

公立はこだて未来大学 2024 年度 システム情報科学実習 グループ報告書

Future University Hakodate 2024 Systems Information Science Practice

Group Report

プロジェクト名

コスメ×データサイエンス

Project name

Cosmetics × Data Science

グループ名

コスメサンプル班

Group Name

Cosmetic Sample Group

プロジェクト番号/Project No.

13

グループリーダー/Group Leader

松岡楓斗 Futo Matsuoka

グループメンバ/Group Member

丸山千寿 Senju Maruyama

伊藤優晴 Yusei Ito

指導教員

佐藤直行 富永敦子 島内宏和 花田光彦

Advisor

Naoyuki Sato Atsuko Tominaga Hirokazu Shimauchi Hanada Mitsuhiko

提出日

2025年1月21日

Date of Submission

January 21, 2025

1. はじめに

1.1 背景・研究動機・目的

消費者は化粧品を購入する際、テスターやサンプルによる疑似体験を行ってから化粧品を購入したいと感じていることが「@cosme(アットコスメ)」の調査(2021)[1]により明らかになっている。ここでのサンプルとは、化粧品メーカーが販売店等で提供するコスメサンプル(試供品)を指す。しかし、現状のコスメサンプルが消費者に与える具体的な効果については、十分に解明されていない。そこで本研究では、コスメサンプル使用の有無による違いを明らかにし、将来的なサービス提案に活用可能な知見を得るため、株式会社アイスタイルが企画・運営する美容系総合ポータルサイト「@cosme(アットコスメ)」に投稿されたクチコミデータを活用し、コスメサンプル使用の有無によるクチコミの内容を比較・分析する。

2. 関連研究

2.1 訪日外国人観光客の口コミ分析による旅館の魅力と類型化

森下(2023)[2]は、訪日外国人観光客に人気の上位20軒の旅館の日本語と英語のクチコミを比較して、旅館の魅力进行分析し、訪日外国人観光客にとっての魅力を論考した。本研究は訪日外国人観光客が旅館で感動し満足する要因を明らかにし、それらをもとに旅館の魅力を体系的に分類することを目的としている。それぞれ1000件以上の日本語と英語のクチコミデータを分析対象とし、KH Coderを使用してクラスタ分析と共起ネットワーク分析を行った。まず、クラスタ分析を実行して、4つのクラスタを定義した。日本語のクチコミのクラスタは「宿泊して朝食」「とても良い従業員の宿」「部屋での食事と風呂」「旅館の温泉露天風呂」と定義した。英語のクチコミのクラスタは「日本の伝統的な旅館での経験」「滞在場所での素晴らしい従業員と時間(夜)」「個室の温泉」「部屋での朝夕食」と定義した。次に、英語のクチコミデータに対応分析を実行した。その結果、日本の伝統とフレンドリーさが特徴と想定される宿のグループA、朝食のご飯がおいしい宿と想定されるグループB、夕食に特徴があることが想定されるグループC、星野グループの宿が挙げられているグループDとなった。この結果を踏まえて、各グループで共起ネットワーク分析を実行した。グループAには「日本の伝統的な和食や旅館滞在での経験」、「所せまいが清潔で快適な部屋の宿」、「親切な従業員と素晴らしい場所」といった日本の和室や和食が評価されている様子がうかがえる。グループBには、「日本の伝統的な旅館の朝食や風呂」、「日本での素晴らしい食や従業員の場所の経験」、「京都“白梅”のともこさん」といった特徴がみられた。「京都“白梅”のともこさん」とは、白梅の女将ともこさんの高い評価が分析結果にも現れていた。グループCには、「部屋での素晴らしい朝夕食の時間」、「日本の伝統」、「フレンドリーな従業員」「個室風呂」といった部屋での過ごし方や温泉を楽しむ体験が多く描写されていた。グループDには「宿や部屋で従業員から受けるサービスの体験」、「日本の伝統」、「渡船からの景色」などサービスやおもてなしに言及するクチコミが多かった。これらのことから訪日外国人観光客は日本人には珍しくない時代遅れの管内や設備、備品に囲まれた和室で浴衣を着て食べる和食など旅館での異文化体験を古き良き日本の伝統として高く評価していることがわかる。訪日外国人観光客の受け入れには、一定レベルの外国語での対応や無料Wi-FiなどのICT化や、日本の伝統、食、サービス(おもてなし)を基本に、旅館独自の特徴を見だし、アピールすべきと示唆している。この研究から本グループにおいてKH Coderを使用して作成した共起ネットワーク図は単語同士の関係进行分析するのに有用であることがわかった。本研究ではクチコミデータを使用しているため、活用できる可能性がある。

2.2 アンケートを用いた景観資源と景観特性の抽出に関する手法の考察

恒松ら(2023)[3]は、秋田市の景観政策に有用なデータと分析手法を明らかにすべく、アンケート調査を通して分析した。アンケート調査では、秋田市内の全域で行政区分ごとに100部、さらに住所区分に準じた地域を設定し、その地域数により100部を除いた値を各地域に均等に配布した。以上、市内全58地

域に対し700世帯を対象に無作為による配布を行った。アンケート内容は、回答者の属性、景観への関心について、景観に対する意識等について、景観に関する秋田市の行政の取り組み等について、景観に関する自由意見の5部構成とし、選択式の設問については平成20年次のアンケートと共通項目を設定した。最終的にアンケートは全260世帯の回答が得られた。まず、選択式の設問では、「景観に関心がありますか」という問いに「関心がある」「少し関心がある」の回答が合わせて8割を超え、秋田市における住民の景観に対する関心が高いことがわかる。さらに、「秋田市全体の景観について、どのように感じますか」という問いに「良い景観だ」「少し良い景観だ」と答えた人が半数を超え、平成20年の結果と比較しても秋田市における住民の景観に対する意識は低下することなく保持されていることがわかった。また、自由記述の設問では2つの問いを用意した。1つ目の問いは「秋田市全体で好きな景観・残したい景観はありますか」という問いで、ここから景観についての固有名詞を集めた。集まった景観資源は建造物に限らず、公園や広場、自然景観などさまざまな景観資源が抽出された。これらをクラスター分析を用いて類型化し、8グループに分類した。分類結果より、点在している景観資源の類似度による構成を定量的に明らかにした。2つ目の問いは、「秋田市の景観について自由にご記入ください」という今後の景観施策で必要とされている景観の特性を把握するための問いだ。総回答数は101で、回答結果の内容としては景観に対して思うことや、景観への取り組みについてなど多様な回答が集まった。この回答はデータクリーニングを行った後、テキストマイニングを用いて分析を行なった。データクリーニングの際には、語句の統一や強制抽出といった語句の検討と、主題ごとに文章を分割し1つのサンプルとするなどサンプルの区切り方の検討を行った。これにより、総サンプル数は132となった。これらをKH Coderを用いて単語抽出を行うと、抽出語の合計は5348語、1サンプルあたり約41語となった。抽出された語を用いて共起ネットワーク図を作成すると、大きな特徴がみられるサブグラフが01, 03, 05の3つあった。まず、サブグラフ01は住民からの希望や要望、市民の協力を必要とする声がみられた。これらの要望を景観施策に付与させていく必要があると考えられる。次に、サブグラフ03は緑化や樹木の維持管理に関する改善を求める意見が多くみられた。一方で秋田市中心部では維持管理されているという意見も含まれていた。最後に、サブグラフ05は田園や山、川が秋田市の景観で重要だと考える一方で、風車や看板などのマイナスな要素であるものが建てられていくことなど、重要であり魅力でもある風景を守るための制限が求められていることがわかった。自由記述の設問では、選択式の設問では掘り下げられない景観に対する視点や要望を含む内容を抽出でき、それらをサブグラフで可視化することができた。これらのアンケートを用いた分析は、景観資源の網羅的抽出に寄与できるものと考えられる。また、景観特性に関しては、景観計画等の景観施策の見直しの際に、その基礎的な資料となると言える。一連のアンケートや分析手法は秋田市での結果が示すように、一手法としての有用性が明らかになったと言える。この研究からデータクリーニングを行う際には語句の検討やサンプルの区切り方に注意しなければならないことがわかった。

2.3. @cosmeからみた、新しい日常における生活者の美容意識の変化

原田ら(2021)[4]は、新型コロナウイルスの登場によって生活者と化粧品のかかわり方に生じた変化を「@cosme」に投稿されたクチコミや会員への意識調査の結果をもとに考察した。まず、コロナ禍における1回目の緊急事態宣言発令を期に、クチコミの中で「衝動買い」というワードの出現率が減少し続け、コロナ前の半分まで落ち込んでいることに着目した。1回目の緊急事態宣言発令をきっかけにEC購入が増加したことや、店頭での買い物においても短時間で済ませたい気持ちが強まったであろうことも影響しているだろう。次に、いつもよりも慎重になっている生活者に安心なイメージをもってもらうには、コロナ禍以上に購入前の使用体験が重要になってくると同時に、店頭でのテスターではなく、できる限り普段の環境下で試したいという生活者の想いも強まってきていると考えた。そうしたニーズへの受け皿として、注目が高まっているのが、本来無償で入手できるサンプルサイズの商品を、有償で購入するという流れだ。「@cosme」では投稿されたクチコミが「現品」か「サンプル」か、そして「購入品」か「プレゼント」かが判別できる仕様になっているのだが、コロナ後「サンプル」の「購入品」クチコミの投稿率は徐々に増加している。このことからサンプルサイズの有償提供は、生活者のニーズにマッチしているのではないかと考え

ている。さらに、コロナ禍の「@cosme」で支持を受けた新商品のランキングに着目した。ランキングでは一般販売に先駆けて、先行販売を実施した商品が多くランクインしていた。先行販売を行うことで一般販売した時に、生活者が購入者のクチコミを見て見当ができる環境を用意できたことが非常に大事な点であったと考えている。また、経済が不安定な時期は、長年生活者から支持されているロングセラー商品が選ばれる傾向があったが、コロナ禍では人気商品上位の中でロングセラーが占める比率は減少傾向にあることが分かった。コロナ後は「ロングセラー商品」ではなく、「ロングセラーのリニューアル商品」が増加傾向にあったのだ。これには、安心感を求めつつも、「閉鎖感のある毎日から何かを変えたい」という生活者の気持ちが反映されていると考える。この気持ちが反映されている商品はほかにもある。例えば、コロナ後の「@cosme」のクチコミでは季節を感じられる植物やイベントを表すワードの出現率が例年以上に多く出現している。これは外出や旅行が制限され、季節の移り変わりを感じられにくくなっているため、化粧品でメリハリをつけているようだ。自分で楽しむことはもちろん、友人や母、パートナーに「プレゼント」を挙げた、もらったというクチコミも増加している。このようなことから、新型コロナウイルスの登場によって生活者の化粧品へのニーズが一変してしまったかのようにも見受けられるが、必ずしもそうではないという見解を示した。この研究から本グループにおいて生活者の美容意識によってクチコミの内容が変化することがあることがわかった。本研究において、クチコミに出てくるワードや内容の変化から生活者の美容意識に変化があると捉えられる可能性がある。

3. データと分析方法

3.1 データ概要

本研究で使用するデータは、株式会社アイスタイルが企画・運営する化粧品コミュニティサイト「@cosme」に投稿されたクチコミおよび、それに付随する情報を基に分析を行った。@cosmeは、化粧品レビューを中心としたプラットフォームであり、サイト開設から約25年が経過し、現在では日本国内において主要な化粧品レビューサイトの一つとして広く利用されている。また、同社はサイト運営のみならず、実店舗での化粧品販売事業や化粧品ECサイトの企画・運営など、多角的な事業展開を行っている。

@cosmeでは、ユーザーが化粧品に関するクチコミや評価を自由に投稿することが可能であり、投稿されたクチコミは、他のユーザーによる化粧品選択の参考として利用される。現在、@cosmeには約40万点の化粧品情報が掲載され、投稿されたクチコミデータは2,000万件に達している。特に、@cosmeの主要な利用者層は日本国内の20代～30代の女性であり、サイトには日々最新の化粧品情報が更新され続けている。

本研究では、@cosmeに投稿されたクチコミのうち、2023年度に投稿された1,852,998件のデータを対象とした。具体的には、各クチコミの本文に加え、クチコミ投稿者に関する情報(性別、年齢、肌質など)および、クチコミに対する他のユーザーの評価データを含む。本研究で用いるデータの構成については、各データにNULLが含まれないもの1851687件を使用した。データの主キーはクチコミIDであり、各クチコミに紐づく詳細情報が付随する構造となっている。

3.2 前処理

以下の手順に沿って前処理を行い分析を進めていく。

(1) データから欠損値(NaN値)を含むデータを除去する。

- (2)クチコミ本文中の単語を基にサンプルコーパスを作成する。
- (3)サンプルコーパスのデータ数を基に商品ジャンルを選定を行う。
- (4)サンプルコーパス・非サンプルコーパスからそれぞれ1,000件を標本として抽出する。
- (5)抽出したクチコミデータから記号・環境依存文字・数字を除去する。

(1)データから欠損値(NaN値)を含むデータを除去することについては、今回使用するカラム、クチコミ本文、商品大カテゴリに含まれるnan値を消去し、全データ1852998件から1851687件へ絞られた。

(2)クチコミ本文中の単語を基にサンプルコーパスを作成する際、提供されたデータには、sample_flag というカラムが含まれている。このカラムは@cosmeにおいてユーザーがクチコミを投稿する際、サンプルを使用した感想であるかどうかを示すフラグに紐づいている。しかし、実店舗でのテスター使用や、他のサービス経由で提供されたサンプルについては、ユーザー間で入力基準が統一されていない。そのため、実際にサンプルを使用した場合でも sample_flag が適切に設定されていないケースが確認された。この問題を補完するため、本研究ではクチコミ本文に「サンプル」「モニター」「提供」という単語が含まれるものを、サンプルに関与するクチコミとみなし、新たにサンプルコーパスを構築し、それに基づくラベルを作成した。なお、サンプルコーパスの単語選定においては、以下の基準に基づき、他の単語を除外した。

- ・「テスター」や「試供品」などの単語は、店頭での利用に関する記述が含まれる可能性が高く、現品購入との区別を明確にする目的に適さないため除外する。

- ・他の類似単語については、使用頻度が極めて低い、または文章の文脈上、サンプルやモニターとして利用したかの判別が困難であるため除外する。

以上の条件によりコーパスを作成し、データ数1851687件のうち133724件をコーパスとして分類した。

(3)本研究でテキストデータを対象とするため、データ量が過剰になると分析の障害となる。そのため、データの絞り込みを商品ジャンルとサンプルフラグに基づいて実施した。

この絞り込みの理由として、商品ジャンルを限定せずに分析を行うと、各商品に特有の単語が分析結果に影響を及ぼし、ノイズとなる可能性があるためである。例えば、ヘアケアや洗髪関連の商品では「髪」「サラサラ」「ダメージ」「トリートメント」などの単語が頻出し、これらが他の単語と独立して現れる傾向がある。これにより、本研究で焦点を当てるサンプルに関する情報と無関係な特徴が強調されてしまう。この影響を軽減するため、本研究では商品ジャンルによる絞り込みを実施した。

商品ジャンルの選定には、提供データ内の商品大カテゴリー(l_item_category_name)を基準とし、サンプルに関与するクチコミ数が最も多い以下の3ジャンルに限定した。

- ・スキンケア・基礎化粧品

- ・メイクアップ

- ・ベースメイク

これらのジャンルに絞ることでサンプルコーパスの数を保持しつつ、133724件から88780件に絞ることができる。

(4)さらに、開発環境と分析にかかるコストから、BERTなどを使用したベクトル化の処理負荷を考慮し、プログラムコードの分析を瞬時に出すことができる1000件のクチコミデータに絞る。そのため、サンプルコーパスでは、88780中1000件。非サンプルコーパスでは、44944中1000件を標本として分析を進めていく。

(5)本データはユーザーが自由に書き込める形式となっているため、装飾文字や不適切な記号表現などが含まれる。これらは、形態素解析や BERT を利用したベクトル化、主成分分析を行う上で障害となりうるため、前処理として削除した。具体的には、以下の処理を適用した。

- ・記号・環境依存文字の削除(句読点, 特殊記号, 絵文字など)
- ・数字表現の削除(商品価格や個数表記が不要な影響を与えないように処理)

3.3分析手法

本研究の分析は、Google Colaboratory 上で Python 3 を用いて実施した。分析の手順は以下のとおりである。

(1)クチコミデータのベクトル化を行う。

(2)ベクトル化したクチコミデータに対し主成分分析(PCA)を適用し、サンプルコーパスと非サンプルコーパスの比較を行う。

(3)クラスタリングを適用し、主成分分析の結果と比較することで、主成分分析によって抽出された特徴について詳細に分析する。

(1)本研究では、クチコミデータのベクトル化に、東北大学が提供する日本語向けの事前学習済み BERT (Bidirectional Encoder Representations from Transformers) モデルを採用した。本研究で使用したモデルは、東北大学が提供する、日本語 Wikipedia を教師データとして訓練された `cl-tohoku/bert-base-japanese` である。本モデルは、既存のtransformerと同様な枠組み(12層のレイヤー、768次元のパラメーター、12個のアテンションヘッド)で構成されており、東北大学によって事前に学習が行われているモデルを使用する。

BERT は Google によって 2018 年に発表された自然言語処理モデルであり、双方向の文脈を考慮して単語の意味を捉える特性を持つ。BERT は計算コストが高いというデメリットがあるが、クチコミデータの分析において TF-IDF や Word2Vec などの手法では単語の頻度や単語間の関連性に強く依存するため、文全体の意味を適切に捉えることが困難である。特に、クチコミデータのような短文の集合においては、単語単位ではなく文脈を考慮したベクトル表現が求められるため、BERT によるベクトル化を採用した。

また、分析過程において文章の視覚化として共起ネットワークを用いた際、化粧品特有の頻出語が多数出現することが確認された。これらの頻出語は分析の精度を低下させる可能性があるため、BERT を用いることでより適切な意味的特徴を抽出できると考えた。

(2)次に本研究では、ベクトル化したクチコミデータ(786次元)を Kernel Principal Component Analysis (KernelPCA)を用いて2次元にプロットし、データの分布を視覚的に分析する。高次元のデータを扱うためデータが非線形な向きにばらつきがある可能性を考え、線形の PCA ではなく KernelPCA を適用し、高次元での非線形な分散を最大化させるように次元を圧縮させていく。これによって、高次元を段階的に圧縮するよりも特徴が潰れることがなく、通常のPCAよりも特徴をつかみやすくなる。

(3)さらに、主成分分析によって圧縮される際に失われた情報を補完するため、主成分分析を適用する

前のベクトルデータに対して K平均法のクラスタリングを実施し、主成分分析では顕在化しなかった特徴についても検討を行う。

本研究では、以下の 4 種類のグラフを作成し、分析結果とした。

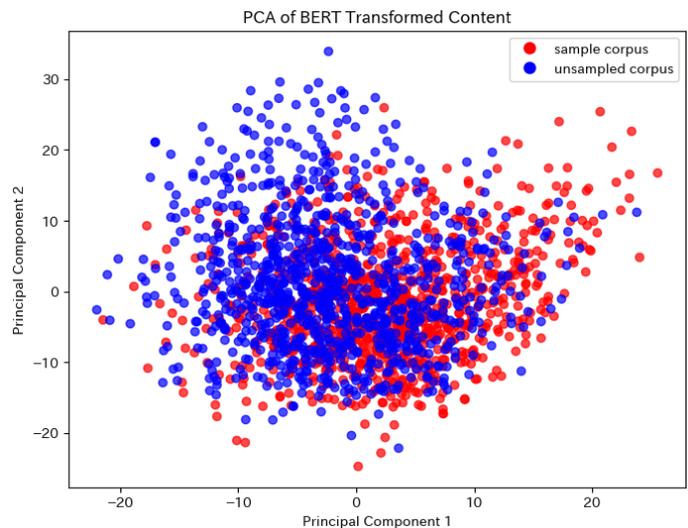
- 主成分とサンプルコーパスの関係図
- クラスタリング結果と主成分の関係図
- 第一主成分と単語数の関係
- 第一主成分と品詞割合の関係

4. 結果

4.1 主成分分析

図1は、ベクトル化した文章に対し KernelPCA を適用し、第一および第二主成分をプロットし、サンプルコーパスの所属を示したグラフである。図1から、大半のデータは中心付近に分布しているものの、赤色のサンプルコーパスに着目すると、やや右方向に偏っていることが確認できる。一方、青色の非サンプルコーパスは左側に強く偏って分布している。

また、第一主成分の 0 を基準としてサンプルコーパスの左右比を比較すると、左右いずれも 6.5:3.5 となっており、



明確な偏りが生じていることがわか

図1

4.2 クラスタリング

図2は、以下の手順で作成した。

(1)クチコミデータをBERTを利用してベクトル化する。(768次元)

(2)ベクトル化したデータに対してK平均法でクラスタリングを行い、各クラスターラベルを作成する

(3) (1)のデータにKernelPCAを適用し、
(2)で作成したラベルを基に色分けを行う

クラスタリングはK平均法を使用し、クラス数はエルボー法及び、シルエットスコアから適切に思われる3分類とした。また、他のクラスタリング手法や非階層クラスタリングについても検討したものの、分類する文章ベクトルが768次元と多く、適切な分類が難しいため、K平均法を用いて分類を行った。

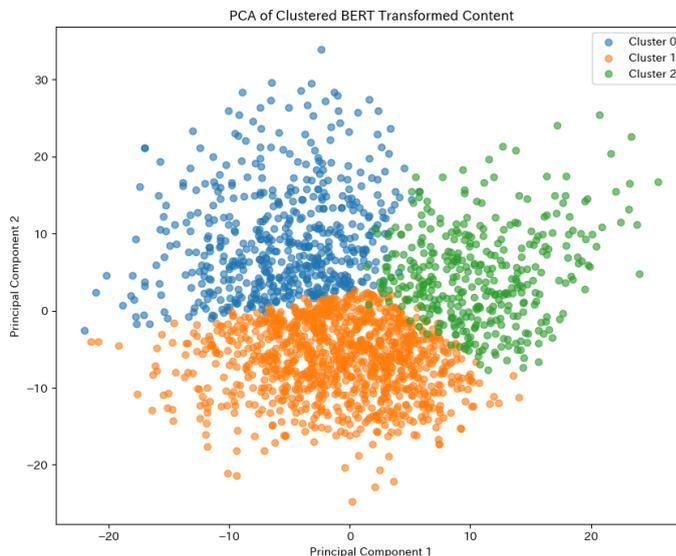


図2

以下に各クラスターの中心に位置するクチコミの例をあげる(表3)。これらは、各クラスター中心からcos類似度を計算しその上位3つのクチコミの元データをあげている。今回のクラスタリング結果から本文を比較すると見えるものとして、クラスター0では比較的専門用語が使用されず、主観的な意見が多く文章も短いように見える。一方、対照的にクラスター2では専門用語や固有名詞が多く使用されており、客観的な説明が多く長い文章になっている。また、他のクラスターと比較し、クラスター1では特別な特徴が見受けられない。

表3 各クラスターのCos類似度上位3件テキストデータ

クラスター0「簡素な群」	クラスター1「率直な群」	クラスター2「詳細な群」
<p>ふってから使うタイプのファンデーションで、のびがいいです。気になる毛穴もカバーしてくれ、気になる部分に重ねづけしてもヨレないところもよかったです。カラーバリエーションが豊富、好みのカラーがきつと見つかるラ ~~~~</p>	<p>一本使いきりました。こちらは洗顔後すぐに使用するものですが、私はビタミンCの後に使用しています。1回3プッシュを朝晩。こちらを使用してから肌の乾きが気にならなくなりました。朝起きると少し乾燥してしまっていたの ~~~~</p>	<p>・エクセルのこの秋のテーマは“MODERN NEUTRAL”★カラーエディットパウダーブロウEP02:サマーネクター, EP04:ウィンターモーヴ 秋を思わせる素敵な配色で見ていただけで気分が上がります ~~~~</p>
<p>ベビーオイル洗顔してます。乾燥しないし毛穴の開きが改善して気に入っています。</p>	<p>焼けるか焼けないかについては長時間日に当たることがないのでわからないのですが、塗り心地が良く、いい匂いがします。肌が擦れると少しモロモロが出てしまう時があります。ニベアの美白ボディクリームを ~~~~</p>	<p>#PR @clearturn_jp ごめんね素肌 クマらないアイクリーム 大人気すぎて品切れ続出したアイクリーム パッケージが可愛すぎ レチノール誘導体やナイアシンアミド配合なのに990円と価格も破格！ 低刺激処方でも目のケア ~~~~</p>
<p>使い心地が好みでした。</p>	<p>・選べるサンプルプレゼントキャンペーンのときに頂いたものです。 他にはSMRコンプレックスかアイクリームでしたが、迷わずコチラで大正解です。 ・・・というか、クチコミを書く今まで知らなかったんですが高いなキミ！ 20mlで14080円って、サンプルのサイズが5mlくらいなので、単純計算で3500円以上・・・、エステイ様よ！え ~~~~</p>	<p>イニスフリーのプロモーションに参加中です。ビタミンCスキンケアですが真夏に限らず使いやすい使用感で、オイルっぽさよりほんのりとろんとしたテクスチャーで広げやすいです。続けて使ってみてもピリピリせず使えています◎ 手持ちのUVやプライマーを選ぶことなく変なべたつきやモロモロみ ~~~~</p>

4.3 Kernel PCA第一主成分に着目した分析

第一主成分について、明確に差が生まれていた要素として、単語数がひとつあげられる。図4は、第一主成分でソートした文章データを50ずつ平均をとり、横軸を第一主成分、縦軸に文章の平均単語数をプロットしたものである。明確に第一主成分の値が増えるほど単語数も増える関係がうかがえる。また、単語数が単調増加していない点についても、第一主成分が単語のみを特徴として分類していないことが明確となり、第一主成分の妥当性と単語との関係性が示唆された。

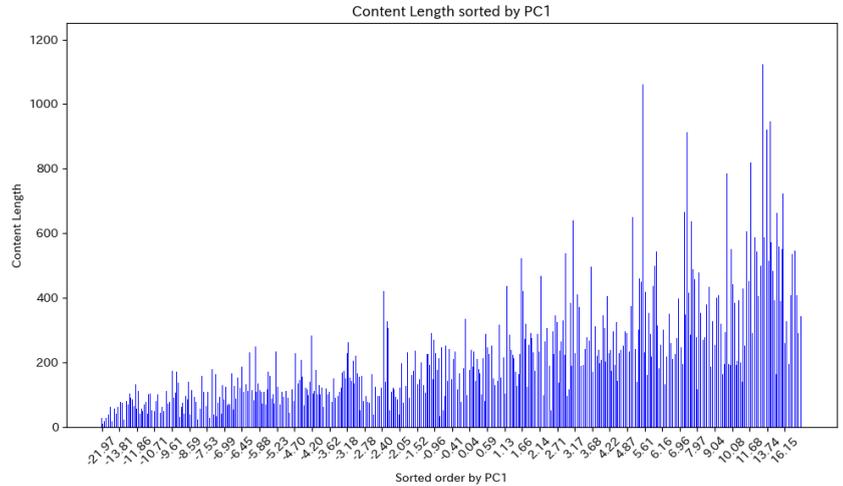


図4

また、図5は同様に第一主成分で小さい値のものからソートされた文章を100ずつグループ化し、それら文章の品詞割合をプロットしたものとなる。左か右にかけて名詞が緩やかに増加し、第一主成分の高い上位3グループからは明確に名詞の割合が増加していることがわかる。また、名詞の変化に比較すると微量ではあるが、助動詞や副詞は第一主成分の増加に合わせて、割合が減少していることが見て取れる。

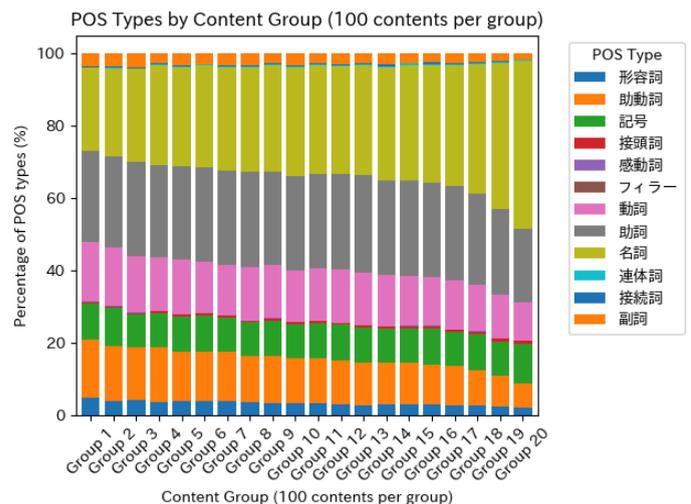


図5

5. 考察

5.1 主成分分析の結果についてのまとめ・考察

この研究は、コスメサンプルを使用したクチコミとそうでないクチコミの内容を比較してどのような違いがあるのかを明らかにすることを目的として実施した。サンプルを使用したユーザーとそれ以外のユーザーにフラグをつけてカーネル主成分分析を行った結果、横軸に焦点をあてた際に左右に分布する傾向が見られた。このような結果から、コスメサンプルは、ユーザーが各クチコミに影響を与えていることが示唆された。この影響については、コスメサンプルを受け取ることによって良いクチコミを書かなければいけないというバイアスがかかるため、起こることが原因であると考えられる。この研究で得られた結果をもとに、クラスタリング、PCA第一主成分に着目した分析を行った。

5.2 クラスタリング結果についてのまとめ・考察

この研究は、5.1節で実施したカーネル主成分分析の結果から明らかになったクチコミ利用者の傾向を把握し、その分析結果をもとにサービス提案につながる知見を得ることを目的として実施した。主成分分析で使ったデータをもとにクラスタリングを行い、各クラスタの特徴を把握するためにcos類似度を計算し、クラスタの中心のクチコミからどのようなクラスタなのかを考察した。その結果、端的なクチコミ、体感したことをそのまま書いているクチコミ、詳細なクチコミの3つのクラスタに分けることができた。

1クラスタ目の端的なクチコミには、一言で完結しているクチコミや「潤う。」などのような、1単語で書かれているようなクチコミが多く見られた。このクラスタのクチコミは、簡潔に商品についての情報が書かれているため、あまり文字を読まなくても気になっている商品について知れることがメリットである。

2クラスタ目の体感したことをそのまま書いているクチコミには、「・・・というか、クチコミを書く今まで知らなかったんですが高いなキミ！」のように話し言葉で個人の感想を書き連ねたようなクチコミをはじめに、効果を主観的に書いているクチコミが多く見られた。使用感などのような主観的感想を求められやすいクチコミサイトでは、このクラスタがユーザーが1番求めている情報に近いものであり、需要があるのではないかと考える。

3クラスタ目の詳細なクチコミでは、「大人気すぎて品切れ続出したアイクリーム パッケージが可愛すぎレチノール誘導体やナイアシンアミド配合なのに990円と価格も破格！」のように、詳細に商品の成分や使い方などの説明や使用感などを書き込んだクチコミが多くみられた。このクラスタのクチコミは、商品について詳しく知った上で購入したいユーザーに需要があると考えられる。

5.3 PCA第一主成分に着目した分析についてのまとめ・考察

この研究では、5.1節で実施したカーネル主成分分析の結果に根拠を持たせるために、第一主成分に対しての品詞分析と単語数の分布についての言及を実施した。単語数の分布に関しては単調増加とはならなかったものの、第一主成分が右側に進むにつれて単語数が増加する傾向が見られた。この結果から、文章量だけが第一主成分の特徴でないことが示唆されたとともに、単語数が第一主成分の特性を把握するための重要な要素であることが示唆された。

さらに、品詞分析を実施した結果、第一主成分が右側に位置する要素では名詞が比較的多く含まれ、助動詞や副詞の割合が減少する傾向が確認された。この結果から、右側に位置するクチコミは成分や商品名などの具体的な名詞を多く含む一方、助動詞や副詞が少ないことで、比較的客観的で情報量の多い文章構成を有していると考えられる。一方、左側に位置するクチコミは、助動詞や副詞の割合が高く、主観的な内容が中心であることが示唆された。このような結果は、第一主成分の意味づけを行う上で有力な知見であるといえる。

この結果は、クラスタリングで私たちが考察したクラスタの特徴にも対応していることが言える。対応させて考察すると、詳細な情報のクチコミとして定義したクラスタは右側に位置していて、比較的客観的で単語数、名詞の豊富さにより情報量が多い文章構成になっている。一方で、端的なクチコミでは、助動詞、助動詞、副詞などが比較的多く、単語数は少なめであるため主観的な内容が簡潔にまとめられている文章構成になっている。体感したことをそのまま書いているクチコミでは、比較的左側に分布していて主観的なクチコミである一方で、端的なクチコミよりは単語量は多くなっている傾向があることが考えられる。

今後、クチコミの言語的特徴をさらに詳細に分析し、ユーザーが投稿するクチコミの多様性をより深く理解することが可能であると考えられる。特に、名詞、助動詞、副詞の使用パターンや文全体の構造的特徴を定量的に評価することで、クチコミの情報密度、信頼性、感情的価値の違いをより高い精度で分析できることが期待できる。また、品詞の特徴と感情分析を組み合わせることで、クチコミが購買意欲や製品評価に与える影響を具体的に明らかにする枠組みを構築することが可能となる。この枠組みを、コスメサンプルについてのクチコミに反映することで、コスメサンプルの有用性について言及することが可能になると考えられる。

6. ビジネス的活用方法の実践について

本研究では、コスメサンプルを使用したクチコミとそうでないクチコミには異なる特徴が含まれていることが示された。また、クチコミの内容を分析することで、それらを「詳細なクチコミ」「簡潔なクチコミ」「体感をそのまま書いているクチコミ」の3種類に分類できた。そして、クチコミを閲覧するユーザーにも、多様なニーズが存在することを考察した。具体的には、詳細な成分情報を把握した上で商品を使用したいユーザーや、簡潔に効果のみが記載されたクチコミを確認したいユーザーがいることなど、いろいろな思考のユーザーがいることが予想される。

このような知見に基づいて、本研究の結果は、誰にでも刺さるクチコミサイトの構築に貢献する可能性があると考えられる。その実現のためには、コスメサンプルを使用することで生じるクチコミの言語学的変化をさらに研究して、得られた知見をコスメサンプル配布方法に反映することが重要である。例えば、クチコミを投稿した顧客にコスメサンプルを配布してクチコミを書いてもらい、サンプルを配布する前後での言語学的特徴の変化について分析する研究である。

活用事例として、詳細な情報を含むクチコミが不足している商品に対し、過去に詳細なクチコミを投稿した履歴のあるユーザーにコスメサンプルを配布する方法が考えられる。このような手法を用いることで、特定の商品におけるクチコミ内容の偏りを抑制し、情報のバランスを改善できる可能性がある。また、得られた知見は、クチコミ欄の情報整理を通じて、ユーザー視点で分かりやすいECサイトの構築に繋がることを期待できる。

さらに、本研究の成果は、ユーザーが求める情報を適切に提示することで、商品の魅力を効果的に伝え、購買意欲を高める仕組みを提供する基盤となる可能性がある。そのため、コスメサンプルの配布方法の工夫が、ユーザー満足度の向上とビジネス価値の向上に繋がることを期待される。

参考文献

- [1]株式会社アイスタイル,「@cosmeにおける二次流通での化粧品購入実態調査」結果, 2021-10-14.<https://www.istyle.co.jp/news/info/2021/10/1014.html>.(参照 2025-01-12)
- [2]森下俊一郎, 訪日外国人観光客のロコミ分析による旅館の魅力と類型化, 日本経営診断学会論集, Vol. 23, pp. 69-75, 2023,
- [3]恒松良純, 鎌田光明, 佐藤涼風(2023)アンケートを用いた景観資源と景観特性の抽出に関する手法の考察. 日本建築学会技術報告集, 29(71):357-361
- [4]和田 安加里, 金 珉智, 2024年, Z世代におけるファッションの消費行動の特徴分析-マーケティング企業調査の結果に基づいて-, 教育経済学研究, 2024年5巻, p54-59