	マルチファンデパフ2P	14/下位	ルージュスターヴァイブラント
5/上位	ルース パウダー	15/上位	ディオールスキン フォーエヴァー スキン コレクト コンシーラー
6/上位	スキנקリア クレンジング オイル アロマタイプ	16/下位	チークブラッシュ
7/下位	リップモンスター	17/下位	クリーミータッチライナー
8/上位	クリーム イン デイ	18/下位	リアルックミラー
9/下位	スピーディーマスカラリムーバー	19/下位	ディオールスキン フォーエヴァー フルイド グロウ
10/下位	ディオール アディクト リップ マキシマイザー	20/下位	ダーマレーザー スーパー VC100 マスク

## 5 分析手法

- ・ **BERTを用いたクチコミのテキストデータのベクトル化**

- ・ 自然言語処理モデルBERTを使用して、収集した化粧品のクチコミのテキストデータ1561件を数値ベクトルに変換

- ・ **主成分分析を用いた次元削減, 可視化**

- ・ 主成分分析によりベクトル化したテキストデータの次元を2次元に削減し、結果を散布図として表した
- ・ 1561個のデータのうち, @cosme SHOPPINGにおける売上ランキング上位30位までに入った化粧品に対するデータを赤色に表示し, 売上との関連性を見た

- **階層クラスタリング**

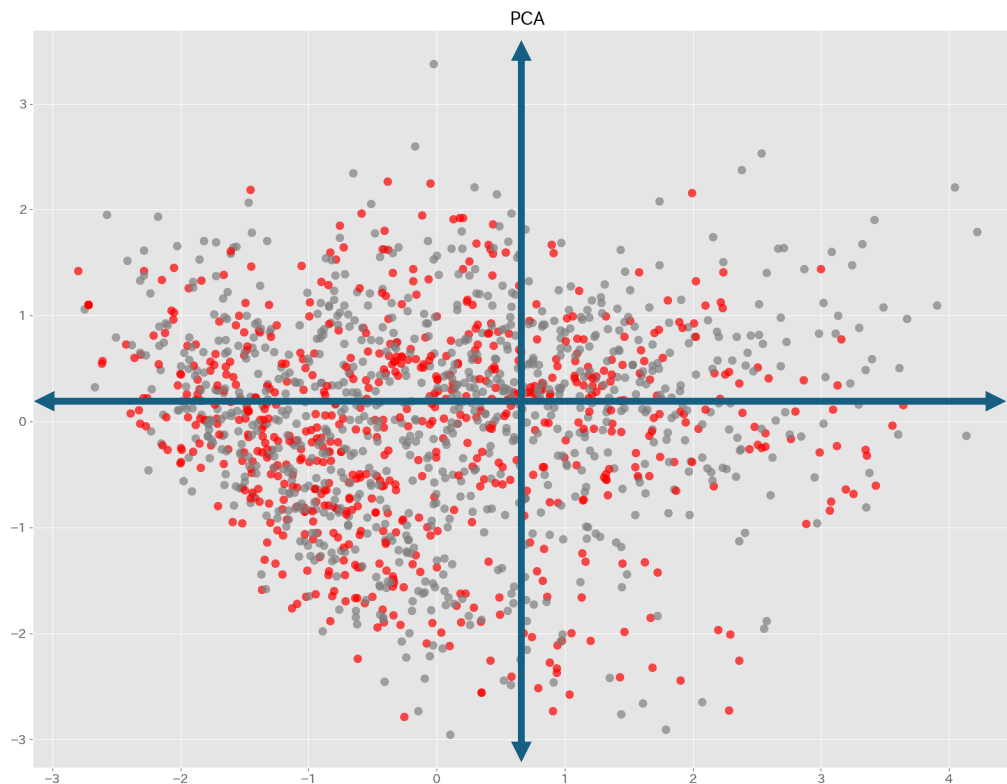
- ・ 主成分分析により得られたデータに対し, 階層クラスタリングを実施し, 6個のクラスタに分類

- **Word Cloudの作成**

- ・ MeCabを用いた形態素解析を行い, 単語の出現頻度に基づく Word Cloudを作成することで分類した各クラスタの特徴を分析
- ・ すべてのクラスタに共通して現れる語彙は除外

## 6 結果

売上ランキングにより色分けした散布図



**マーカー：**

ベクトル化された各クチ  
コミのデータ

**赤色：**売上上位群

**灰色：**売上下位群

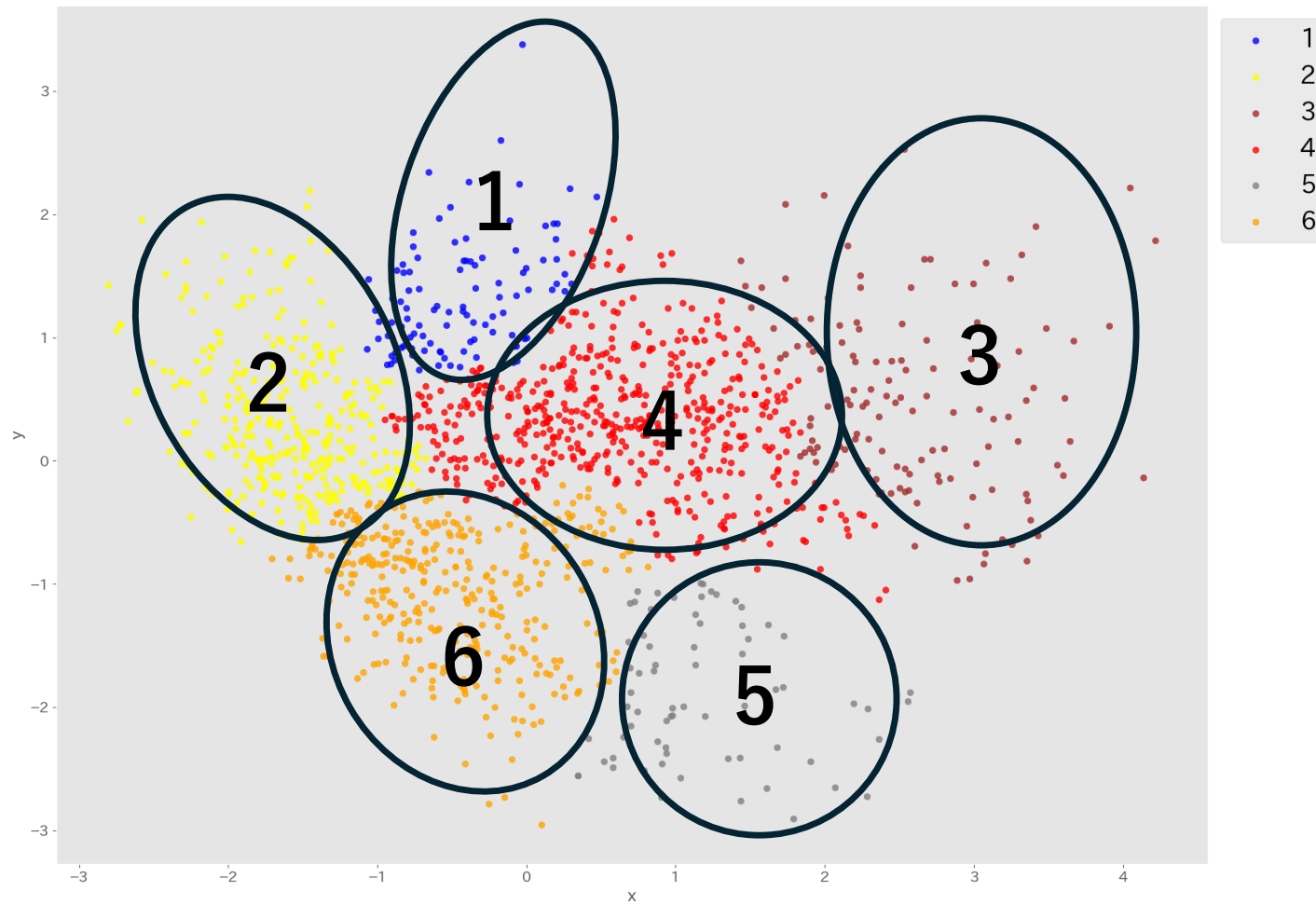
**縦軸：**第一主成分

**横軸：**第二主成分

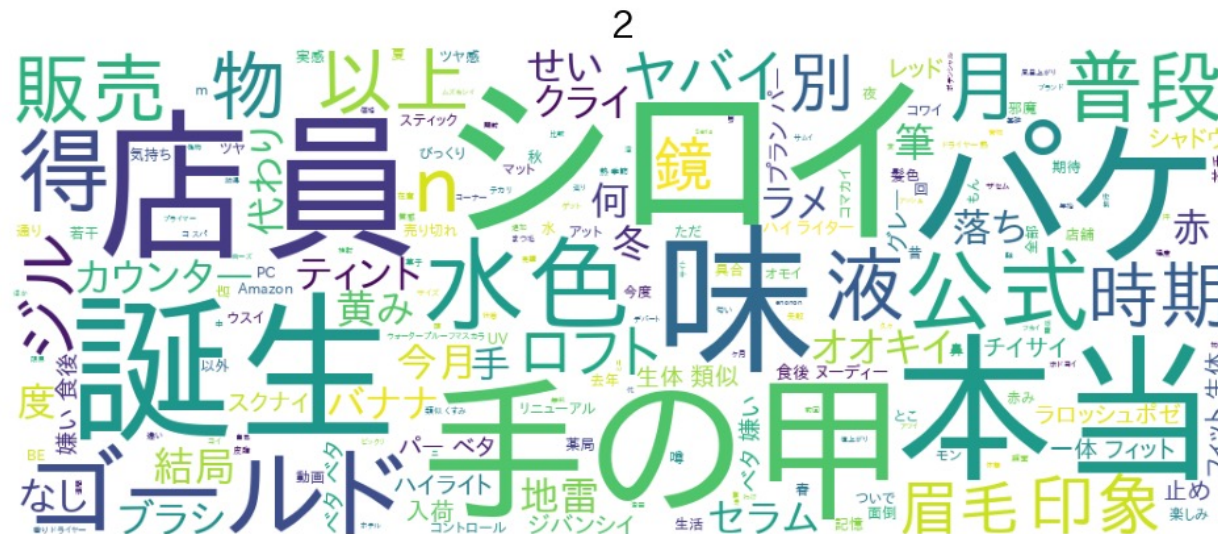
色の偏りが見られない

→売上とクチコミの関係  
性は見られない

## 階層クラスタリング結果の散布図

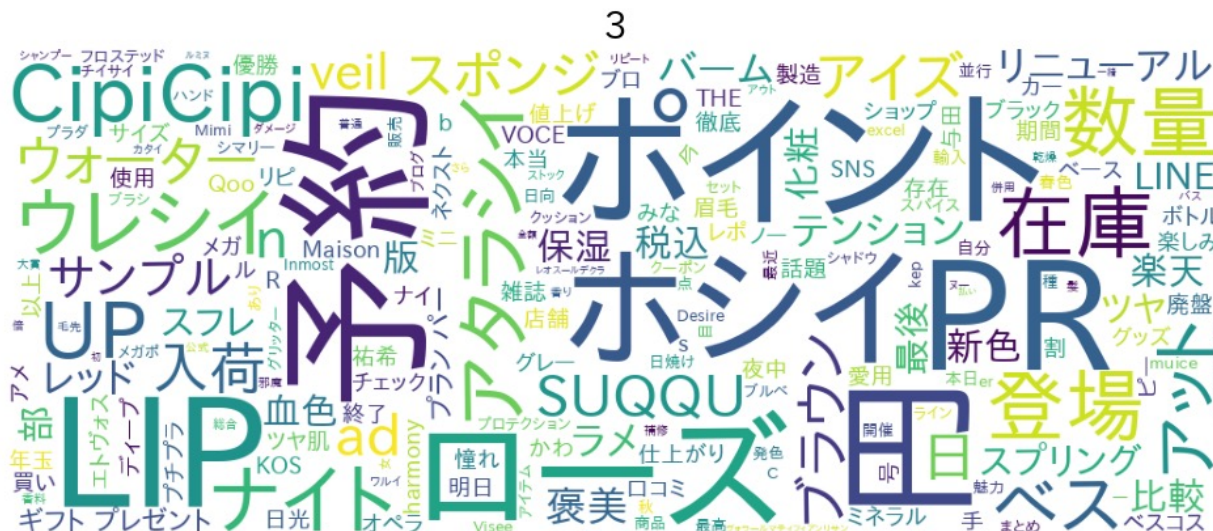






- ・「店員，手の甲，カウンター，ロフト」  
→店頭での体験
- ・「ゴールド，シロイ，クライ，  
黄み，びっくり，ベタベタ」  
→色や使用感





- ・「在庫，数量，新色，廃盤，予約」  
→在庫や商品展開，未購入者の投稿
- ・「バーム，スポンジ，スフレ，プランパー」  
→化粧品の状態







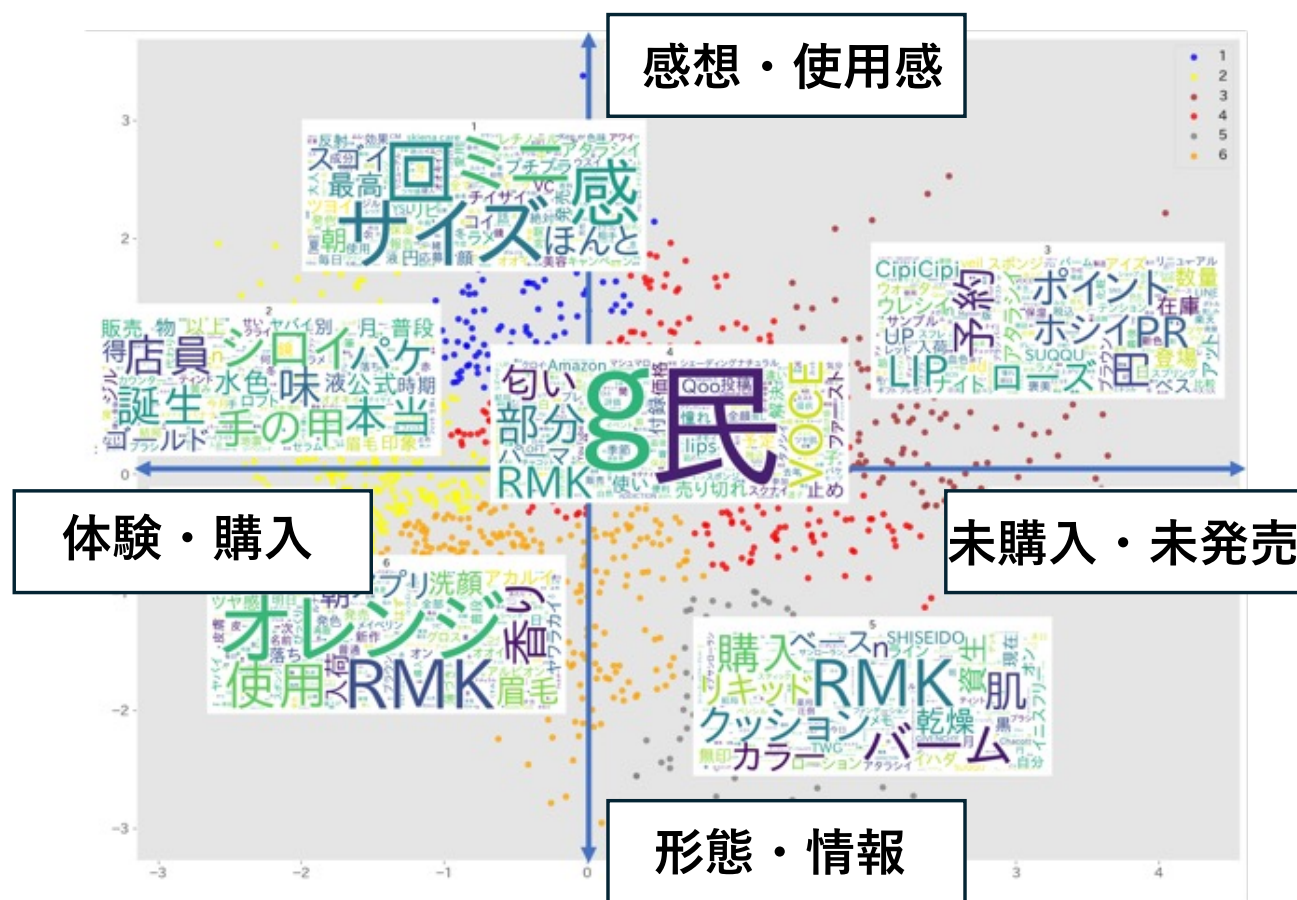
A scatter plot showing 6 classes of data points in a 2D space. The x-axis ranges from -3 to 4, and the y-axis ranges from -3 to 3. The classes are represented by different colors: 1 (blue), 2 (yellow), 3 (red), 4 (orange), 5 (grey), and 6 (dark orange). A black ellipse highlights a region containing points from classes 4 and 6.



## クラスタごとのクチコミの特徴

クラスタ	クチコミの特徴
1	常用, 使用感
2	色, 使用感
3	在庫, 商品展開, 広告関連, 購入予定
4	周辺情報, 欠品
5	形態, 分析対象外のブランド名
6	色, 形態, 欠品

## クチコミの分布傾向



- ・ **第1主成分（縦軸）**  
「形態・情報-感想・体験談」によるクチコミの性質の違い
- ・ **第2主成分（横軸）**  
「購入 - 未購入」によるクチコミの性質の違い



## 7 考察と展望

- ・ X上のクチコミは商品の「形態・情報-感想・体験談」の軸，「購入報告 - 購入困難報告」の2軸で要約可能
- ・ 分析されたクチコミの特徴と，売上との関連性は見られない
- ・ 今後の研究の展開
  - ・ 売上データについて30位よりも低い化粧品や同商品に関するクチコミの記述データ，Xのユーザーデータを分析に追加し，クチコミ投稿者やクチコミ内容の特徴を用いることで，売上を予測できるモデルを検討

## 8 参考文献

- [1]臼井浩子: クチコミの促進要因に関する先行研究の整理と今後の研究課題, マーケティングジャーナル, Vol.34, No.1, pp. 158-169, (2014)
- [2]株式会社アイスタイル: @cosme「化粧品の情報収集実態」調査結果～信頼できる情報源として「サンプル・テスター」などリアルな体験が重視されるも、購入の決め手となるのは「クチコミ」がトップに～,  
<https://www.istyle.co.jp/news/info/2023/09/0925.html>, (2023)
- [3]白田由香利, 橋本 隆子, 久保山 哲二: インターネット上の口コミサイトにおける化粧品の評判分析, 学習院大学計算機センター年報, Vol. 33, pp.2-7, (2012)
- [4]峰滝和典: レディース・ファッションに関するインターネット上のクチコミについてのデータ解析, 商経学叢, Vol. 69, No.2, pp.295-316, (2022)
- [5]峰滝和典: テキストマイニングと機械学習による化粧品のクチコミとレーティングの分析, 日本情報経営学会第84回全国大会予稿集, pp. 85-88, (2022)