

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

クチコミ可視化班

Group Name

Review Rating Group

プロジェクト番号/Project No.

13

グループリーダー/Group Leader

小原叶夢 Knamu Obara

グループメンバ/Group Member

伊藤妃雪 Hiyuki Ito

橋本海羽 Miu Hashimoto

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

概要

消費者は商品購入の際、適切な選択を行うためにオンラインのクチコミを参考にすることが多い。特に化粧品業界では、@cosmeが日本最大級のクチコミサイトとして広く利用されており、2000万件以上のレビューが蓄積されている。しかし、レビュー件数の増加により、消費者が必要な情報を効率的に取得することが難しくなり、情報過多の問題が発生している。

本研究では、この問題を解決するためにクチコミデータの可視化を行い、消費者がひと目で商品の特徴やクチコミの傾向を把握できる仕組みの構築を目指した。@cosmeの約138万件のクチコミデータの中から、化粧水に関するレビュー約84,000件を対象にテキストマイニング手法を用いた分析を実施した。データスクリーニングを行い、化粧水のテクスチャーや効果を表す特徴語を抽出し、Pythonを用いて解析した。抽出されたデータを円グラフで可視化し、全体のクチコミ傾向と各商品のクチコミ内容の比較分析を行った。

分析の結果、商品ごとに異なる特徴語が浮かび上がり、消費者の認識が広告の表現と一致する場合としない場合があることが確認された。本研究の結果は、クチコミデータの可視化が消費者の意思決定を支援し、情報収集の効率化に貢献する可能性を示唆している。今後の課題として、抽出する特徴語の範囲を拡張し、リピーターと非リピーターのクチコミの違いを分析することが挙げられる。

1. はじめに

近年、クチコミサイトが急速に普及した結果、商品選択時の参考となるクチコミは膨大な量にのぼっている。奥村(2013)は、ソーシャルメディア(cQAサイト、クチコミサイト、マイクロブログなど)に投稿される膨大なテキストをすべてユーザーが読み切ることは難しく、クチコミの要約が必要であることを強調している。実際、ユーザーが興味のある商品に関するクチコミをくまなく読み込むのは困難であるため、効率的に情報を抽出し、短時間で全体像を把握できる仕組みの整備が求められている。こうしたクチコミの増加と要約の必要性が指摘される一方で、実際に消費者は商品を購入するときに適切な選択のため、各商品のクチコミを読み比べて検討することが多い。佐々木(2004)は、評価サイトは雑誌などに比べ購買の決め手となる情報を提供する情報源としても、商品を理解する情報を提供する情報源としてもその相対的有効度を上げていることが判明したと述べている。相対有効度とはある対象の要素や因子が、他の要素や因子と比較してどれだけ重要であるかを相対的に評価するための指標である。つまり、雑誌などに比べてクチコミは消費者の購買行動に大きく影響を与えることが分かった。クチコミが頻繁に投稿される商品として化粧品が挙げられる。化粧品クチコミサイトの中でも、株式会社istyleが運営するクチコミサイト「@cosme」は日本最大級の化粧品のクチコミサイトであり、クチコミ件数は2,000万件を超えている。クチコミサイトのクチコミ件数が増加する一方、クチコミ件数が多すぎて情報過多に陥り、商品決定を満足に行うことができないという問題があると考えた。そこで大量のクチコミを読む負担を軽減しつつ、ひと目で商品の特徴やクチコミの概要を把握できる仕組みがあれば、自身に合う化粧品を探す時間の短縮につながると考えた。クチコミを読み比べる時間が減ることでユーザーに時間を提供することになり、最終的にユーザーの幸せにつながると予想される。そこで本グループの目的は、@cosmeから提供されたクチコミデータを分析し、クチコミの可視化することでユーザーの幸せにつなげることにした。今回、@cosmeから提供されたデータは約138万件あり、全ての商品を分析するのは難しかったため化粧水のクチコミにカテゴリを絞って分析を行った。化粧水を対象にした理由は、クチコミの文章内にテクスチャや効果が多く含まれ、商品ごとの特徴が明確に抽出しやすいためである。次の章から関連研究、目的、方法、結果、考察の順で説明する。

2. 関連研究

2.1 ソーシャルメディアを対象としたテキストマイニング

奥村(2013)は、ソーシャルメディア(cQAサイト、クチコミサイト、マイクロブログなど)を対象としたテキストマイニングを用いた研究を行い、クチコミの信頼性や有益性を測定する手法を提案した。特に、クチコミサイトの投稿内容を解析し、クチコミの集合の要約をすることで、消費者が求める情報の信頼性や有益性を測定する手法が2つ提案されている。一つは属性ごとの肯定的・否定的意見数を集計する「属性に基づく意見要約」であり、もう一つは、レビューの中から重複を避けつつ重要情報を含む部分集合を選択するという手法である。この研究では、クチコミの特徴語を抽出し、消費者が注目するポイントを可視化することで、購買行動の分析にも活用できる可能性が示唆されている。

2.2 商品購買における評価サイトの有効度

佐々木(2004)は、消費者の購買行動において評価サイト、雑誌、クチコミがどの程度有効な情報源となるかを検証した。特に@cosmeやAmazonレビューなどの評価サイトが、消費者の意思決定にどのような影響を与えるかが分析された。評価サイトの「相対的有効度」(他の情報源と比較した重要度)を分析し、消費者が購買の意思決定において評価サイトをどのように活用しているのかを考察している。特に、評価サイトの情報が商品理解と購買決定の両面において有用であることを示し、従来の広告・メディアに比べて、評価サイトの信頼性や影響力が高まっていることを明らかにした。これは企業のマーケティング戦略において、評価サイトの活用が重要な要素となることを示唆しており、消費者の購買意思決定を支援するためには、信頼性の高いレビューの提供が不可欠であることを強調している。

2.3 口コミが行動変容促進型商品の評価に与える影響

木暮ら(2021)は、行動変容型商品(例:化粧品、ダイエット食品、サプリメント)におけるクチコミの影響を調査し、ポジティブなクチコミとネガティブなクチコミが購買決定に与える影響について分析を行った。特に、ポジティブなクチコミとネガティブなクチコミが購買決定に与える影響を分析し、消費者がどのように情報を解釈し行動を変えるのかについて言及している。例えば、広告では「肌荒れを防ぐ敏感肌特化化粧水」と表現されている場合、クチコミでは「肌トラブル」や「肌荒れ」といった異なる言葉に言い換えられることがあり、この言い換えが消費者の意思決定に影響を与える可能性が指摘されていた。

3.目的

本プロジェクト全体の目的は、生活者の幸せに繋がる提案としてデータ分析を行うことである。そこで、膨大なクチコミデータをもとにデータサイエンスの手法を用いて、生活者の幸せに繋がる知見を得ることを目指した。

生活者の幸せについて考えたとき、クチコミの可視化が実現すれば、消費者は多くの時間をクチコミを読むことに費やすことなく商品選択を行うことができる。そのため、クチコミ可視化班は、クチコミの可視化によって、消費者がひと目で商品の特徴と概要が分かることを目的として分析を行った。

4.方法

4.1 データの選定

本研究では、化粧品に関する評価サイト「@cosme」(<https://www.cosme.net>)を調査対象とした。「@cosme」は1999年12月に開設された日本最大級の化粧品評価サイトの一つである。2023年に投稿されたクチコミデータ1,384,745件を対象とし、その中から「化粧水」に関するデータ83,829件を抽出した。2023年のデータには52,135種類の化粧品が含まれており、製品ごとに性質や用途が大きく異なることを考慮し、本研究の調査対象として「化粧水」2,506種類を選定した。

4.2 特徴語について

4.2.1 特徴語の定義

本研究では、化粧水の「感触」や「効果」を表現する語を「特徴語」と定義する。例えば「さらさら」「しっとり」、「美白」などが特徴語である。これらの語は、消費者が化粧水を選択する際の判断基準となる情報を提供し、クチコミ分析において重要な役割を果たす。

4.2.2 特徴語の抽出方法

特徴語の抽出に際し、まず化粧品関連のWebサイトを参照し、化粧水のテクスチャーや使用感を表現する語句を収集した。特に、ナールスエイジングケア(Nahls)が提供する「化粧品のテクスチャーに関する情報」(ナールスエイジングケア, 2023)を参考にし、一般的に用いられる表現を26語リストアップした。次に、これらの語句の一般的な使用頻度を検証するため、アンケート調査を実施した。SNS上にアンケートを流し、82名に回答してもらった。性別や年代を回答する項目を設けなかったため、年代や性別は不明であるが、女性のグループメンバーの友人達がアンケートに回答していたため、20代の女性が多く回答しているのではないかと推察される。アンケートの結果、使用頻度が10件未満の語句は十分な一般性がないと判断し、分析対象から除外した。結果、「保湿」、「とろみのある」、「しっとり」、「浸透」、「さらさら」、「さっぱり」、「肌荒れ」、「美白」、「もちもち」、「しゃばしゃば」、「伸びが良い」、「みずみずしい」、「ライト」、「肌トラブル」、「かるい」、「リッチ」、「柔らかい」、「重い」、「アンチエイジング」、「ふわふわ」、「こっくり」の21単語を特徴語として選定した。その特徴語を選定したアンケート結果を図1に示す。また、「さらさら」などの言葉は「さらさら」と「サラサラ」というように書き方がクチコミ投稿者によって変わることがあるのでクチコミから特徴語を抽出する前にクチコミ内の表記の違いをPythonを使用して全て統一してから抽出を行った。例えば、「サラサラ」は「さらさら」に統一、「モチモチ」は「もちもち」に統一した。抽出した特徴語を用いて、化粧水のクチコミを分析した。

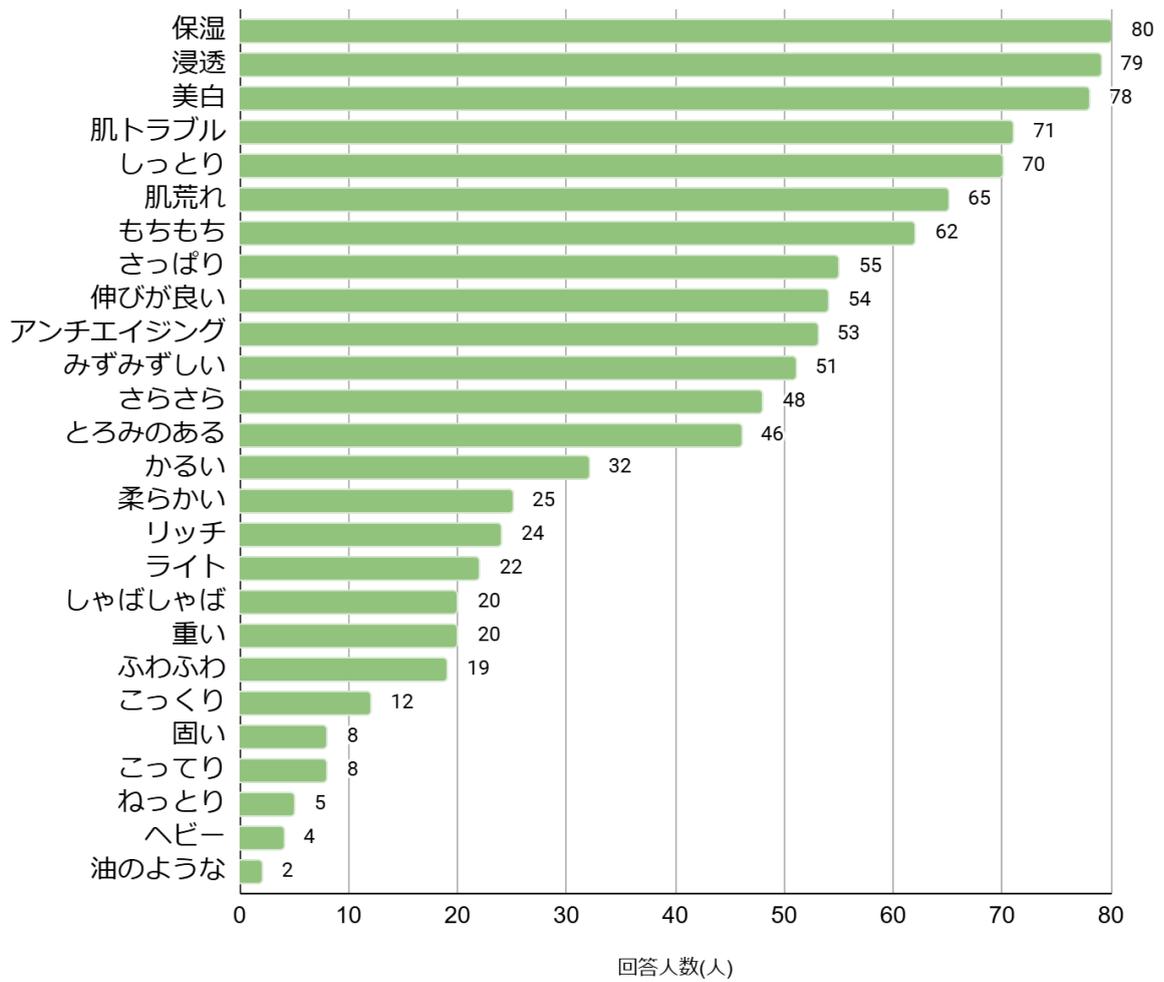


図1.特徴語を選定するために行ったアンケートの結果

4.3 特徴語の可視化

4.3.1 特徴語の集計方法

抽出した特徴語をクチコミデータから抽出し、Bag-of-Words (BoW) 方式で集計した。本研究では、全化粧水のクチコミのうち、ランダムに抽出した10,000件を分析対象とし、加えて個別の化粧水ごとのクチコミデータも併せて集計した。

4.3.2 可視化方法

特徴語の分布を視覚的に把握するため、円グラフを作成した。ただし、化粧水ごとのクチコミ件数には偏りが見られたため、各特徴語の使用頻度を比率で算出し比較を行った。この手法により、クチコミ件数の影響を排除し、特徴語の使用傾向を正確に把握できる。

5.結果

化粧水に関するクチコミデータを対象に、各商品のクチコミにおける特徴語の割合を可視化した。商品の選定方法は、2023年度のクチコミ件数上位5位の商品とした。その結果、以下のような特徴が見られた。

5.1 ランダムに10,000件抽出したデータ

スクリーニングしたデータからランダムに10,000件抽出したデータでは、「保湿」、「とろみのある」、「しっとり」といった特徴語が多くみられた(図1)。「保湿」の割合は19.9%と全体の約1/5を占めている。その次に多いのが、「とろみのある」で13.7%、「しっとり」は14.0%、「浸透」は11.0% となっている。

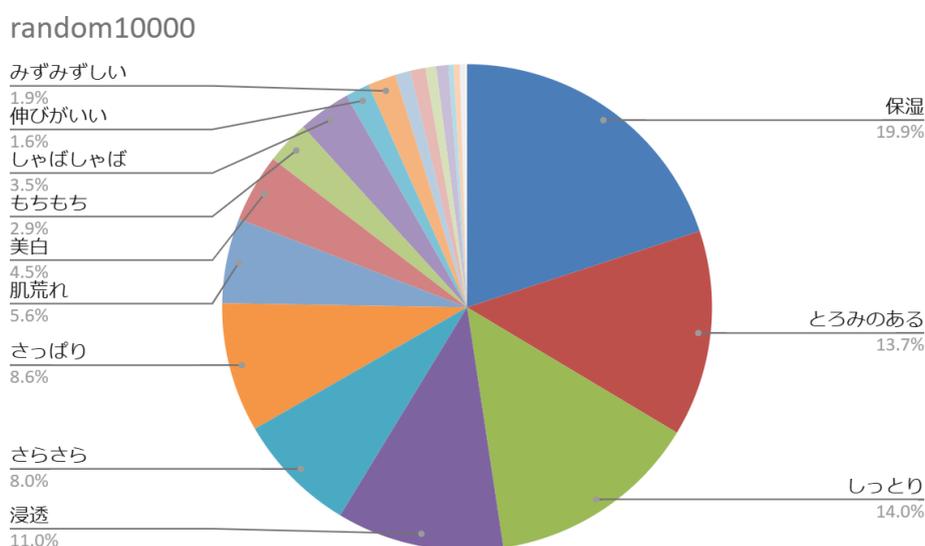


図2.ランダムに10,000件抽出したデータを分析した円グラフ

5.2 エリクシールのブライティングローションWT II

「ブライティングローションWT II」のクチコミ件数は965件だった。クチコミには、「美白」、「ライト」、「透明感」といった特徴語が多く含まれていることが確認された(図2)。特に、「美白」の割合は全体平均よりも高く、全体平均が4.5%であったのに対し、「ブライティングローションWT II」では19.9%の割合を示した。また、「ライト」の出現率は全体平均の1.0%に対し、該当商品のクチコミでは5.7%であった。

エリクシール

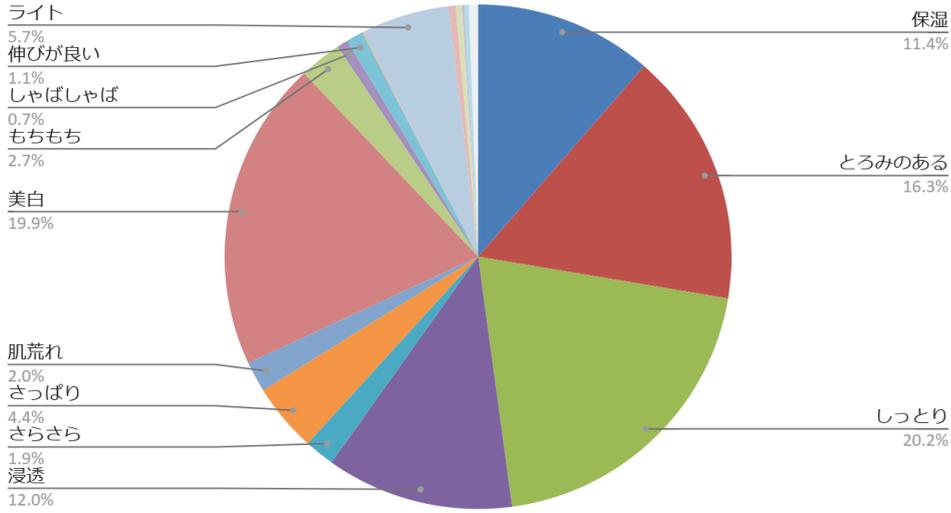


図3.エリクシールのブライトニングローションWT IIを分析した円グラフ

5.3 クレスクのジェリーコンディショナー

「ジェリーコンディショナー」のクチコミ件数は491件だった。クチコミには、「リッチ」、「みずみずしい」、「伸びがいい」といった特徴語が頻出していることが確認された。「みずみずしい」の割合は全体平均の1.9%に対し、該当商品のクチコミでは16.7%と大幅に高い割合を示した。

クレスク

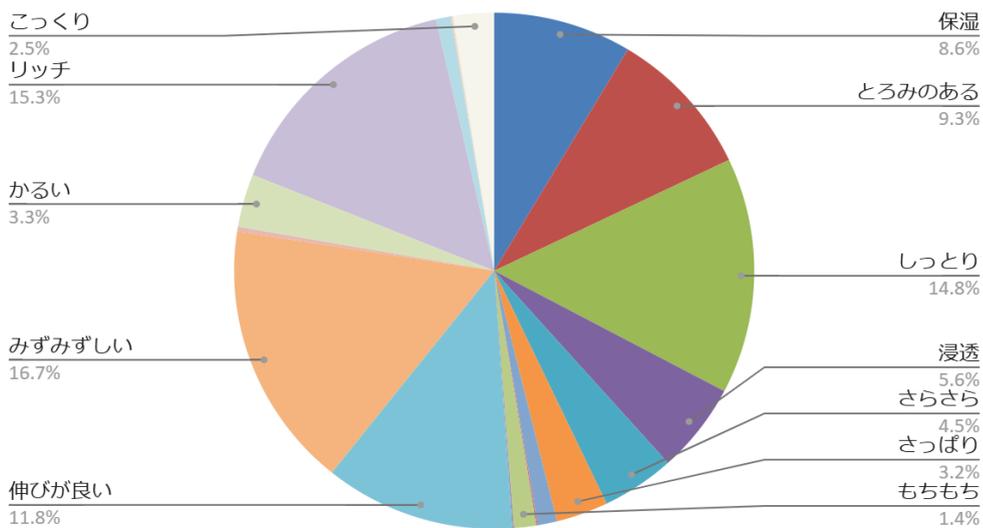


図4.クレスクのジェリーコンディショナーを分析した円グラフ

5.4 SK2のフェイシャルトリートメントエッセンス

「フェイシャルトリートメントエッセンス」のクチコミ件数は789件だった。クチコミでは、「しゃばしゃば」、「浸透」、「もちもち」といったテクスチャーに関する特徴語が頻出した。「しゃばしゃば」の割合は全体平均が3.5%であるのに対し、該当商品のクチコミでは11.3%を示した。特徴語の割合が大幅に増加しているのが分かる。

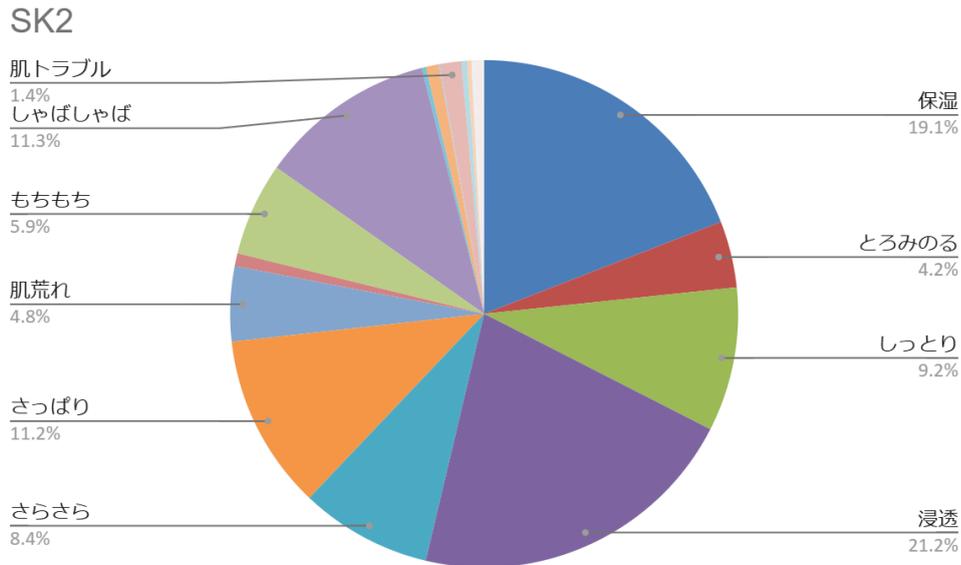


図5.SK2のフェイシャルトリートメントエッセンスを分析した円グラフ

5.5 ドクターシーラボの薬用ローションスーパーセンシティブEX

「薬用ローションスーパーセンシティブEX」のクチコミ件数は80件だった。クチコミには、「さらさら」、「みずみずしい」、「かるい」、「肌荒れ」といった特徴語が多く見られた。「さらさら」の割合は全体平均の8.0%に対し、該当商品のクチコミでは26.6%と、全体平均の割合を大きく上回る結果になった。また、「みずみずしい」の割合も全体平均の1.8%に比べ、9.3%と特徴語が多くみられたことが分かる。

スーパーセンシティブEX

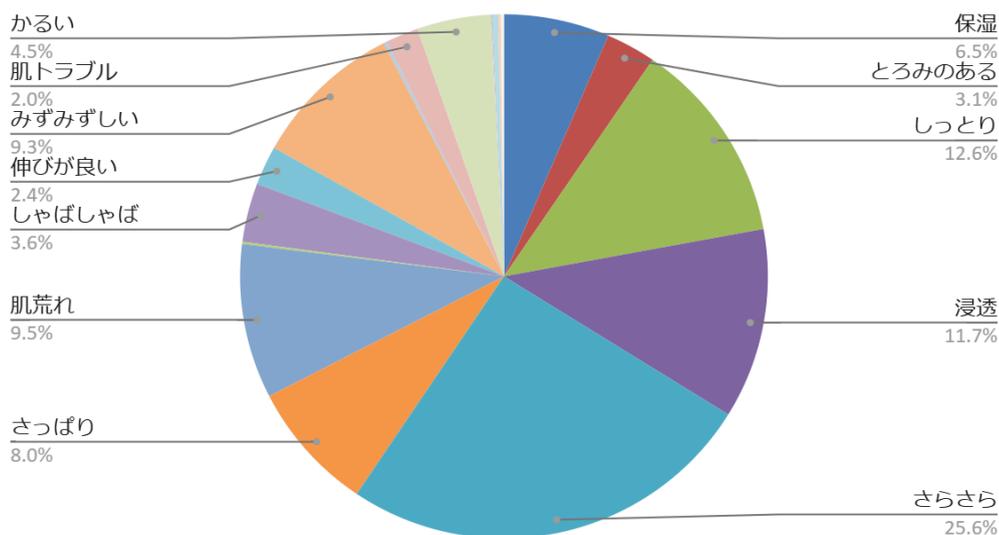


図6. ドクターシーラボの薬用ローションスーパーセンシティブEXを分析した円グラフ

5.6 iPSAのザ・タイムRアクア

「ザ・タイムRアクア」のクチコミ件数は1199件だった。クチコミでは、「しゃばしゃば」、「さらさら」、「肌荒れ」、「保湿」といった特徴語が多く出現した。「しゃばしゃば」の割合は全体平均の3.5%に対し、該当商品のクチコミでは9.5%を示した。また、「さらさら」の出現率も14.0%と、全体平均と比較すると6.0%割合を上回っていることが確認できる。

ザ・タイムRアクア

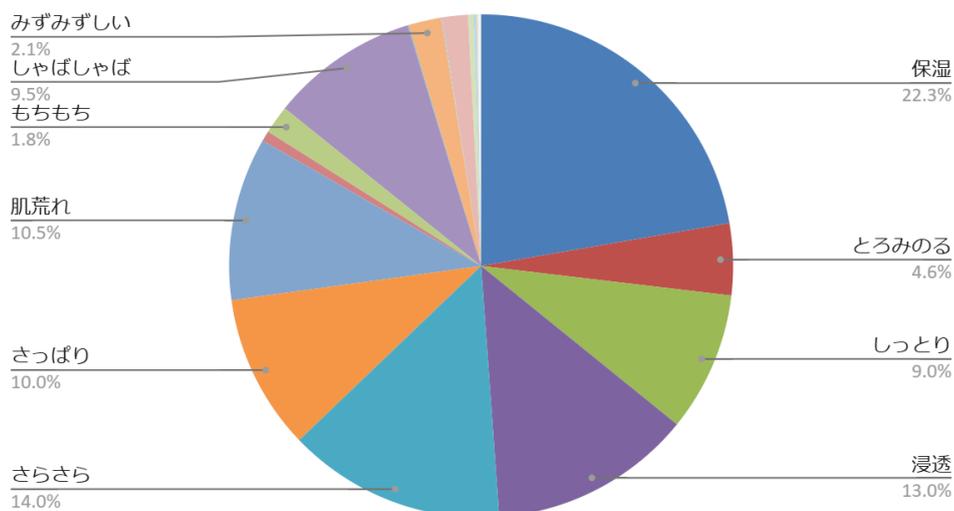


図7. iPSAのザ・タイムRアクアのクチコミを分析した円グラフ

これらの分析結果から、全体平均においては「保湿」「とろみのある」「しっとり」といった保湿効果を示す特徴語が多く出現する傾向が確認された。一方、特定の商品ごとのクチコミデータでは、商品ごとに異なる特徴語が目立っていた。

6章 考察

本研究ではクチコミの可視化を目的とし、クチコミデータをアンケートで選定した特徴語をもとに分析して化粧品全体のクチコミと商品ごとのクチコミの特徴語の割合を円グラフで出力し比較した。その結果、全体平均と商品ごとの割合を比較すると違いがあることが一目で分かるようになった。商品によってよく使われる特徴語には違いがある。この章では、結果をもとにそれらの特徴語が、商品を販売している企業が宣伝している内容と一致しているのかについて考察する。企業が宣伝したい言葉の定義は企業が広告画像に記載されている言葉をそのまま抜き出し、企業が宣伝したい言葉とした。分析から分かった特徴語と広告が宣伝したい言葉を商品のキャッチコピーとして表1にまとめた。

6.1 エリクシールのブライティングローションWT IIについて

全体平均とブライティングローションWT IIを比較すると平均より「美白」、「ライト」の割合が多いことが分かった。実際にブライティングローションWT IIの広告を見ると「美白」、「透明感」、「ハリ」について宣伝している。このことから、「美白」については企業が宣伝したいことがクチコミの内容に反映されていると考えられる。「透明感」と「ハリ」については、今回は特徴語に選定されていなかったことから、その2つの言葉は特徴語として表れなかったと考えられる。

6.2 クレスクのジェリーコンディショナーについて

全体平均とジェリーコンディショナーを比較すると平均より「リッチ」、「みずみずしい」、「伸びがいい」の割合が多いことが分かった。実際にジェリーコンディショナーの広告を見ると「みずみずしいジェリーテクスチャー」について宣伝している。このことから、「みずみずしい」については企業が宣伝したいことがクチコミの内容に反映されていると考察される。加えて、クチコミで使われていた「伸びがいい」という表現は、「みずみずしい」に近い表現であることから商品のテクスチャーについて、企業の宣伝したいことがクチコミに反映されていると考えられる。

6.3 SK2のフェイシャルトリートメントエッセンスについて

全体平均とフェイシャルトリートメントエッセンスを比較すると平均より「浸透」、「しゃばしゃば」、「さっぱり」の割合が多いことが分かった。実際にフェイシャルトリートメントエッセンスの広告を見ると「水分補給」、「保湿」、「ツヤ」、「アンチエイジング」について宣伝している。このことから、企業が宣伝に使われている言葉がクチコミの内容にそのまま反映されていないと考えられる。「ツヤ」と「アンチエイジング」については、今回は特徴語に選定されていなかったことから、その2つの言葉は特徴語として表れなかったと考えられる。また、フェイシャルトリートメントエッセ

ンスの広告はテクスチャーについて言及をしていないため、テクスチャーに関する表現が多い特徴語の中では反映されにくかったとも考えられる。

6.4 ドクターシーラボの薬用ローションスーパーセンシティブEXについて

全体平均と薬用ローションスーパーセンシティブEXを比較すると平均より「さらさら」、「みずみずしい」、「かるい」、「肌トラブル」、「肌荒れ」の割合が多いことが分かった。実際に薬用ローションスーパーセンシティブEXの広告を見ると「肌荒れを防ぐ敏感肌特化化粧水」、「べたつきを最小限に抑えたテクスチャー」について宣伝している。このことから、表現は異なるが肌荒れ改善とテクスチャーについて企業が宣伝したいことがクチコミの内容に反映されていると考察される。「肌荒れを防ぐ敏感肌特化化粧水」という広告の言葉は、「肌トラブル」、「肌荒れ」という言葉に、「べたつきを最小限に抑えたテクスチャー」という広告の言葉は「かるい」、「さらさら」という言葉にそれぞれユーザーが使いやすい言葉に言い換えられていると考えられる。

6.5 ipsaのザ・タイムRアクアについて

全体平均とザ・タイムRアクアを比較すると平均より「さらさら」、「しゃばしゃば」、「肌荒れ」の割合が多いことが分かった。実際にザ・タイムRアクアの広告を見ると「潤い」、「みずみずしい」、「肌トラブル」について宣伝している。このことから、「肌荒れ」については企業が宣伝したいことがクチコミの内容に反映されていると考察される。「潤い」については、今回は特徴語に選定されていなかったことから、特徴語として表れなかったと考えられる。また、ザ・タイムRアクアの広告はテクスチャーについて言及をしていないため、テクスチャーに関する表現が多い特徴語の中では反映されにくかったとも考えられる。

	クチコミで言及	商品のキャッチコピー
ブライティングローションWT II	「美白」、「ライト」	「美白」、「透明感」、「ハリ」
ジェリーコンディショナー	「リッチ」、「みずみずしい」、「伸びがいい」	「みずみずしいジェリーテクスチャー」
フェイシャルトリートメントエッセンス	「浸透」、「しゃばしゃば」、「さっぱり」	「水分補給」、「保湿」、「ツヤ」、「アンチエイジング」
薬用ローションスーパーセンシティブEX	「さらさら」、「みずみずしい」、「かるい」、「肌トラブル」、「肌荒れ」	「肌荒れを防ぐ敏感肌特化化粧水」、「べたつきを最小限に抑えたテクスチャー」
ザ・タイムRアクア	「さらさら」、「しゃばしゃば」、「肌荒れ」	「潤い」、「みずみずしい」、「肌トラブル」

表1.クチコミで言及されている言葉と企業が言及している言葉の違い

以上が各商品の考察である。商品によって、クチコミに広告の内容が反映されている商品もあれば反映されていないと考えられる商品もあった。反映されていない商品は、特徴語が企業が広告で宣伝する効果を表す言葉よりも、テクスチャーに関連する言葉を多く特徴語として選定したことも反映されなかった原因と考えられる。今後の課題として、特徴語を拡張して効果を表す言葉を追加することや、リピーターと非リピーターとの間にクチコミの違いがあるのか分析することなどが挙げられる。また、現実的な実装についても考える必要がある。今回の活動では、図8のような実装を考えた。色味を既存のサイトで使用されている色に合わせて、グラフを限られたスペースに配置し見やすいように配置した。



図8.可視化の実装イメージ

参考文献

- 奥村 学 (2013). 「ソーシャルメディアを対象としたテキストマイニング」. 電子情報通信学会 基礎・境界サイエンス Fundamentals Review, 6(4), 285.
- 木暮 美菜, 諸上 茂光 (2021). ロコミが行動変容促進型商品の評価に与える影響— 自己の行動変容への類推に着目して —. マーケティングレビュー, 2(1), 2021.003
- 佐々木 裕一 (2004). 商品購買における評価サイトの有効度—評価サイトユーザーにおける評価サイト/雑誌/ロコミの有効度比較—. 情報メディア研究, 3(1), 29-42.
- ナールスエイジングケア (2023). 化粧品テクスチャーに関する情報. ナールスエイジングケア公式サイト.
<https://www.nahls.co.jp/eijingukea/cosmetics/eijingukea-cosme/texture/>(閲覧日: 2024年1月8日).