

2022 年度 修士学位請求論文

化粧品クチコミに特化した
信憑性判断支援の研究

明治大学大学院先端数理科学研究科

先端メディアサイエンス専攻

濱野 花莉

概要

肌質や化粧の仕上がりの好みは人によって異なることもあり、多くの消費者の需要に応じて多種多様な化粧品が開発されている。例えばファンデーションのみに絞っても、シミを隠すカバー力を重視したもの、崩れにくさを重視したもの、自然由来の成分にこだわったもの、ツヤのある肌に仕上がるものなど様々な種類の商品がある。

ここで、化粧品の選択を誤ると肌荒れや仕上がりの悪さに繋がるため、化粧品の購入に失敗したくないと考える消費者は多く、インターネット上のクチコミを参考にすることがある。しかし、化粧品のクチコミは本当に利用しているのかを肌の写真からは確認しづらいことや併用した化粧品の相性で仕上がりが変化することなど信憑性の判断を難しくする要因を抱えている。そこで本研究では、化粧品に特化した信憑性評価軸を検討し、その評価軸を用いてクチコミの信憑性を評価するシステムを提案する。本研究の対象は化粧品のクチコミ閲覧時に信憑性を判断することが苦手な人であり、そうした人が信憑性を意識しつつクチコミを閲覧できるよう支援することを目的とする。

まず、化粧をする人への基礎調査を行い、化粧品のクチコミの利用状況やどのような点で信憑性を判断しているのかを調査した。その結果、クチコミの信憑性を判断する基準は文章・画像・投稿者の大きく3つに分類できることを明らかにした。この中でも、文章に関する回答では、人によって信憑性が高いと感じるクチコミの特徴が大きく異なることが明らかになったため、これ以降は文章に着目し、研究を進める。

次に、クチコミの信憑性データセット構築を行い、信憑性評価軸を検討した。データセット構築では、実際にインターネット上に投稿されたクチコミ300件を収集し、どういった記述があるクチコミが信用されやすいのかを調査した。このデータセットを分析し、信憑性評価軸の候補を検討した。

そして、この評価軸の有用性を検証するために、使用状況別にクチコミを323件作成し、評価軸に沿ったアノテーションを付与した。機械学習を用いて81.7%の精度で化粧品の使用状況を判別可能であると示した。

次に、検討した信憑性の評価軸を利用し、化粧品に特化したクチコミ信憑性評価システムを開発した。また、このシステムをユーザに使用してもらうことで、システムを利用した際の信憑性への意識の変化を調査するとともに、システムの改善点について検討した。その結果、元々信憑性を意識していなかった実験協力者が信憑性の提示により信憑性を意識するようになることや信憑性提示の見方にも個人差が観察されたため、その個人差に対応可能なシステムにする必要があることが明らかになった。また、得られたフィードバックからシステムの改良を行い、再度システムの評価実験を行った。2度目の実験では、システム利用中に画面録画を行い、システム内の機能をどのように使用しているのかを調査した。

本研究の提案システムは化粧品に特化したものであるが、他のジャンルのクチコミへの応用可能性や化粧品のクチコミがもつ独自性について議論するためにアンケート調査を行った。調査の結果、20～24歳の若い世代では化粧品のクチコミを頻繁に閲覧する人が多く、閲覧にはSNSが利用されやすいことや、フォロワー数の多いユーザーによるクチコミが参考にされやすいといった化粧品のクチコミがもつ特性が明らかになった。この結果をふまえて、今後はSNSでのクチコミに対応可能なシステムにしていくことやフォロワー数の多いインフルエンサーのクチコミの扱い方に関する検討が必要であることが示された。また、参考にするクチコミの特徴はどのジャンルにも共通するものが多く、現在検討した評価軸を他に応用できる可能性が示唆された。

目次

第1章	はじめに	1
1.1.	化粧の日常性と化粧品の購入における困難	1
1.2.	化粧品についての情報収集と信憑性	1
1.3.	化粧品のクチコミ特有の問題	2
1.4.	本研究の目的	2
1.5.	本稿の構成	3
第2章	関連研究	4
2.1.	化粧に関する研究	4
2.2.	クチコミに関する研究	5
2.2.1.	クチコミを用いたマーケティング	5
2.2.2.	クチコミの分析	6
2.2.3.	化粧品のクチコミ	7
2.3.	信憑性に関する研究	7
2.4.	マイクロタスクに関する研究	8
第3章	化粧をする人を対象とした基礎調査	10
3.1.	調査目的	10
3.2.	調査概要	10
3.3.	結果	12
3.4.	KH Coderによる分析	14
第4章	文章に着目したデータセット構築	17
4.1.	化粧品の選定とクチコミの収集	17
4.2.	ユーザによる信憑性判断	18
4.3.	結果と考察	19
4.3.1.	データセット構築者ごとの結果・考察	19
4.3.2.	クチコミごとの結果・考察	21
4.3.3.	信憑性評価についての結果・考察	22

4.3.4.	商品ごとの結果・考察	25
第5章	信憑性評価軸の提案と有用性検証	30
5.1.	信憑性評価軸に関する検討	30
5.2.	使用状況別のクチコミデータセット構築	30
5.3.	クチコミに対する軸評価データセット構築	33
5.3.1.	概要	33
5.3.2.	分析・考察	35
5.4.	機械学習を用いた判定精度の検証	36
5.5.	信憑性評価軸についての議論	38
第6章	プロトタイプシステムの構築と評価実験	39
6.1.	システム概要	39
6.2.	クチコミに対する軸評価	41
6.3.	クチコミの信憑性スコア算出	43
6.4.	システム評価実験	43
6.4.1.	実験概要	43
6.4.2.	事前アンケートの結果	43
6.4.3.	事後アンケートの結果	45
6.5.	考察	47
第7章	システムの改良と評価実験	48
7.1.	システムの改良	48
7.2.	実験概要	49
7.3.	結果	49
7.3.1.	事前アンケートの結果	49
7.3.2.	画面収録を用いた分析と結果	50
7.3.3.	事後アンケートの結果	51
7.4.	考察	53
第8章	本システムの応用可能性の議論	54
8.1.	大学生・大学院生を対象とした予備調査	54
8.1.1.	調査概要	54
8.1.2.	結果	54

8.2. Yahoo!クラウドソーシングを利用した調査.....	58
8.2.1. 調査概要.....	58
8.2.2. 結果.....	60
8.2.3. 考察.....	64
第9章 制約と展望.....	66
第10章 おわりに.....	67

第1章 はじめに

1.1. 化粧の日常性と化粧品の購入における困難

顔は全身の魅力評価における重要な要素であり[1], 化粧をすることで顔の印象を大きく変化させることができる[2]. そのため, 化粧は理想の自分を実現する手段として用いられており, どのような顔になりたいかによって, 眉やアイラインの描き方の調節や, アイシャドウやチーク, リップの色や質感など様々な要素を使い分ける必要がある. また, 仕事やデート, 女子会といった日常の場面に合わせて化粧を変化させる人も多く, 実際に雑誌などでTPOに合わせた化粧に関する特集記事が組まれている[3]. そのほかにも, シミやそばかす, 傷跡など自身のコンプレックスを隠すために化粧をする人もいる.

このような様々な理由から, 多くの人にとって化粧は日常的なものである. ポーラ文化研究所が2021年に行った調査[4]によると女性の81.8%, 男性の16.6%が化粧を行っており, スキンケアに関しては女性の93.1%, 男性の55.5%が行っている. 消費者の多種多様な要望に応えるため, 化粧品メーカーは様々な特徴をもった化粧品を開発・発売している. 例えば, ファンデーションであれば, 皮脂や時間経過による崩れを防止するもの, マット肌・ツヤ肌など質感にこだわったもの, 肌に優しい素材を使って作られたものなどがある. また, 化粧は質感, 色合いやメイクの方法等において流行の移り変わりが激しく, それに対応した新商品の発売も頻繁に行われる. こうした背景から, 現在化粧品売り場では膨大な種類の化粧品が販売されており, その中から自身に合う化粧品を見つけるのは困難である.

ここで, 化粧品は継続的に直接肌につけるものであるため, 肌に合わない化粧品を使用すると肌荒れを引き起こすことがある. また, 値段も安くはないため, 購入に失敗したくないと考える消費者は多い. しかしNTTコムオンライン・マーケティング・ソリューション株式会社が2022年に行った調査[5]によると, 化粧品の購入において「失敗した」と感じたことがある人は60.6%にのぼるなど, 化粧品選びは容易ではないことがわかる.

1.2. 化粧品についての情報収集と信憑性

化粧品の購入に失敗しないために, 情報収集は重要である. 化粧品に関する情報を得る手段として, 店舗で商品の実物やPOPを見ることの他に, 友人や家族, 化粧品売り場の美容部員からのアドバイスを受けること, 化粧品の公式サイトや雑誌・テレビを見ること, クチコミサイトやECサイト, SNSなどインターネット上に投稿されたクチコミを見ることなどが挙げられる.

特にインターネット上のクチコミは多くのユーザの意見を手軽に閲覧することができるという利点があり, @cosme[6]やLIPS[7]といった化粧品専門のクチコミサイトや, Twitter[8], Instagram[9]などのSNSに多くのクチコミが投稿されている. しかしその一方で, 誰でも書

き込むことができるため、質の低いクチコミやステルスマーケティングによるクチコミが投稿されている可能性がある。質の低いクチコミの例としては、専門知識をもたないユーザによる正しい知識に基づかないクチコミ、SNS で自分の投稿が話題になってほしいと考えるユーザによる誇張表現を含むクチコミなどがある。こうした背景から、化粧品の購入に失敗しないために、クチコミサイトや SNS に投稿されたクチコミの信憑性をユーザ自身で見極める必要がある。

1.3. 化粧品のクチコミ特有の問題

これまで、クチコミの信憑性判断に関する研究[10]やシステムの開発[11]が行われてきたが、これらは化粧品に特化したものではない。著者は、化粧品のクチコミがもつ多くの特性を考慮したうえで、化粧品に特化したクチコミの信憑性判断支援が必要であると考えた。ここで、化粧品のクチコミがもつ特性とそれによって引き起こされる問題を挙げる。まず、化粧下地の上にファンデーションやフェイスパウダーを重ねるなど、同じ部位に複数の化粧品を併せて使うことがある。そのため、例えばファンデーションのクチコミであっても、その効果がファンデーションによるものなのか、併用した他の製品によるものなのかが判断できないといった問題が起こりうる。また、化粧品の単体での効果だけでなく、併用した化粧品によっては相性の良し悪しが効能や仕上がりに影響を及ぼすこともある。さらに、化粧品は投稿者の肌質や好みなどの個人差によってクチコミの意見が大きく変わることや、投稿者が本当にその化粧品を使っているのか、どの程度使っているのか、どのように使ったのかなどをクチコミだけでは判断できない。特にスキンケア化粧品では長期的に使用することで効果が得られるものも多いが、試用品の配布や数日分のミニサイズで販売が行われていることもあり、短期間の使用によって正当な評価のされていないクチコミが投稿されることも珍しくない。

このように、化粧品のクチコミには特有の問題があるため、特に信憑性判断の支援が必要である。しかし、クチコミサイトや SNS などインターネット上の情報を真実かどうか完全に判断することは現実的でない。こうした問題を踏まえ、本研究では客観的な信憑性判断に加え、信憑性判断が得意なユーザの評価基準を共有する、化粧品に特化したクチコミの信憑性判断支援システムを提案し、このシステムを用いて信憑性判断が苦手なユーザを支援することを目指す。

1.4. 本研究の目的

本研究では、化粧品購入の際に参考にされることの多いクチコミについて、誰でも投稿できるため信憑性が低いうえ、化粧品特有の性質により信憑性判断が難しいという問題に着目する。そこで、本研究では信憑性判断が苦手なユーザを支援することを最終的な目標とす

る。この最終目標に向け、クチコミの信憑性データセットを構築・分析し、信憑性の評価軸を検討する。評価軸を用いてクチコミの信憑性判定精度を確認したのち、この評価軸を利用した信憑性判断支援システムを構築する。

1.5. 本稿の構成

本稿は、本章を含む全10章から構成される。まず本章で化粧品のクチコミがもつ特性とそれに起因する問題点について述べた。これ以降、2章では化粧品及びクチコミ分析という観点から関連する研究を取り上げ、本論文の位置付けを明確化する。続いて3章では、化粧品に対する意識や化粧品のクチコミの利用状況を調査するために行った化粧品をする人を対象とした基礎調査とその結果について述べる。4章では、基礎調査の結果をもとに文章に着目した化粧品のクチコミデータセット構築を行い、その分析結果について述べる。5章では、データセットの分析結果から検討した信憑性評価軸の候補について述べたのち、候補となる評価軸を用いた時の信憑性判定精度評価実験を行い、その結果と新たに検討した評価軸について述べる。6章では、検討した評価軸を用いてプロトタイプシステムを実装し、評価実験を行い、その結果を述べる。7章では、6章の実験を踏まえ改善を行い、提案システムを用いた評価実験について述べる。8章ではクチコミに関する調査を実施し、システムの応用可能性について議論を行う。9章では提案システムの制約と今後の展望について述べ、最後に第10章で本稿のまとめを行う。

第2章 関連研究

2.1. 化粧に関する研究

化粧は顔の見た目を大きく変化させるため、それに伴う印象の変化に関する研究が行われている。九島ら[2]は、化粧が対人関係に及ぼす影響について検討している。この研究では、まず基準となる顔画像と成熟・性別特徴を踏まえた顔画像を作成、そしてメイクについても同様に成熟・性別特徴を踏まえたものを作成した。その後、それぞれ特徴の異なる顔・メイクを組み合わせることで、どの特徴の顔でもメイクの成熟・性別特徴の印象に影響を受けることを確認した。また、加藤ら[12]は、目と眉の間隔に着目し、間隔の違いが顔の印象に及ぼす影響について調査している。この研究ではまず、女子大学生41名の顔画像から平均顔を作成し、その平均顔の目と眉の間隔を操作した画像を複数枚用意した。これらの画像に対しSD法を用いた印象評価を実施した。調査の結果、目と眉の間隔は洗練度や女性らしさといった印象に影響を与えることを示し、眉メイクやアイメイクへの応用可能性について述べている。このように化粧は顔の印象に大きく影響をするため、自身が目指す顔や印象に合う化粧品を選んで使用することが重要である。化粧によって自身の理想像を追求することや、使用する化粧品を変えるだけで簡単にイメージチェンジを楽しむこともできる反面、使用する化粧品を間違えると意図しない印象を与えてしまう可能性がある。そのため、理想とする印象を実現できる化粧品を正しく選択することが必要となる。本研究では、化粧品の購入における失敗を防止し、こうした化粧の楽しさをより多くの人に感じてもらうことを目指す。

化粧は多くの人にとって日常的なものであるため、化粧支援に関する研究は数多く存在する。高木ら[13]はメイクアップテクニックの説明やユーザが自分の顔に行ったメイクアップに対するアドバイスを出力するシステムを提案している。このシステムでは、顔データを3次元で扱っており、正面からだけでなく横や斜めから見た顔のイメージも考慮できる。また神武ら[14]は、個人の好みを反映したメイクアップ支援のためのシステムを提案している。このシステムは、ユーザの顔画像と好みの顔画像を入力することで、ユーザの肌の色と好みを考慮したメイクアップを顔に施したシミュレーション画像を出力し、そのメイクアップを実現するための商品を推薦してくれる。Chongら[15]は、トランスジェンダーのための化粧支援システムを構築している。このシステムでは、顔写真と認識されたい性別を入力することで、化粧のシミュレーションと他人から自分の望む性別として認識される可能性をスコアとして提示する。Kajitaら[16]は、リアルタイムでファンデーションの塗布状態を可視化するシステムの実現を目指し、スマートフォンのカメラで撮影した写真から機械学習を利用してファンデーション塗布画像と素肌画像を82.3%の精度で判別可能であることを明らかにした。Jiyoung[17]は、消費者の好みに合う商品を推薦することで満足度を高めることを目指し、ビッグデータに基づいた化粧品推薦システムを開発した。このシステムでは、化

化粧品を 110 のクラスターに分類し、各化粧品がどの肌合うかをスコアリングしたデータを用いた。本研究では化粧品使用時の支援ではなく、Jiyoung と同様に購入する化粧品を選ぶ段階からユーザの支援を行うことを考えている。

2.2. クチコミに関する研究

2.2.1. クチコミを用いたマーケティング

クチコミはマーケティング戦略として注目されており、インターネット上のクチコミについての研究も行われている。Trusov ら[18]は、クチコミによるマーケティングとイベントマーケティングやメディア出演といった従来のマーケティング手法を比較し、クチコミによるマーケティングは従来手法よりも影響力が強く、長く持続すると結論づけた。Cheung ら[19]は、インターネット上のクチコミがもつ特徴とユーザが知覚する信憑性の関係や、知覚した信憑性が購買への意思決定に及ぼす影響について調査、検討を行った。その結果、クチコミがもつ様々な特徴がクチコミの信憑性に影響を与え、信憑性が高いと認識されたクチコミが購入の意思決定にプラスの影響を与えることが示された。しかし、否定的なクチコミや二面的なクチコミは信憑性が高いと認識されるという仮説は支持されなかった。また You ら[20]は、過去の研究で報告されたクチコミのもつ影響力、製品特性、業界特性、プラットフォーム特性に関するデータを調査し、インターネット上のクチコミが売上に与える影響についてメタ分析を行った。その結果、クチコミが売上に与える影響が、製品、業界、プラットフォームの特性に依存することを示した。さらに、製品の特性や競争の激化、否定的な評価を含めることでクチコミのもつ影響力に変化を与えるといった、過去の研究とは対照的な知見を得た。Heejae ら[21]は、インターネット上のクチコミが消費者の意思決定に与える効果について、事前態度やクチコミの内容によってどのように変化するかを調査した。調査の結果、良好な態度のユーザが肯定的なクチコミを受け取ったときに商品を選択する確率が増加することや、否定的なクチコミをメッセージ受容性の高いユーザが閲覧すると、商品を選択する確率が減少することを明らかにした。Brandes ら[22]は、気候とクチコミの関係性を調査し、悪天候の日はクチコミ件数と文章量が増加することとクチコミ内容がネガティブになることを示した。この結果をふまえて、企業は天候などのコンテキストを利用し、クチコミの投稿を促すなどのマーケティングアプローチが可能になると主張している。また、Sokolova ら[23]は YouTube や Instagram における美容・ファッション系インフルエンサーによるマーケティングに着目し、投稿者の魅力や自身との類似性が投稿に対して知覚する信頼度やパラソーシャルインタラクション（投稿者と対話しているかのように感じる感覚。以下 PSI とする）に及ぼす影響や、その信頼度や PSI が購入意思に及ぼす影響について調査した。その結果、自身との類似性と PSI および信頼度や PSI と視聴者の投稿意思に正の相関があることを明らかにした。Castillo ら[24]は、化粧品・スキンケア業界におけるインフ

ルエンサマーケティングが、特に10代から20代前半のフィリピン人に対し有効であることを示した。これらの研究でも明らかになっているように、インターネット上のクチコミは従来のマーケティングよりも影響力が強く、クチコミがもつ様々な要因によって消費者の購入時の意思決定に大きな影響を及ぼしている。したがって、消費者がクチコミの信憑性を適切に判断することが可能であれば、消費者が自分に合う化粧品を購入することが容易になると考えられる。本研究では、消費者がクチコミを適切に扱うために必要となる、化粧品におけるクチコミの信憑性を評価する基準を提案する。

2.2.2. クチコミの分析

また、クチコミの信憑性・信頼性や有用性に着目し、分析を行った研究も存在する。Leeら[25]は、Instagramにおけるインフルエンサらのプロモーション投稿について、スポンサー情報の開示・投稿者の信頼性・ブランドの信頼性の観点からプロモーション効果への影響を調査した。その結果、インフルエンサ自身の信頼性はプロモーション効果に影響せず、ブランドの信頼性が影響することを明らかにした。また、Oh[26]はスキンケア製品のネット通販におけるレビューについて、レビューの特徴と有用性の関係を調査した。その結果、レビューの長さ・投稿された時期・肌トラブルについての言及などが有用性に影響を及ぼすことを明らかにした。Rofiantoら[27]は、動画でのクチコミに着目し、開封動画において信頼性や有用性の形成とクチコミ採用率の関係について調査した。その結果、発信者の専門性や信頼度から形成される動画の信頼性は採用率に影響を与えるが、動画内での論証品質がもたらす動画の有用性は採用率に影響しないことを明らかにした。さらに、Mukherjeeら[10]は、ECサイトのレビューアークラップから偽のレビューアークラップを検出することを目指し、ラベル付きデータセットの作成、偽のレビューアークラップを判別するための指標や偽のレビューアークラップを検出するためのアルゴリズムの提案を行った。この研究では、グループ単位で偽のレビューアークラップを検出することは、個々のレビューに注目して偽のレビューアークラップを検出するより容易であることも明らかにしている。Minら[28]は、YouTubeにおけるインフルエンサらの動画に対し視聴者が感じる信頼性に影響を与える要因や、動画内で言及された商品・ブランドに対する意識の変化について調査した。その結果、インフルエンサがもつ専門性や信頼性が、動画に対する信頼性に影響し、動画の信頼性を高く認識すると動画の楽しさやブランドの魅力を感じやすくなることを明らかにした。Learら[29]は、クチコミの物語性に着目し、クチコミの物語性の程度について分析を行い、消費者行動との関連を調査した。その結果、魅力的で説得力のあるクチコミは物語性が高いことを示した。しかし、これらの研究は化粧品のクチコミの特性を考慮したものではないため、他のカテゴリとは異なる独自の特性をもつ化粧品のクチコミについて信憑性を評価する際には不十分である可能性がある。本研究では、化粧品のクチコミの特性を踏まえたうえで、信憑性を評価する基準を提案することを目指している。

2.2.3. 化粧品のクチコミ

本研究で着目する化粧品のクチコミについても、多く分析が行われている。Sumita ら[30]は、化粧品メーカーがインターネット上で製品を宣伝する際に使用しているキーワードとクチコミサイトで、消費者が使用しているキーワードがどのように重なり合うかについて検討を行った。この研究では日本の化粧品市場に焦点を当て、化粧品専門のクチコミサイトである@cosme[6]からクチコミを収集している。Pugsee ら[31]は、化粧品に対するクチコミから、ポジティブな内容のものとネガティブな内容のものを自動分類する手法を提案した。この手法では、まずクチコミの単語にそれぞれ品詞をタグ付けし、その単語を化粧品に特化した感情語彙のリストである the cosmetic sentiment lexicon に含まれる語彙と比較することで、そのクチコミが化粧品についての情報のみで構成されているものか、主観的な表現が含まれているものかを識別する。その後、主観的なクチコミから形容詞、副詞、一部の動詞を特徴量とした分類モデルを作成している。また、Matsunami ら[32][33]は、効果的な化粧品の推薦を目指し、化粧品のクチコミの文章から、効果について算出する自動スコアリング方式の開発に取り組んだ。そして、このスコアリング結果を用いて効果機能タグ推薦手法の提案、開発を行った。Thinukool ら[34]は、インターネット上のクチコミを分析し、化粧品の顧客満足度に影響を与える要因について調査した。この研究では、様々なソーシャルネットワークサイトから化粧品のクチコミを収集し、データクレンジングと線形正規化を行った後、文章中の頻出単語を可視化するワードクラウドという手法を用いてテキストマイニング分析の結果を可視化している。分析の結果、製品に対するクチコミには製品を販売する企業にとって広告やキャンペーンなどに応用できる有益な言葉を含んでいることを明らかにした。これらの研究では、主にマーケティングへの応用を目的としてクチコミについて分析している。本研究の最終目的である化粧品のクチコミの信憑性評価基準の提案も、これらの研究と同様にマーケティングや適切な化粧品推薦といった応用が可能であると考えられる。

2.3. 信憑性に関する研究

インターネットの普及により、ウェブ情報の信憑性を評価することは重要な課題となっている。Wathen ら[35]は、インターネット上の情報の信頼性に着目し、文献レビューを行った。その後、ユーザが情報の信頼性を判断する方法のモデルを提案した。藤本ら[36]は、災害時など不測の事態においても臨機応変に信憑性の高い情報を収集することを目指し、Twitter[8]上で信頼性の高いユーザを特定する手法を提案した。これは、過去の投稿と情報のリアルタイム性をもとにユーザを分類する手法であり、検証の結果、信頼性の高いユーザを特定できることを明らかにした。また、Johnson ら[37]は、政治における SNS の影響力に着目し、SNS 上の政治情報に対する信憑性の認識や SNS に対する依存度の関係を調査した。その結果、SNS は他の情報源に比べ信憑性が低いと評価されている一方、ニュースや政治

情報を SNS に依存しているユーザは、比較的 SNS を信用しやすいことを明らかにした。Moran ら[38]は、SNS 上のクチコミに着目し、クチコミの信憑性に関する文献の再検討を行ったうえで、オンライン環境におけるクチコミの信憑性を構成する要因をまとめたフレームワーク 4Cs(Community, Competence, Content, Consensus)を提案した。このフレームワークは、クチコミの分析を支援し、クチコミをより有意義なものにできると考えられる。また、池田ら[39]は、ソーシャルメディア上でのデマの拡散を防止するために情報拡散のメカニズムを理解することを目標とし、エージェントベース情報拡散モデル (AIDM) を提案した。この提案モデルを用い、デマ拡散の再現実験を行った結果、拡散のピークが一度だけ存在するシングルバースト型デマ拡散の再現に成功した。中島ら[40]は、ブロガーのマニア度を算出するマニア指向ブログランキング方式を提案した。この手法を用いることで、情報を閲覧したユーザ自身がその信頼性を判断することが可能になると主張している。Huang ら[41]は、誇張された情報や実証されていない情報を含む広告を問題視し、そういった広告を判定するために畳み込みニューラルネットワークを利用したシステムを提案し、高精度に判定できることを示した。Nakamura ら[42]は、検索エンジンの利用状況に関する調査と、ウェブ検索を行った際の検索結果に対する信憑性を判断する要因について分析を行った。さらに、信憑性に関わる要因となる検索結果の情報を計算し表示することで、ユーザ自身が信憑性を判断するためのシステム開発も行っている。山本ら[43][44]は、分析対象のウェブ情報とその関連情報をデータ対として表現し、support 関係を分析することで対象となる情報の信憑性を評価するモデルを提案した。この提案モデルの有効性を確認する評価実験では、適合率 65%以上で自動判定が可能であることを明らかにした。また、正確なウェブ情報を収集する能力である「ウェブアクセスリテラシー」を測定する尺度を提案し、この尺度の有効性についても確認している。さらに、情報の信憑性を判断する際、ユーザはウェブページ作成者がどのような人物かということにあまり注意を払わないことも明らかにしている。本研究は、こうしたインターネット上の情報の中でも化粧品のクチコミに特化して信憑性を判定していくものであるが、これらの手法と組み合わせることでさらに有用なシステムの実現につながると考えている。

2.4. マイクロタスクに関する研究

本研究で構築するシステムではマイクロタスク的にクチコミの特徴をユーザに評価してもらうことを検討している。ここで、マイクロタスクを用いた研究は様々行われている。Iwamoto ら[45]は、クラウドソーシングにおけるワーカ確保という問題に着目し、床面にタスクを提示することで、通行人がタスクに容易にアクセスできる手法を提案し、データの品質を管理するための対策を検討した。その結果、通行人の観察結果に基づいた分類器によってデータの品質を向上させた。また、三輪ら[46]はマイクロタスクを行う機会の少なさに着目し、参加を促す手法として、音楽ゲームに埋め込むことを提案した。その結果、ゲームを

邪魔することなくマイクロタスクへの回答を促すことを可能とした。一方、佐々木ら[47]はクライアントの要望に則したイラストを、マイクロタスクを用いて生成する手法を提案している。この手法では、描画能力を問わない複数人のワーカが少しずつ線を描画することで、一枚の写真からイラストの生成が可能であることを示した。また、三原ら[48]は、マンガのコマ領域の判定を目的として、画像認識とクラウドソーシングを組み合わせたマイクロタスクを実装した。これにより、マンガメタデータ作成にかかる手作業の工数を削減し、従来の計算機のみでのコマ抽出よりも精度を向上させることに成功した。著者はユーザがクチコミを表示した際、マイクロタスク的に評価軸にまつわる簡単な質問を提示し、その結果を集約・蓄積することで信憑性評価を行うことを検討している。

第3章 化粧をする人を対象とした基礎調査

3.1. 調査目的

山本ら[43][44]や Cheung ら[19]は、人は情報の信憑性についてどのように評価しているのか、どのような要因に影響されるのかということをも明らかにしてきた。しかし、化粧品の場合は1.3節で述べたような特性があるため、通常のインターネット上の情報とは信憑性の評価基準が異なると考えられる。そこで、クチコミの信憑性について分析していくにあたり、まず化粧をする人が化粧に対してどのような印象をもっているのか、化粧品を購入する際にどのような情報を重視しているのか、何により信用できると判断するのかといった化粧品に関する事前調査を行った。この調査では、化粧に対してポジティブな感情をもっているのか、あるいはネガティブな感情をもっているのかといった根本的な化粧に対する意識も調査し、それらの意識と化粧品のクチコミの利用状況の関連についても分析する。

3.2. 調査概要

本調査は、2020年4月24日～25日に「化粧をする人を対象としたアンケート」として、2000名(男性622名、女性1318名、不明60名)を対象にYahoo!クラウドソーシング[49]を用いて実施した。アンケート項目を表1に示す。なお、Q3ではCOVID-19の流行前の状況での頻度を回答するよう指示した。

表 1 化粧をする人を対象とした調査のアンケート項目と回答形式

Q1	性別を回答してください	3 択 (男性, 女性, 不明)
Q2	年齢を回答してください	6 段階 (10 代, 20 代, 30 代, 40 代, 50 代, 60 代以上)
Q3	普段メイクをしますか?	5 段階 (毎日する (週 7 日程度), ほぼ毎日する (週 5 日程度), 時々する (週 3 日程度), あまりしない (週 1 日程度), ほとんどしない (週 1 日未満))
Q4	化粧をするのは好きですか?	5 段階 (とても好き, 好き, どちらでもない, 嫌い, とても嫌い)
Q5	あなたにとって化粧とはなんですか?	自由記述
Q6	化粧の仕方をどうやって学びましたか?	自由記述
Q7	あなたが化粧品の情報を得る手段をすべて回答してください	8 択 (友人・家族などからの口コミ, テレビCM, 紙媒体(雑誌)の情報, SNS(Twitter, Instagram, Facebook 等), YouTube, 口コミサイト (アットコスメや LIPS 等)でのレビュー, 特に参考にしているものはない, その他 (自由記述))
Q8	「あなたが化粧品の情報を得る手段」で選択した情報を利用する理由をすべて回答してください	9 択 (手軽だから, 見やすいから, 調べやすいから, 商品が魅力的に感じられるから, 様々なユーザのレビューを見ることができるから, 信頼できる情報だから, 商品について詳しく知ることができるから, 意識しなくても情報が入ってくるから, その他 (自由記述))
Q9	化粧品を買うときに参考にする情報をすべて回答してください	8 択 (Q7 と同様)
Q10	「化粧品を買うときに参考にする情報」で選択した情報を利用する理由をすべて回答してください	9 択 (Q8 と同様)
Q11	化粧品についてのインターネット上のレビューで信頼できるものの特徴を回答してください	自由記述
Q12	化粧品についてのインターネット上のレビューで信頼できないものの特徴を回答してください	自由記述
Q13	過去に口コミを参考にして失敗した経験があれば回答してください	自由記述

3.3. 結果

今回の調査では不真面目な回答者が含まれたため、そうした回答者を除外する必要がある。ここで本調査は、化粧をする人を対象としたものであるため、Q3で「ほとんどしない（週1日未満）」と回答した426名、自由記述で「化粧をしない」と記入した2名、質問とは無関係な単語や意味のない文字列を記入した11名、すべての質問で同一の回答をしていた13名、回答の辻褄が合っていない1名の計453名を不適切回答者と判断し、分析から除外した。除外した結果、1547名（男性283名、女性1224名、不明40名）が分析対象となった。アンケートの回答の一部を図1~4に示す。なお、以降の図表内では、「紙媒体（雑誌の情報）」を雑誌、「SNS（Twitter、Instagram、Facebook等）」をSNS、「口コミサイト（アットコスメやLIPS等）でのレビュー」を口コミサイト、「特に参考にしているものはない」を特にないと表記している。

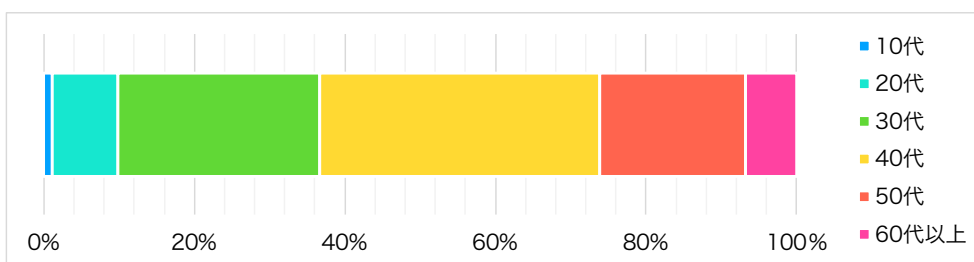


図1 実験協力者の年齢分布 (Q2)

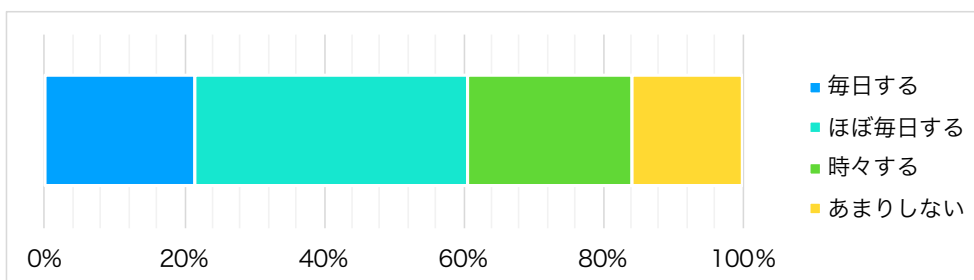


図2 化粧をする頻度 (Q3)

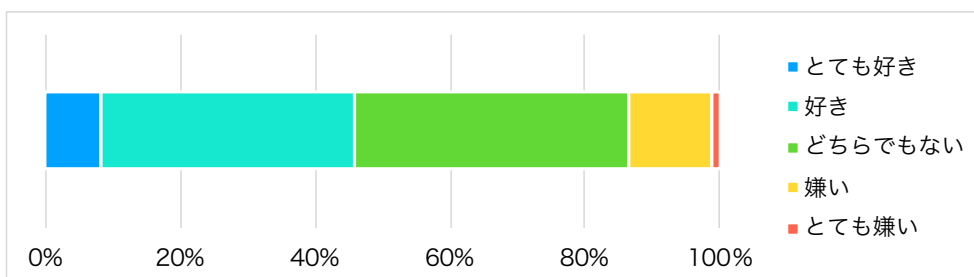


図3 化粧に対する好き嫌い (Q4)

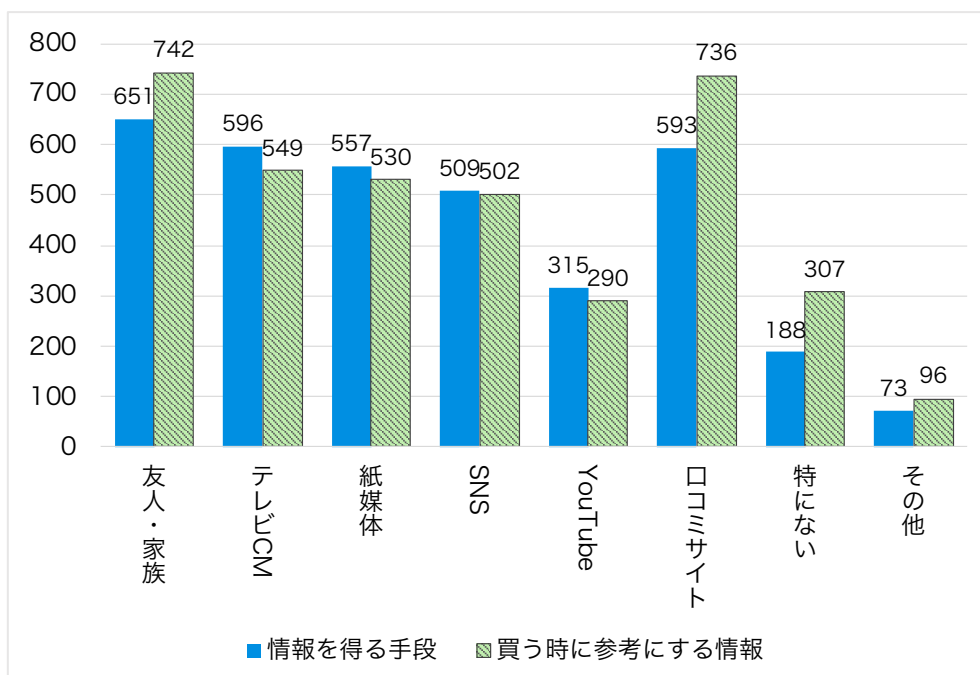


図4 化粧品の情報を得る手段 (Q7, Q9)

図1より、回答者の多くが30～50代であることがわかる。図2より、化粧の頻度については「毎日する(週7日程度)」、「ほぼ毎日する(週5日程度)」と回答した人が60.6%となった。また、図3より、化粧に対する好き嫌いについての質問では、「とても好き」、「好き」と回答した人が45.9%、「どちらでもない」と回答した人が40.5%、「とても嫌い」、「嫌い」と回答した人は13.5%であった。

図4より、化粧品を購入する際の情報収集では、化粧品の情報を得る時よりもクチコミサイトを閲覧して情報収集を行う傾向があることがわかる。つまり、テレビCMや雑誌などで化粧品の存在を知ることが多いが、化粧品購入時には、友人・家族からのアドバイスを聞く、クチコミサイトにアクセスするなどして、購入するか判断するための情報を新たに得ていると考えられる。

ここで、Q4とQ9の回答について、クロス集計を行った結果を表2に示す。表中1列目の括弧内の数字は、Q4の各項目を回答した人数を示す。また、背景色が灰色のセルは、化

表2 化粧の好き嫌いと購入時参考にする情報 (Q4, Q9)

	友人・家族	テレビCM	紙媒体	SNS	YouTube	口コミサイト	特にない	その他
とても好き(126)	70	43	45	54	32	62	6	2
好き(584)	266	181	207	217	121	291	32	32
どちらでもない(627)	240	187	180	144	72	222	105	32
嫌い(194)	62	40	48	33	17	88	45	11
とても嫌い(16)	3	2	1	4	3	5	8	0

粧に対する好き嫌い度合い別（Q4）に、化粧品を買うときに参考にする情報（Q9）として回答者が選択した上位2つを示している。表2より、Q4で「とても嫌い」と回答した人を除くと、化粧の好き嫌いに関わらず、クチコミサイトは化粧品購入時に多くの人が参考にしていることがわかる。

3.4. KH Coder による分析

Q11, Q12 の、化粧品についてのインターネット上のレビューで信頼できる情報および信頼できない情報の特徴に関する回答を可視化するために、共起ネットワーク図の描画を行

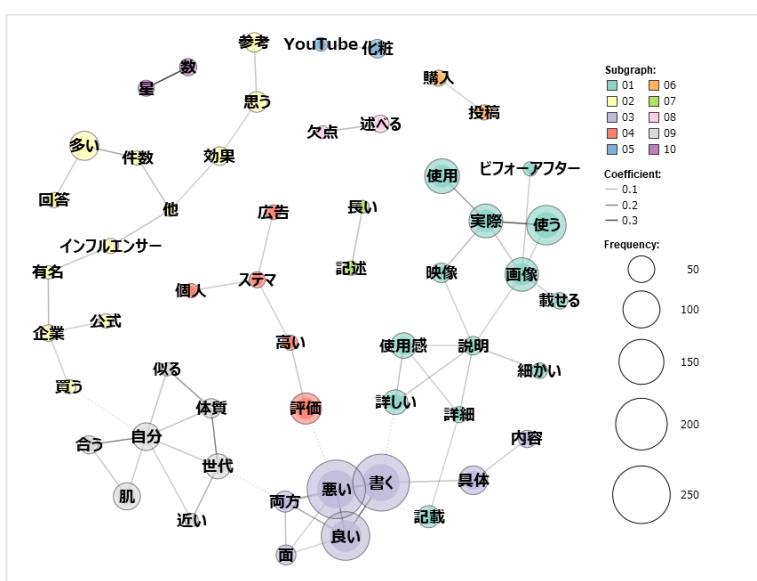


図5 信頼できるクチコミについての共起ネットワーク

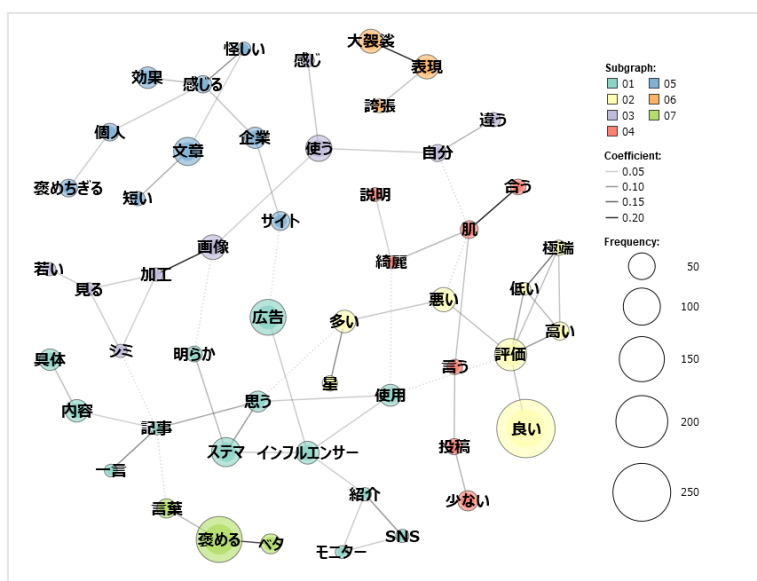


図6 信頼できないクチコミについての共起ネットワーク

った。その結果を図5, 6に示す。なお、描画はKH Coder[50]を用いて行っている。描画した共起ネットワーク図において、円の大きさは単語の出現回数の多さ、線の濃さは単語同士の繋がり強さを表している。今回は、共起係数の算出にJaccard係数を使用し、抽出語の最小出現回数を10、描画する共起関係の絞り込みを描画数60に設定している。図5, 6より、抽出語がいくつかのグループに分かれていることがわかる。

そこで、どのようなグループになるのかを詳しく見ていくために、KH Coderを用いて文書ごとのクラスター分析を行った。クラスター分析の方法は分類感度が良いとされるWard法を採用し、距離の算出にはJaccard係数を使用している。クラスター数は併合水準を考慮し、それぞれQ11では13、Q12では10に設定した。クラスター分析から得られた各クラスターの特徴と特徴語を表3, 4に示す。

図5, 表3から、ユーザが化粧品のカチコミを信用できると感じる要因として、マイナスな情報を含むこと、プラスな情報とマイナスな情報の両方を含むこと、内容の具体性などのカチコミ本文に関する特徴のほか、画像を含むこと、評価の高さ、カチコミ投稿者とユーザ自身との共通性などのカチコミ自体がもつ特徴が挙げられることが明らかになった。ここで、Q11の回答の中から評価の高さについての記述を見ると、評価が高いものや評価のばらつきが少ないもの、評価が高すぎないものなど、ユーザによって評価の高さについての認識が異なっていた。つまり、同じ評価のカチコミを見ても、信憑性についての判断は異なる可能性があることが示唆される。

また、図6と表4から、ユーザが化粧品のカチコミを信用できないと感じる要因として、プラスな情報、もしくはマイナスな情報のみで構成されていること、文章の長さ、大袈裟な表現を含むこと、砕けた文章や文法上不適切な文章を含むこと、広告であると明記していること、具体性がないことなど、カチコミ本文に関する特徴が多く挙げられた。さらに、発信者がインフルエンサーであること、評価が極端なこと、画像の有無や加工に関することなども挙げられていた。

Q11, 12の分析結果から、信憑性を判断する要因は文章、画像、カチコミ投稿者の3つに分類できると考えられる。また、ユーザはこの3つの要因がもつ様々な特徴から化粧品のカチコミの信憑性を判断しているが、その判断基準は人によって異なることが示唆された。そこで、この3つの要因について細かく調査する必要があると考えた。特に、信頼できないカチコミの特徴に関する回答の中で文章に触れているものに注目すると、文章が短いことを挙げる人もいれば、逆に文章が長いことを挙げる人も存在した。また、悪い点の記述があれば信頼できると考える人もいたが、悪い点のみの記述は信頼できないと考える人もいた。このように、信憑性が高いと感じるカチコミの文章の特徴は人によって真逆のものである可能性があるため、文章についてより詳細に調査、分析を行うことが重要であると考えられる。

表3 信頼できるレビューについてのクラスター分析結果 (Q11)

クラスター	特徴	特徴語
1	マイナスな意見が書いてあるもの	悪い, 書く
2	プラス・マイナス両方の意見が書いてあるもの	良い, 悪い, 書く, 両方, 記載, 面
3	実際に使用した人の意見	使う, 実際, インフルエンサー, 書く
4	クチコミサイトの意見	クチコミサイト
5	クチコミ件数が多い商品	多い
6	画像付きのもの, 画像が魅力的なもの	画像
7	具体的なもの	書く, 具体, 細かい
8	実際に使用した人の意見	使用, 実際
9	肌についての記述	肌
10	クチコミ件数が多いユーザ, 商品	多数
11	評価の高さ	評価, 高い, 悪い
12	自分の体質, 年齢に近いユーザの投稿	自分, 世代, 体質, 近い, 肌
13	詳しい使用感, 成分の記載, 公式のHP など	詳しい, 具体, 使用感, 情報, 企業など

表4 信頼できないレビューについてのクラスター分析結果 (Q12)

クラスター	特徴	特徴語
1	プラスな意見ばかり書いてあるもの	良い, 言う
2	ベタ褒めしているもの	褒める, ベタ
3	広告	広告
4	ステルスマーケティング	ステマ
5	評価が極端なもの	評価, 低い, 良い
6	大袈裟なもの	表現, 大袈裟
7	効果を強調しているもの	効果
8	マイナスな意見のみのもの, 文章の長さなど	悪い, 使う, 文章, 若い, 企業など
9	日本語が文法上不適切なもの	文章
10	画像がないもの, 加工された画像	画像, 加工

第4章 文章に着目したデータセット構築

4.1. 化粧品の選定とクチコミの収集

3章の化粧をする人を対象とした基礎調査より、化粧品購入時にクチコミを参考にする人が多いことや、人によって信用できると感じる基準が異なることがわかった。そこで、基礎調査で得られた知見をもとに、化粧品のクチコミの信憑性に関するデータセットを構築する。

データセット構築にあたり、クチコミを収集する化粧品としてファンデーション、化粧水、リップを選定し、この3カテゴリについて、各5商品、計15商品のクチコミを収集した。ファンデーションを選定した理由は、色や質感が写真のみではわかりにくく、時間経過や汗による崩れなどの体験に基づくクチコミが豊富であることから、クチコミの文章が重要視されると期待したためである。化粧水を選定した理由は、素肌に直接つけるため、自身の肌との相性や成分についての記述が重視されることや、継続利用することで効果を実感できることから使用期間が重要であると考えたためである。また、リップを選定した理由は、ファンデーションや化粧水と比較すると、百貨店で販売されている高級ブランドの商品でも安価である場合が多いが、インターネット上の画像が実際の発色と異なる場合もあり、発色についての記述が重視されると考えたためである。各カテゴリ内の商品については、価格帯が偏らないよう考慮し、化粧品専門のクチコミサイトである@cosme[6]内でランキングが上位でありクチコミ件数が多いものを選定した。以下にクチコミの収集で使用した商品とクチコミ収集時点での金額を示す。

ファンデーション

- オンリーミネラル「ファンデーション」(¥2,500)
- マキアージュ「ドラマティックパウダリー UV」(¥3,300)
- ランコム「タンアイドル ウルトラ ウェア リキッド」(¥6,600)
- シャネル「ルブラン コンパクト ラディアンス」(¥8,250)
- SUQQU「エクストラ リッチ グロウ クリーム ファンデーション」(¥11,000)

化粧水

- ナチュリエ「ハトムギ化粧水」(¥715)
- 肌ラボ「極潤プレミアム ヒアルロン液」(オープン価格, ¥970 前後)
- イプサ「ザ・タイム R アクア」(¥4,400)
- SK-II「フェイシャル トリートメント エッセンス」(¥9,350)
- ランコム「クラリフィック デュアル エッセンス ローション」(¥12,100)

リップ

- セザンヌ「ラスティング リップカラーN」(¥528)
- rom&nd「ジューシーラスティングティント」(¥1,320)
- B IDOL「つやぶるリップ」(¥1,540)
- アディクション「ザ マット リップ リキッド」(¥3,520)
- ランコム「ラブソリュ ルージュ」(¥4,400)

各商品に対するクチコミを20件ずつ、計300件を収集した。収集元の内訳は@cosme から236件、LIPS[7]から8件、Twitter[8]から56件である。@cosmeやLIPSは日本国内において代表的な化粧品専門のクチコミサイトであり、Twitterは化粧品のクチコミを閲覧できるSNSでありユーザー数も多いため、データセット構築ではこれらのサイトのクチコミを使用することにした。ここで、化粧品のクチコミ投稿が盛んなInstagram[9]などのSNSを用いなかったのは文章ではなく画像や動画がメインのクチコミが多いためである。収集元の内訳の割合に偏りがあるのは、@cosmeは1つの化粧品に対するクチコミを投稿する形式であるのに対して、LIPSやTwitterなどは複数の化粧品に関する投稿になっていることや、化粧品の正式名称を書いていないことが多く、クチコミを探すのが困難であったためである。また、商品に対する先入観を除くため、クチコミ内の商品名、ブランド名、プロデューサーなどの個人名は「製品A」というように伏せた状態に変更した。最後に、各クチコミについて、クチコミ本文に加え、その化粧品の良い点、悪い点、使用感、投稿者の体質、化粧品に含まれる成分、他の化粧品との比較、リピート購入の有無、企業からのプレゼントやPRであるかどうかの8項目の特徴について、記述の有無を筆者自身が付与した。これらの8項目は、クチコミの内容に含まれ、信憑性評価に影響を与える可能性があるとして期待し選定した。なお、クチコミを収集する際は、クチコミ内での化粧品に対する評価や、前述の特徴がなるべく偏らないように留意した。

4.2. ユーザによる信憑性判断

本データセット構築を行うにあたり、Vue.jsを用いたWebアプリケーションをアノテーション付与システムとして実装した。システムの画面を図7に示す。システムでは、クチコミを1件ずつ表示し毎回回答を記録することで、実験協力者が好きなタイミングで回答を中断できるよう設計した。本調査における各クチコミへの調査項目を以下に示す。

- Q1. このクチコミはどの程度信頼できますか？
- Q2. この商品を買いたいと思いますか？
- Q3. この商品を使用したいと思いますか？
- Q4. 評価の基準（なぜそのように評価したか）を教えてください

クチコミに関する調査

●これはファンデーションについてのクチコミです

ブラシで塗ってつけています。
お肌なじみがよく、色味も良いと思います。
全く磨れないわけではないけど、磨くだけでもぐっすり知らずで、お肌が磨れにくいと思います。
ケースも定期的にかかっているのが出てくるのも魅力だと思います。

Q1. このクチコミほどの程度信頼できますか？

信頼できない	-2	-1	0	1	2	信頼できる
	○	○	○	○	○	

Q2. この商品を買いたいと思いますか？

買いたいと思わない	-2	-1	0	1	2	買いたいと思う
	○	○	○	○	○	

Q3. この商品を使用したいと思いますか？

使用したいと思わない	-2	-1	0	1	2	使用したいと思う
	○	○	○	○	○	

Q4. 評価の基準（なぜそのように評価したか）を教えてください。

次へ
回答を終了する

図 7 本調査システム画面

ここで、Q1～3については5段階のリッカート尺度（-2～+2）で評価してもらい、Q4については自由記述で回答してもらった。

また、本調査は女子大学生13名に依頼した。事前に注意事項としてすべての質問が回答必須であること、好きなタイミングで回答を中断・再開できること、回答期限などを伝えた。最初に練習用のクチコミ3件について回答を送信してもらった後、300件のクチコミについて回答してもらった。

4.3. 結果と考察

4.3.1. データセット構築者ごとの結果・考察

実験協力者13名のうち、9名（A～I）が回答期限内に300件のクチコミすべてに回答した。ここでは300件すべてに回答したこの9名のデータを用いて分析を行う。

まず、9名について一元配置の分散分析を行ったところ、有意差が認められた（ $p < 0.01$ ）。したがって、実験協力者9名の間でも化粧品のクチコミに対する信憑性の評価基準が異なると考えられる。これは3章の調査と同様の結果である。また、実験協力者A～Iの9名の信憑性評価（Q1）についてcos類似度を総当たりで求めた結果を表5に示す。cos類似度の平均値は0.60であった。特にcos類似度が低かった組み合わせとして、0.50未満の数値を太字で示している。表5より、実験協力者A、G、Iはcos類似度が低いことがわかる。このことから、この3名は評価基準が他の実験協力者とは異なっている可能性がある。また、9名の信憑性評価の平均は0.79であり、各実験協力者の分散の平均は0.86であった。これに対し、cos類似度が低かった実験協力者A、G、Iの信憑性評価の平均はそれぞれ0.61、0.47、

表 5 実験協力者 9 名の cos 類似度

	A	B	C	D	E	F	G	H	I
A		0.61	0.61	0.43	0.48	0.56	0.49	0.55	0.48
B	0.61		0.78	0.65	0.73	0.70	0.52	0.68	0.54
C	0.61	0.78		0.64	0.74	0.70	0.60	0.75	0.58
D	0.43	0.65	0.64		0.77	0.77	0.40	0.72	0.51
E	0.48	0.73	0.74	0.77		0.80	0.50	0.72	0.54
F	0.56	0.70	0.70	0.77	0.80		0.46	0.69	0.60
G	0.49	0.52	0.60	0.40	0.50	0.46		0.52	0.46
H	0.55	0.68	0.75	0.72	0.72	0.69	0.52		0.49
I	0.48	0.54	0.58	0.51	0.54	0.60	0.46	0.49	

0.54 であり、分散は 1.53, 1.19, 1.24 で、信憑性評価の平均が低く、分散が大きいことがわかる。このことから、この 3 名はクチコミの信憑性を低く見積もるが、信憑性が高いと感じたものには非常に高く評価する傾向があると考えられる。一方で、実験協力者 D は信憑性評価の平均が 1.11 と高く、分散が 0.56 と小さいことから、多くのクチコミについて信憑性が高いと評価する傾向があると考えられる。ここで、実験協力者 A, G, I が高く評価したクチコミや実験協力者 D が低く評価したクチコミについてさらに分析を行う。

実験協力者 A, G, I の 3 名のうち、2 名以上が信憑性の評価を 2 としたクチコミは 31 件あった。この 31 件についての評価理由を詳しく見ると、「購入した経緯があり、また淡々と評価をしていて信頼できました」、「併用した商品まで事細かに書いてあるので、信用できると思った」、「この人なりの良い使い方が書かれていて参考になる」など購入時や使用時の具体的なエピソードが含まれるクチコミについて信憑性を高く評価していた。また、「自分の肌が何が合わないタイプかについて記述されていて、合わなかった原因も考察されていたので、信用できると思った」、「使用した期間があったので信頼できました」といった投稿者についての情報、「良い点と悪い点両方わかりやすくまとめてあって信頼できました」といった両面性をもったクチコミ、「別日に追加で記述されていたから」といった追記を含むクチコミなども信憑性を高く評価していた。

実験協力者 A, G, I と他の実験協力者 6 名の評価理由を比較すると、A, G, I の中では着眼点が一致しているが、他の実験協力者は同じクチコミに対して異なる評価理由を述べていることもあった。例えば、実験協力者 A, G が購入経緯についての記述を根拠に信憑性を評価していたクチコミに対して、他の実験協力者はその化粧品の効果についての記述や、リピート購入について記述に着目していた。したがって、実験協力者内で A, G, I の 3 名は信憑性が高いと判断する基準に共通性があると考えられる。しかし、表 5 で示した信憑性評価の cos 類似度は低いため、信憑性を低いと評価する基準は異なると考えられる。

また、実験協力者 D が信憑性の評価を-2としたクチコミは2件あった。2件についての評価理由は、「書き方、どの商品に対しても批判が非常に目立つ点で、ただのクレームのように思え、信頼はできない」、「客観性が全くなく、商品に関する情報も限られている。商品を薦めているのはわかるが、書き手の立場も分からず信頼度は低い」というもので、商品に対する批判が多いものや、商品や投稿者に関する情報が少ないものを低く評価していた。実験協力者 D が信憑性を低く評価したクチコミの件数が少ないため、他の実験協力者と比較することは困難であるが、2件に関しては他の協力者と評価理由の一致が見受けられたため、信憑性が低いと評価する観点に独自性は認められなかった。

4.3.2. クチコミごとの結果・考察

次に、クチコミごとに信憑性評価 (Q1) の回答の分散を算出した。ここで、分散が大きいクチコミは人によって信憑性の評価が異なるということである。この調査の目的であるシステムが支援すべき箇所の検討を行うためには、人による信憑性の評価にブレが生じるクチコミに着目することが重要であると考えた。そこで、分散の大小に着目し、さらに分析を行った。ここでは、分散が平均値の 0.56 以上であった 122 件と、0.56 未満であった 178 件のクチコミについて、クチコミを収集する際に付与した特徴 8 項目を用いてクチコミの特徴を分析した。8 項目の特徴について、その特徴を含むクチコミ件数の割合を表 6 に示す。表 6 では、「投稿者の体質」を体質、「化粧品に含まれる成分」を成分、「他の化粧品との比較」を比較、「リピート購入の有無」をリピート、「企業からのプレゼントや PR であるかどうか」を PR とそれぞれ表記している。表 6 より、変化量が特に大きいのは、悪い点についての記述と企業からのプレゼントや PR の記述である。分散が大きいものに絞るほど、悪い点についての記述が含まれる割合は小さく、プレゼントや PR の記述が含まれる割合は大きくなった。ここで、ある特徴をもつクチコミについて、すべてのクチコミに対する割合よりも分散が大きいものに絞ったときの割合の方が大きいということは、その特徴をもっていると分散も大きくなりやすいと言える。つまり、実験協力者間で評価のブレが生じやすいということである。これらのことから、商品の悪い点が含まれているクチコミはユーザによる信憑性評価にブレが生じにくく、PR の目的で投稿されたクチコミは評価にブレが生じやすいと解釈できる。評価にブレが生じにくく、その評価が正しいものであれば、システムで支援する際の優先度は低いと考えられる。一方、評価にブレが生じやすいクチコミは、正しく信憑性を評価できないユーザもいると考えられるため、システムによる支援が必要である。なお、良い点、体質、成分、比較を含むクチコミの割合は分散が大きいクチコミに絞っても大きな変化はなかった。これらの特徴は、評価の分散に影響を与えてないと考えられる。

表 6 各特徴を含む割合

	良い点	悪い点	使用感	体質	成分	比較	レポート	PR
全体(300件)	0.94	0.36	0.59	0.31	0.20	0.13	0.19	0.12
分散小(178件)	0.93	0.45	0.64	0.34	0.20	0.15	0.24	0.04
分散大(122件)	0.95	0.23	0.52	0.26	0.19	0.10	0.13	0.22
変化量 (分散小-分散大)	-0.02	0.22	0.12	0.08	0.01	0.05	0.11	-0.18

4.3.3. 信憑性評価についての結果・考察

また、信憑性の評価 (Q1) について正の値・負の値にそれぞれ分類し、評価理由 (Q4) の結果について分析を行った。まず、信憑性の評価が正の値であったもの、負の値であったものの評価理由について、それぞれ KH Coder[50]を用いて共起ネットワーク図を描画した。描画結果を図 8, 9 に示す。今回は、最小出現数の設定を信憑性の評価が正の値では 25, 負の値では 5 とした。また、描画する共起関係の絞り込みを描画数 60 に設定している。さらに、文書ごとのクラスター分析も行った。分析結果を表 7, 8 に示す。共起ネットワーク図内で「信頼」と繋がりがある単語が、信憑性の評価において特に重要なキーワードであると考えたため、図 8 で「信頼」とつながりがある単語を表 7 の特徴語の欄に太字で示し、図 9 も同様に表 8 に示した。これらの単語に関連する内容は、クチコミ閲覧者が信憑性を評価する際に重視する可能性が高い。

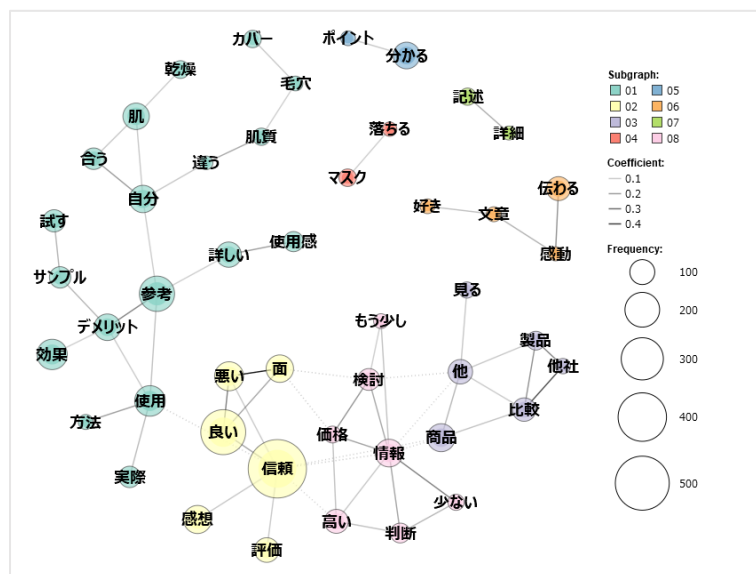


図 8 信頼できると評価した理由の共起ネットワーク

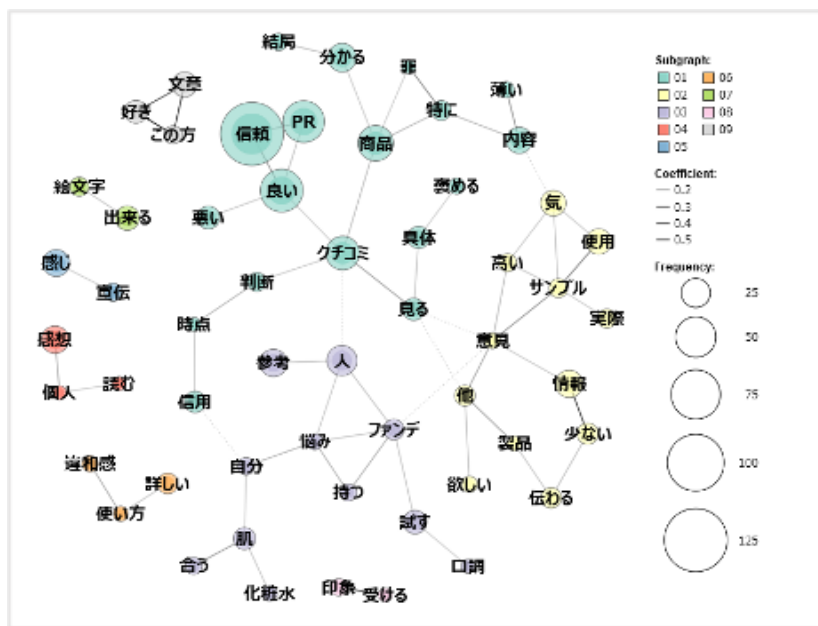


図 9 信頼できないと評価した理由の共起ネットワーク

表 7 信頼性の評価が正の評価理由クラスター分析結果

クラスター	特徴	特徴語
1	商品の良さが伝わるもの	良い, 商品, 伝わる, 使用
2	プラス・マイナス両方の意見が書いてあるもの	悪い, 良い, 面, 判断
3	情報が少ないもの	情報, 価格, 少ない, 検討, 判断
4	他の商品との比較があるもの	比較, 商品, 他社, 製品, 他, 伝わる
5	色についての記述があるもの	色, 自分
6	成分についての記述があるもの	成分
7	投稿者の肌質について記述されているもの	肌, 自分, 合う, 肌質, 違う, 乾燥
8	個人の感想が記述されているもの	感想
9	効果についての記述があるもの	効果, リピート
10	商品の情報がわかりやすいもの, 値段が高いものなど	分かる, 伝わる, 評価, 人, 高い
11	使用方法についての記述があるもの	使用, 方法, 実際, サンプル, 参考
12	マイナスな情報が書かれているもの	デメリット, 参考, サンプル, 使用
13	記述が詳しいもの	参考, 詳しい, 使い方, PR, 使用感

表8 信頼性の評価が負の評価理由クラスター分析結果

クラスター	特徴	特徴語
1	良さがわからないもの、PRと明記されているもの	良い, 低い, PR
2	PRと明記されているもの	PR
3	文章が好みでないもの	文章, 好き
4	情報量が少ないもの	少ない, 情報
5	絵文字を使用しているもの	絵文字, 出来る
6	投稿者と自身の体質が合わないもの	人, 参考, 使用, 他, 肌
7	広告のような文章, 文章が薄いものなど	良い, 感じ, 感想, 気, 試す, 悪い
8	わかりにくいもの	分かる
9	内容が薄いもの	内容
10	商品の魅力を感じられないもの	商品, 魅力, 分かる

表8より、唯一特徴語に「文章」が含まれていたクラスター3に着目し、言葉遣いについて分析を行った。クラスター3に分類された評価理由は主に、このクチコミの文章が好きではないというもので、この評価理由がついているクチコミの信頼性評価の平均は-0.07であった。すべてのクチコミに対する信頼性評価の平均は0.79であったことから、言葉遣いが好みでないクチコミは、信頼性も低く評価されやすいと考えられる。評価理由で文章が好みでないと言われていたクチコミを見ると、「少量でオッケ👍ってぐらいのびが最高😍😍マスクしててもそんなによれないし！個人的には、今まで使ってきたファンデの中で1番👍👍👍👍👍👍」、 「まず、パッケージがかわいい!!!持ちたくなる可愛さです!!!お肌に優しいだけでなく、薄化粧 ステキ(^)/ 使用道具のお勧めはパフではなく、筆です☆いつかは、ノーファンデになるくらいにしたいからこそ、なるべく肌へのストレスは普通の化粧品から減らしたいですよねーっ たっぷり基礎化粧品した後に、こちらのパウダーファンデーション、目元など気になるところにリキッドやコンシーラーを重ねるだけで、肌が仕上がります(^♪)のように絵文字や顔文字、記号を多用しているものや、「化粧品販売を覗いた際にオススメされて買ってしまった商品ですわ。ファンデーションにほんのりA社っぽい匂いがついていますわ。色のバリエーションもいくつかありますわ。そして日焼け止めの成分もついていますがあまり当てにはなりませんわ。かなり薄づきで人によってはフェイスパウダー?って思うかたもいるのではないのでしょうか。私は厚化粧なので少し不向きかもしれませぬわ。見た目はとてもお洒落で可愛いですわ。」といった特徴的な語尾の文章であった。こういったクチコミは、文章が実験協力者の好みに合わないという理由で正当に信頼性を評価されなかった可能性がある。しかし、このクチコミが実際には有益なものであった場合、その商品についての情報を得る機会を損なってしまうと考えられる。

次に、信憑性評価において実験協力者9名全員が正の値で評価したクチコミに着目する。全員が共通して信憑性を高く評価したクチコミの特徴は、客観的な信憑性の評価軸として利用可能であると考えられる。そこで、実験協力者全員が信憑性を正の値で評価したクチコミ90件について、8項目の特徴それぞれを含む割合を集計した結果を表9に示す。表9より、PRを含むクチコミが極端に少ないことがわかる。PRを含むクチコミで全員が信憑性を高く評価したのは2件であり、その2件についての評価理由は、「企業からもらったと書いてあったけど、他社との比較や、率直な感想が書いてあったので信頼出来ると思った」「合わなかったが、その会社に対して感謝している点などから中立的な立場から書かれていることが分かり、信頼できると思った」などであった。つまり、企業がPRを目的としてユーザーにプレゼントした商品であっても、他の製品との比較や、商品の悪い点について率直な意見を述べていれば、信憑性を高く評価されやすいと考えられる。これらの着眼点は9名にほぼ共通しており、客観的な信憑性評価軸として利用可能である。なお、今回のデータセット構築では全員が信憑性について負の値で評価したクチコミは存在しなかった。

表9 各特徴を含む割合

	良い点	悪い点	使用感	体質	成分	比較	リピート	PR
全体(300件)	0.94	0.36	0.59	0.31	0.20	0.13	0.19	0.12
正の評価(90件)	0.94	0.56	0.76	0.44	0.26	0.21	0.32	0.02

4.3.4. 商品ごとの結果・考察

今回の調査で用いた化粧品カテゴリについて、それぞれのカテゴリについてクチコミ内での記述内容が異なると期待し、ファンデーション、化粧水、リップを選定した。そこで、記述内容に差があるのかを検証するために商品ごとにKH Coder[50]を用いた分析を行った。

まず、各カテゴリのクチコミについて、最小出現数を10、描画する共起関係の絞り込みを描画数60とし、共起ネットワーク図を描画した。描画結果を図10~12に示す。図10~12から、各カテゴリでそれぞれクチコミ内容に特徴があることがわかる。図10より、「カバー」、「ツヤ」、「毛穴」、「シミ」といった効果に関する単語や、「スポンジ」、「下地」、「コンシーラー」といった使用方法に関する単語、「崩れる」、「直し」のような経験に関する単語を含むクチコミがファンデーション特有であると考えられる。図11より、「透明」、「綺麗」、「肌」など効果に関する単語や、「重ねる」、「少量」、「手」など、使用方法についての単語、「オイル」、「酸」、「配合」など成分についての単語を含むクチコミは化粧水特有であると考えられる。また、図12より、「色」、「ピンク」、「発色」などの色についての単語、「指」、「ティッシュ」、「オフ」などの使用方法についての単語、「マスク」、「付く」、「落ちる」、「荒れる」といった経験に関する単語を含むクチコミがリップ特有であると考えられる。

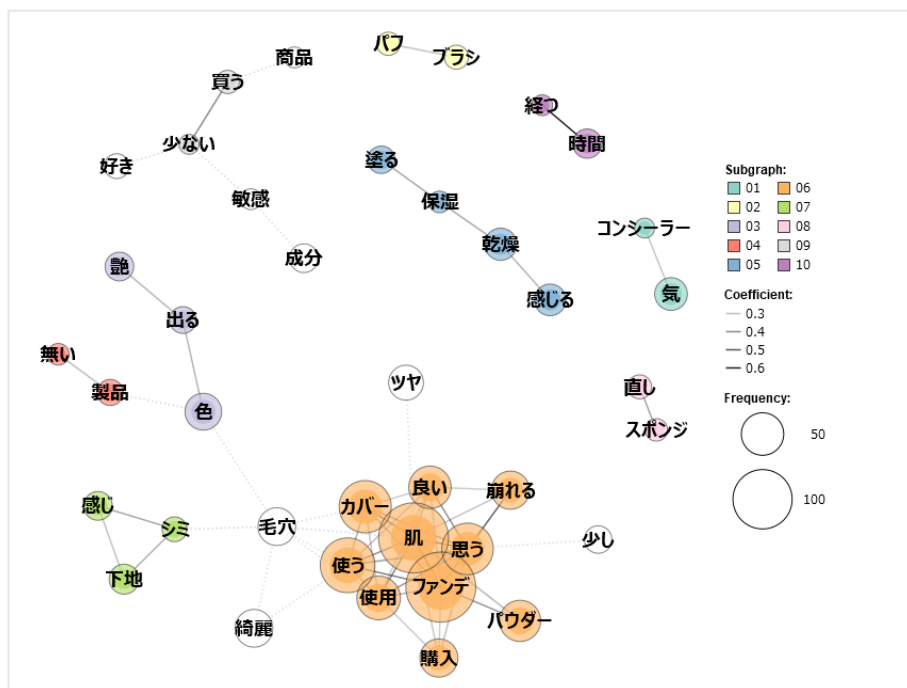


図 10 ファンデーションのクチコミの共起ネットワーク

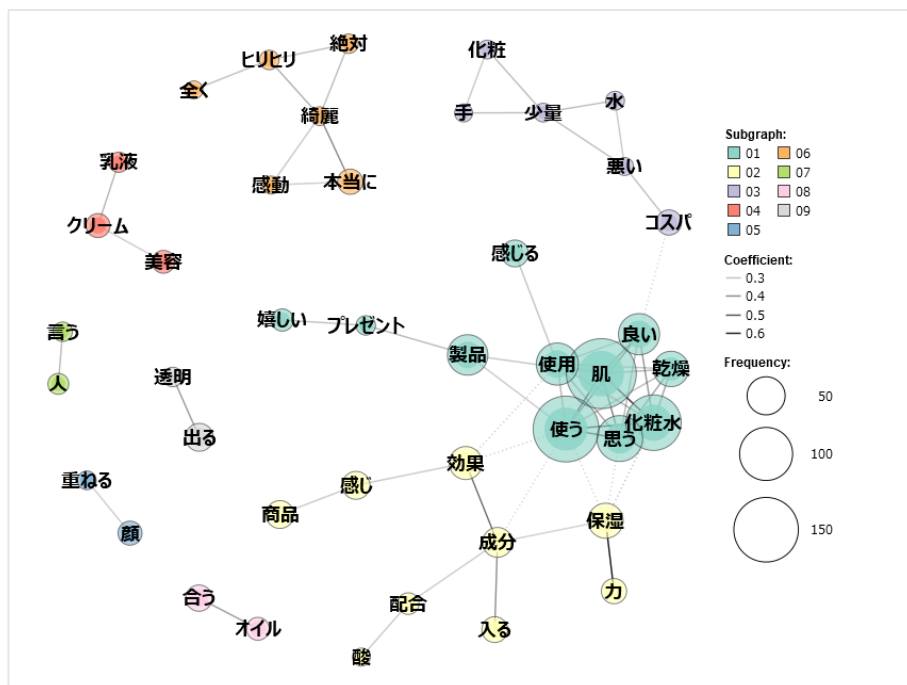


図 11 化粧水のクチコミの共起ネットワーク

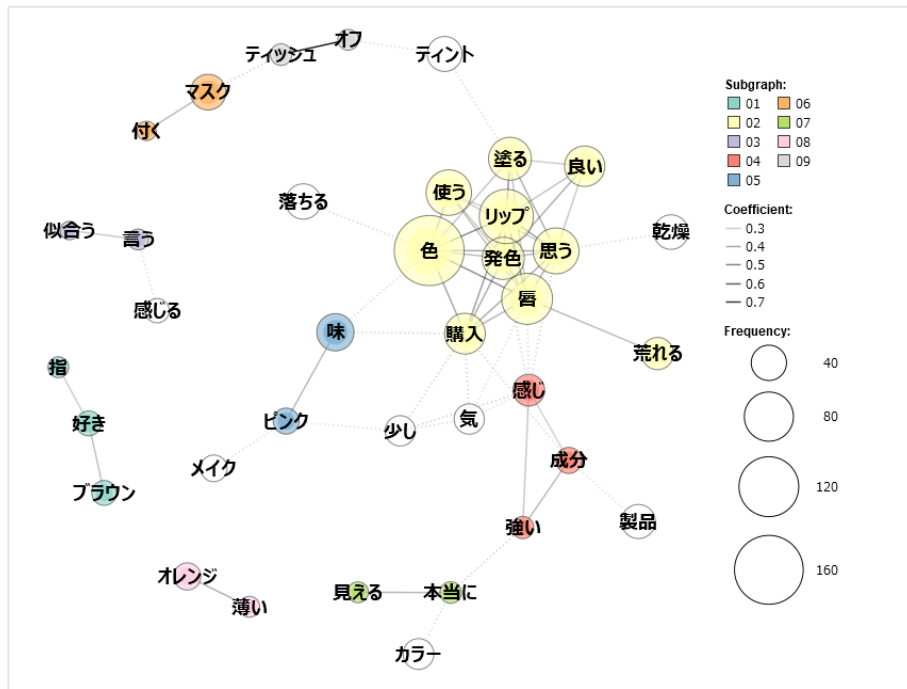


図 12 リップのクチコミの共起ネットワーク

次に、各カテゴリのクチコミについての評価理由 (Q4) について、最小出現数を 15、描画する共起関係の絞り込みを描画数 60 とし、共起ネットワーク図を描画した。描画結果を図 13~15 に示す。図 10, 13 の単語を比較するとファンデーション特有の単語では「カバー」、「毛穴」が共通しており、ファンデーションのクチコミをユーザが閲覧する際は、効果についての記述を重視している可能性が示唆された。また、「使用」といった単語も共通することから、使用方法についての記述も重視されると考えられる。図 11, 14 を比較すると、化粧水特有の単語では「肌」が共通しており、図 14 には「肌質」も含まれている。したがって、化粧水のクチコミをユーザが閲覧する際は肌についての記述が重視され、特にユーザ自身と投稿者の肌質を比較することが多いと考えられる。また、図 12, 15 を比較すると、リップ特有の単語では、「色」、「発色」、「落ちる」、「荒れる」などが共通しており、リップについてのクチコミをユーザが閲覧する際は、色についての記述や経験に基づく記述が重視されると考えられる。また、図 15 では「使い方」、「使用」が含まれていることから、リップについてのクチコミでも使用方法についての記述は重要であることが示唆された。

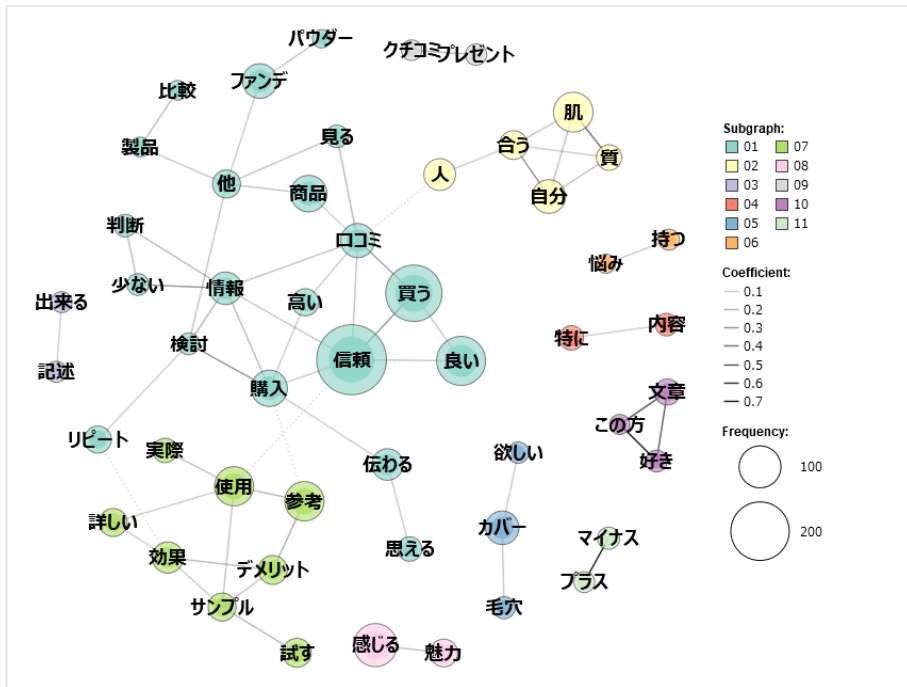


図 13 ファンデーションの評価理由の共起ネットワーク

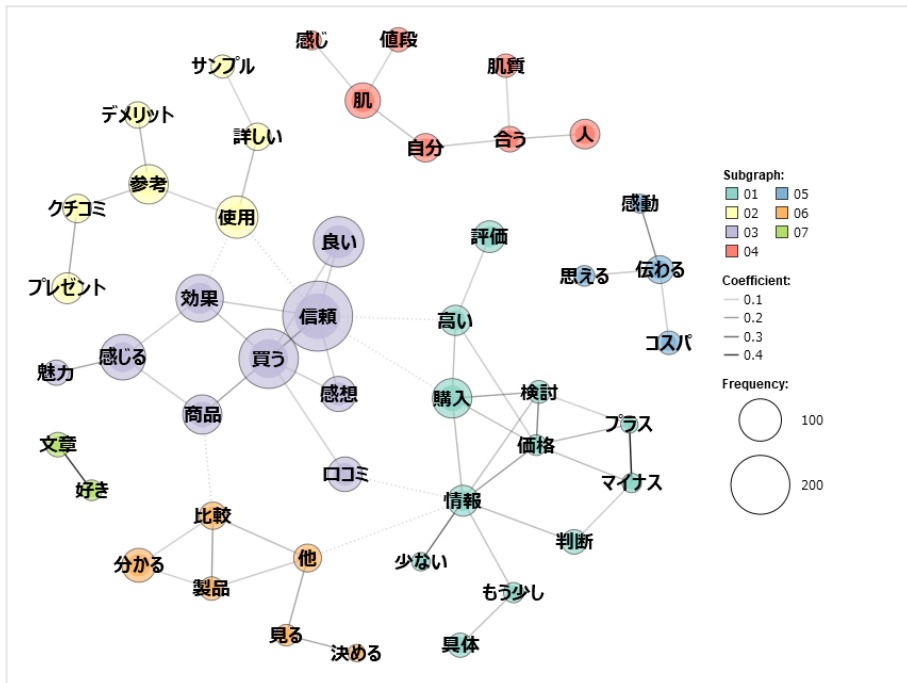


図 14 化粧水の評価理由の共起ネットワーク

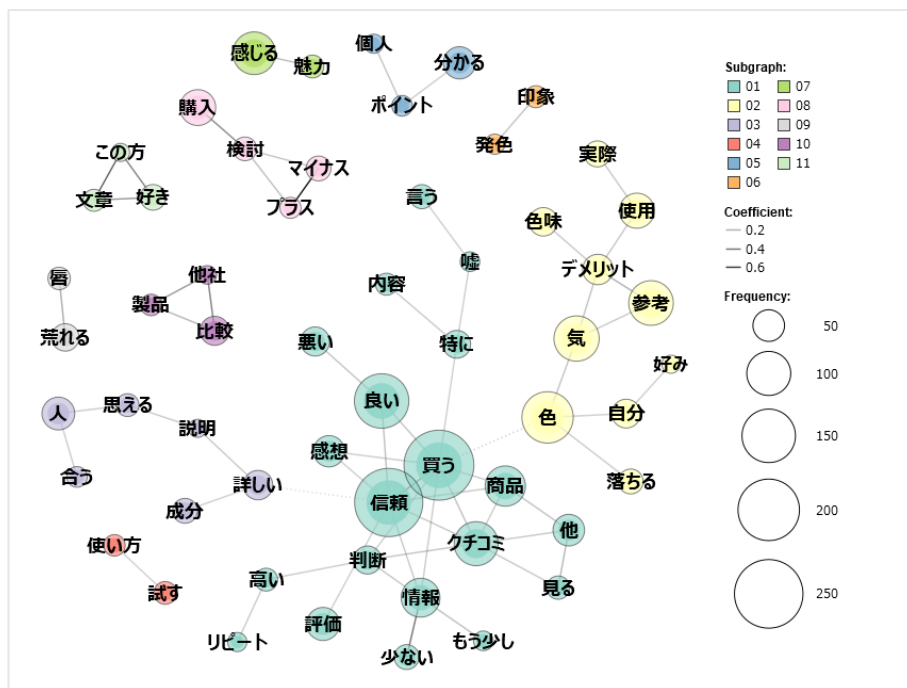


図 15 リップの評価理由の共起ネットワーク

第5章 信憑性評価軸の提案と有用性検証

5.1. 信憑性評価軸に関する検討

4章の調査より、信憑性の評価軸としていくつかの候補が考えられる。本章では、この候補について述べる。

まず、4.3.2項で述べた、商品の悪い点が含まれるクチコミやPRと明記されているクチコミが評価軸の候補として挙げられる。特にPRを含むクチコミは、実験協力者間で信憑性評価にブレが生じており、正しい信憑性評価が困難であると考えられる。そのため、システムを介して評価を支援する必要がある。さらに、4.3.3項で述べたように、PRであっても率直な意見や他の製品と比較が含まれるなど複合的な評価軸にすることで、より高精度な信憑性判定を可能にできる。

次に、表7、8内に太字で示した単語や、4.3.4項で示したそれぞれの化粧品カテゴリにおいて重要であると考えられる単語を用いた評価軸の設計も候補として挙げられる。これらの単語は、クチコミ閲覧者が信憑性を評価する際に重視する可能性が高く、システムを介して客観的に単語の有無を示すことは有用であると考えられる。また、文章が好みでないという理由で正当に信憑性を評価されなかったクチコミにおいても、単語の有無を提示することで有益な情報を得る機会を損なわずに済むと期待される。加えて、化粧品の種類によって評価軸に用いる単語を変更することで、より高い精度で信憑性を評価できると考えられる。

また、4.3.1項で述べた、実験協力者A、G、Iのみがもっていた評価基準を評価軸として採用することで、より多面的な信憑性の評価が可能となり、高精度で評価が行えると考えられる。

しかし、これらの評価軸候補は「人がどのようなクチコミを信用しやすいのか」がもとになっているため、この軸で信憑性の評価が可能かを検証し、さらなる評価軸の検討を行う必要がある。

5.2. 使用状況別のクチコミデータセット構築

検討した信憑性評価軸について、有用性の検証を行う。具体的には、使用状況の異なるクチコミを用意し、評価軸を用いてその使用状況を判別できるかを検証する。これを行うために、化粧品のクチコミを使用状況別（「実際に長期的に使用して投稿されるクチコミ」「一度だけしか使用せずに投稿されるクチコミ」「全く使用していないのに投稿されるクチコミ（嘘のクチコミ）」）に分類したデータセットの構築が必要となる。これまで構築してきたデータセットでは、インターネット上のクチコミを使用していたが、これらは投稿者が本当にその化粧品を使用したのかを判断することが不可能である。そのため、使用状況別のクチコミを新たに作成した。

本章ではこれまでと同様、ファンデーション・化粧水・リップについてのクチコミを作成する。それぞれの使用状況別でのクチコミ作成は以下の通りである。

- 実際に長期的に使用して投稿されるクチコミ：20～25歳の女性8名に、自身が使用している化粧品について商品名とクチコミをできるだけ多く作成するよう依頼した。
- 一度しか使用せずに投稿されるクチコミ：Kajitaら[16]の12種類のファンデーション塗布実験（一度だけそのファンデーションを顔に塗ってもらうというもの）における実験協力者に依頼し作成してもらったクチコミ194件のうち、短すぎるものを除くために50文字以上のものだけを選定した。
- 全く使用していないのに投稿されるクチコミ（嘘のクチコミ）：20～25歳の7名（男性3名、女性4名）に、指定した化粧品60点のうち、使用したことがないものについて100文字以上のクチコミを書くよう指示した。ここで指定した化粧品は、@cosme[6]内の各化粧品カテゴリのランキングから、上位のものを選定した。なお、クチコミ作成者には、「あなたはある企業から、『これから提示する商品の説明ページを見て、できるだけ使っていないことがばれないようにクチコミを書いてクチコミサイトに投稿してください。謝礼として、商品券を差し上げます』と依頼されています」という設定を伝えた。ここで、実際のサクラによるクチコミも全く化粧品を使用していない人が投稿している可能性があるため、作成依頼において化粧品の使用歴は不問とした。クチコミ作成の際には、化粧品メーカーの公式サイトから化粧品の情報を提示し、その情報を見て自由に書いてもらった。

このクチコミデータセット構築で、323件（実際に使用50件・一度だけ使用70件・未使用203件）のクチコミを得た。作成されたクチコミの例を表10に示す。

表 10 収集・作成したクチコミの例

使用状況	クチコミ
実際に使用	<p>乾燥肌なので保湿力がありながら、カバー力もあるファンデーションが欲しくて購入しました。面倒くさがりなのでクッションなのも決めてです。SPF50、PA++++で日焼け止め効果は申し分ありません。色は今回リネンを選びました。前回ジンジャーを間違えて購入してしまい黄味が強くて失敗してしまいましたが、リネンは私の肌色にちょうどあっている感じでした。日本のファンデーションの黄味が気になる人には良いと思います。ただ最近はマスクをつけての生活なので、少しよれやすさが目立つ気がします。ツヤ感が売りなのにパウダーのせまくってマットぼくしても鼻周りはやっぱりよれてしまいますね。お直し面倒なので、マスク向けの商品に乗り換えようかなと考えています。</p> <p>肌のトーンアップをしてくれて、厚塗り感が無くサラッとした使用感です。自分は肌が弱くて他社の安いファンデーションでは肌荒れしてしまいましたが、この商品では肌荒れしませんでした。</p>
一度だけ使用	<p>イプサで肌診断をしていただいた際にサンプルを頂きました。すごい評判がいいと知っていたので期待していたのですが、悪くはないけど、すごくいいという感じもわからなかったです。何か肌に良い変化も起こらなかったもので、これならいつも使っている薬局で買えるプチプラのものでいいかと思ってしまうような感じでした。</p> <p>手に出した感じは色みが暗くみえるのですが、塗ってみるとそこまで暗くなかったです。</p> <p>毛穴が消えてびっくりしました。</p>
未使用	<p>20代後半の女性です。普段から愛用させていただいています！このオルビスユーローションは使用するときもベタベタせず、肌にはとっても潤いを与えてくれるのでこれからも愛用していこうと思っています。価格的にも買いやすくて感謝です。</p> <p>少し前に友人から貰ってからずっと愛用しています！</p> <p>私が持っているリキッドファンデの中ではかなり肌に馴染みやすく、少量でも十分なのでコスパも良く重宝しています。ただ、その分薄く塗りがちで紫外線を防いでいるのかについては分からないのでスプレータイプの日焼け止めと併用しています</p>

なお、クチコミデータセット構築においては、クチコミ作成者に対し、クチコミ作成前に化粧品のクチコミを閲覧するよう指示した。これは、化粧品のクチコミをあまり閲覧しない作成者がいることを考慮し、化粧品のクチコミにどのようなことが書かれるかを知ってもらうためである。

また、化粧品の使用状況やクチコミの閲覧頻度によって作成したクチコミが変化する可能性もあると考え、実際に使用と未使用のクチコミ作成者には、化粧品の購入とクチコミの利用状況に関するアンケートに回答してもらった。このアンケートでは、普段クチコミサイト、EC サイトや SNS で化粧品のクチコミを閲覧・投稿するか、ファンデーション・化粧水・リップの購入前にどの程度クチコミを調べるか、どのような記述を重視するかといったクチコミの利用状況と、ファンデーション・化粧水・リップをそれぞれいくつ持っているか、衝動買いをするかといった購入についての項目を用意した。アンケートの結果、クチコミ作成者のうち、化粧品のクチコミを投稿する人は実際に使用・未使用のクチコミ両方を作成した1名であり、他の作成者は全く投稿しないと回答していた。閲覧については、全くしないと回答したのは男性1名のみであった。また、化粧品の購入前にどの程度クチコミを調べるかという項目については、数日調べる、100件以上閲覧するといった人もいたが、30分程度と回答した人が多かった。さらに、化粧品によっても差があり、ファンデーションの購入前は女性の作成者全員がクチコミを調べると回答したが、化粧水やリップの場合はあまり調べないという回答も少数ではあるが存在した。この理由として、化粧水の場合は肌が弱いから決まったものしか使用しないことが挙げられており、リップの場合は衝動買いが多いことが挙げられていた。また、男性は3名とも化粧水のみを使用しており、ファンデーションとリップは購入したことがないと回答していた。

5.3. クチコミに対する軸評価データセット構築

5.3.1. 概要

構築したクチコミデータセットについて、軸評価データセットの構築を行う。軸評価データセットの構築は大学生・大学院生17名（男性8名、女性9名）に依頼した。まず回答者には、性別・化粧の頻度・化粧品への興味・クチコミの閲覧頻度についてのアンケートに回答するよう指示した。その後、データセット構築システム（Vue.js を用いた Web アプリケーションとして動作）に移動し、軸の評価を始めてもらった。実際のシステムの画面を図16に示す。本システムでは、クチコミがランダムに1件ずつ表示され、毎回回答を記録し、回答者の好きなタイミングで回答を中断できるよう設計した。回答項目を以下に示す。

- 製品の悪い点が書かれている
- 他の製品との比較が書かれている
- 購入経緯が書かれている

- 併用した製品が書かれている
- クチコミ投稿者なりの使用方法が書かれている

これらの項目は5.1節で述べた評価軸をもとに作成したもので、クチコミの特徴としてはまるものすべてを選択してもらった。また、この5項目以外に、クチコミについて気づきがあれば自由記述欄に記入するよう指示した。

一方、5.1節で検討した評価軸に含まれていた、「それぞれの化粧品において重要な単語（カバー・肌質・発色など）の有無」については、単語が含まれているかを機械的に判別し、単語がいくつ含まれるかをカウントすることで軸評価とした。使用した単語を化粧品別に表11に示す。ここに示した単語は、4章のデータセットを分析し、共起ネットワークを描画した際に登場した単語のうち、その化粧品特有のものという観点で選定した。

クチコミに関する調査

●これはファンデーションについてのクチコミです
現在7/323件完了しています

そこまで厚塗りをしなくても、1日中化粧落ちせず、肌にあったナチュラルな色でもとても大好きなファンデーションです。コロナ禍でマスク生活も増えている中で、マスクをしても落ちたりすることが無いファンデーションはやはり重要すべきだと思いました。

このクチコミに当てはまる特徴にチェックをつけてください

製品の悪い点が書かれている

他の製品との比較が書かれている

購入経緯が書かれている

併用した製品が書かれている

クチコミ投稿者なりの使用方法が書かれている

クチコミについて特に気づいたことがあれば教えてください。

こちらに記入してください

次へ 回答を終了する

図 16 軸評価データセット構築システムの画面

表 11 評価軸として使用した単語

化粧品	単語
ファンデーション	カバー、毛穴、肌質、ツヤ、パフ、ブラシ、スポンジ、直し、シミ、下地、成分、乾燥、保湿、時間、崩れ
化粧水	乾燥、保湿、肌、成分、配合、効果、オイル、乳液、クリーム、コスパ、手、値段
リップ	ティッシュオフ、マスク、発色、色、荒れ、乾燥、成分、ピンク、オレンジ、ブラウン、リピート、印象、落ち、指

5.3.2. 分析・考察

評価者17名のうち複数選択を1度もしていなかった女性2名について、回答方法を理解していないものとして分析から除外した。また、回答にばらつきがみられたため少数派回答（評価者の中で同じ評価をした人が3名以下）をしている回数をカウントし、外れ値（ $\text{mean} \pm 1.5\text{SD}$ ）となる男性2名も分析から除外した。結果として、13名（男性6名、女性7名）の回答3332件が得られた。

ここで、今回の軸評価データセット構築では、クチコミの特徴について当てはまるものを選択する形式だったため、回答には個人差が出ないと予測していた。しかし、回答を分析すると、評価者の意見が二分しているものや、評価者の一部のみが選択した少数派回答も存在していた。回答のばらつきを、表12に示す。87.3%のクチコミについて全員の評価が一致している軸もあれば、48.6%のクチコミでしか評価が一致していない軸もあった。また、13.6%のクチコミで評価者の意見が二分した軸も存在しており、軸によって評価の難易度が異なることが示唆された。なお、ここでは4名以上が軸について多数派と逆の回答をした場合二分しているとカウントした。

今回のデータセット構築では323件と多くのクチコミを評価してもらったため、疲れの影響で後半になるにつれて評価が雑になる恐れがある。それを確認するために13名のうち全てのクチコミを評価した8名のデータを用いて、前半で評価した100件と後半で評価した100件における選択した項目数の平均を算出した。その結果前半100件では1.14、後半100件では1.16となり、323件の評価を依頼したが疲れによる大きな影響はないことが明らかになった。これは任意のタイミングで中断できる設計にした効果と考えられる。

また、使用状況によってクチコミに対する評価に差があるのかを明らかにするために、得られた評価をもとに各軸におけるスコアを使用状況別に算出したものを表13に示す。表中のスコアは、「それぞれの化粧品において重要な単語」の軸についてはクチコミに含まれる重要な単語の数の平均、それ以外の軸については、クチコミがその軸に当てはまると評価した評価者の人数割合である。表13より、「製品の悪い点が書かれている」「他の製品との比較が書かれている」以外の軸では実際に使用のクチコミのスコアが高い。次に、評価軸の重みを均等に扱ったときの信憑性スコアを算出した。このスコアは、各軸に0.2点ずつ割り当て、クチコミがその軸に当てはまると評価した評価者の人数割合との積を合計したもので、軸は5つあるため最大で1.0の値をとる。算出した結果、実際に使用のクチコミで0.30、一度だけ使用のクチコミで0.17、未使用のクチコミで0.21となり、実際に使用のクチコミがもっとも高いスコアであった。しかし、実際に使用のクチコミも0.30とスコアが高いとはいえ、適切な方法で軸を用いて信憑性を評価することが必要である。また、一度だけ使用のクチコミよりも未使用のクチコミのスコアがわずかに高い結果となり、表13でも一度使用のクチコミは、「製品の悪い点が書かれている」の軸のみもっともスコアが高く、他の軸ではもっともスコアが低い。これらの結果は、一度だけ使用のクチコミについては、Kajita

表 12 各評価軸の回答のばらつき

評価軸	全員一致	二分
製品の悪い点が書かれている	52.3%	9.6%
他の製品との比較が書かれている	53.9%	8.0%
購入経緯が書かれている	61.6%	6.5%
併用した製品が書かれている	87.3%	1.9%
クチコミ投稿者なりの使用方法が書かれている	48.6%	13.6%

表 13 各評価軸の使用状況別スコア

評価軸	実際に使用	一度使用	未使用
製品の悪い点が書かれている	0.38	0.43	0.24
他の製品との比較が書かれている	0.09	0.04	0.10
購入経緯が書かれている	0.39	0.03	0.30
併用した製品が書かれている	0.06	0.01	0.02
クチコミ投稿者なりの使用方法が書かれている	0.22	0.09	0.09
それぞれの化粧品において重要な単語	3.56	2.64	2.80
合計	4.70	3.24	3.55

ら[16]の実験内で作成しており、クチコミを書くことが実験のメインではなかったため、少し文章が雑なものや、端的すぎるものが多かったことが原因と考えられる。そのため、これ以降の分析では「実際に長期的に使用して投稿されるクチコミ」「全く使用していないのに投稿されるクチコミ」のみを使用する。

5.4. 機械学習を用いた判定精度の検証

評価軸の有用性を明らかにするために、ランダムフォレストを使用し、5分割交差検証を行った。なお、学習データを80%、テストデータを20%とした。その結果、80.2%の精度で分類できることがわかり、それぞれの重要度は表14のように求められた。表14より、「クチコミ投稿者なりの使用方法が書かれている」がもっとも重要度が高いことがわかる。これは、使用方法は実際に使用しないと具体的に書くことができず、未使用のクチコミには具体的な使用方法があまり書かれていなかったことが要因と考えられる。一方で、「他の製品との比較が書かれている」がもっとも重要度が低い結果となった。この結果は、そもそも他の製品と比較しているクチコミが少なかったことが原因と考えられる。今回は大学生および大学院生にクチコミ作成を依頼したため、比較するほど多くの化粧品を使った経験がない

可能性もある。そのため、今後はさらに広い世代から多くのクチコミを収集し、同様の分析を行う必要がある。

データセットの構築の際に記述されたクチコミについての気づきに注目すると、今後についてのことが書かれているというものがあつた。そこで、「今後のこと（リピート購入、季節が変わってからの使用など）について書かれている」という軸を追加した。その結果、精度は80.5%となり、大きな変化はなかつた。この軸の重要度は0.03であり、信憑性に大きく影響しないことがわかる。また、クチコミに対する気づきとして「どのような人におすすめか書かれている」というものもあつた。この特徴を軸として追加すると、精度は81.0%となった。また軸の重要度は0.09であつた。なお、二つの新しい軸を両方追加すると、精度は81.7%となった。

ここで、化粧品によってクチコミの内容も大きく変わるため、化粧品の種類別の分析も行った。分析結果を表15にまとめる。表中では「ファンデーション」を「ファンデ」と表記する。表15より、特にファンデーションについて高精度で信憑性を判別できており、化粧水の精度がもっとも低いことがわかる。そのため、信憑性の評価をより高精度にするためには、化粧品の種類ごとに適した評価軸を用意する必要があり、現在の評価軸はファンデーション

表14 各評価軸の重要度

評価軸	重要度
製品の悪い点が書かれている	0.18
他の製品との比較が書かれている	0.07
購入経緯が書かれている	0.17
併用した製品が書かれている	0.14
クチコミ投稿者なりの使用方法が書かれている	0.28
それぞれの化粧品において重要な単語	0.17

表15 化粧品の種類別の精度と各評価軸の重要度

評価軸	ファンデ	化粧水	リップ
精度	87.7%	79.3%	80.5%
製品の悪い点が書かれている	0.17	0.24	0.19
他の製品との比較が書かれている	0.17	0.10	0.09
購入経緯が書かれている	0.11	0.10	0.19
併用した製品が書かれている	0.06	0.15	0.04
クチコミ投稿者なりの使用方法が書かれている	0.18	0.19	0.23
それぞれの化粧品において重要な単語	0.12	0.14	0.15
今後のことについて書かれている	0.05	0.05	0.05
どのような人におすすめか書かれている	0.14	0.04	0.07

ョンの評価に向いていることが示唆された。ここで、ファンデーションにおいては「どのような人におすすめか書かれている」の重要度が 0.14 と他の化粧品よりも比較的高く、これがファンデーションの高精度での判別に繋がったと考えられる。また、評価軸の重要度に関しては、「クチコミ投稿者なりの使用方法が書かれている」がどの化粧品においても高く、化粧品の種類に関わらず有用な評価軸であることがわかる。

普段化粧をしない人（男性など）と、日常的に化粧を行う人とは、未使用のクチコミに差が生じると考えられる。そこで、ほぼ毎日、または週3~4回化粧をすると答えた女性のみクチコミ作成者を限定して機械学習および判定を行ったところ、精度は 75.2%となり、男性など日常的に化粧をしない人が作成したクチコミが含まれているときよりも精度が下がることが明らかとなった。このことから、日常的に化粧をする女性は、未使用の化粧品であっても普段の経験や化粧特有の用語を用いたクチコミを作成できるが、日常的に化粧をしない人のクチコミの信憑性判定はやや容易であることが示唆される。

5.5. 信憑性評価軸についての議論

今回の軸評価データセットの構築で得られたクチコミに対する気づきより、「今後のことについて書かれている」「どのような人におすすめか書かれている」の軸について検討を新たに行った。その結果、ファンデーションにおいて、「どのような人におすすめか書かれている」という軸は重要であることが明らかとなった。今後、化粧水やリップに対しても、より効果的な軸を探索することによって、さらに高精度な信憑性評価を行うことが可能であると考えられる。

また、本研究で検討した評価軸を用いると、ステルスマーケティングにおいて未使用の化粧品であると気づかれにくいようなクチコミを投稿することが可能である。この問題については、Mukherjee ら[10]が提案していた「クチコミが投稿された時期」を評価軸に組み込むことで、ステルスマーケティングによる組織的なクチコミを判別することが可能であると考えられる。

著者は、クチコミの特徴の回答には個人差が出ないと予測していたが、今回の軸評価データセットの構築では評価者の意見が二分しているものや、評価者の一部のみが選択した少数派回答も存在しており、軸によって評価の難易度が異なることが示唆された。さらに、現在検討している評価軸の中には、機械学習などで判別するのは難しいものも含まれている。そこで、この評価軸を用いたシステムを実現する際には、ユーザが表示したクチコミについて、マイクロタスク的に特徴を選択してもらうことで、多くのデータを収集し、より正確にクチコミの評価を行うことを検討している。また、ユーザが容易に特徴の選択を行えるように、軸については議論を重ね、可能な限り評価の難易度が低いものを採用することも検討している。

第6章 プロトタイプシステムの構築と評価実験

6.1. システム概要

実現を目指す化粧品クチコミの信憑性判断支援システムは、クチコミにおいて他の商品との比較がなされているか、その商品の悪い点も書かれているかといったクチコミに対する各種評価軸に基づき行うものである。また、クチコミに対する軸評価はマイクロタスク化し、膨大なユーザの手で軸の有無などを評価してもらい、そのデータを集約・蓄積し、信憑性評価を行う予定である。

本章では、これまでに検討した化粧品クチコミの信憑性評価軸を用いた信憑性評価システムを、Vue.js を用いた Web サービスとして実装した。実際に構築したシステムを図 17 に示す。画面左半分にクチコミが表示され、その右側に「信憑性評価項目」「信憑性ランク」として評価軸と評価結果 (S・A・B・C) を提示する。信憑性評価項目は後述の軸評価の際当てはまると判断した評価者の人数に応じて色の濃さを変化させている (図 18, 19)。システムで用いた信憑性評価軸を以下に示す。

- それぞれの化粧品において重要な単語の数
- 製品のデメリットについての記述の有無
- 他の製品との比較の有無
- 購入経緯についての記述の有無
- 併用した製品についての記述の有無
- 投稿者なりの使用方法 (HOWTO) についての記述の有無
- 今後のことについての記述の有無
- どのような人におすすめかについての記述の有無

画面右上の②マークをクリックすると信憑性評価項目に関する説明が表示される (図 20)。また、化粧品の種類の絞り込み機能も実装しており、画面中央上部のラジオボタンを用いて選択した化粧品についてのクチコミのみを表示できる。



図 17 システム画面



図 18 信頼性評価項目（信頼性ランク S の場合）



図 19 信頼性評価項目（信頼性ランク C の場合）

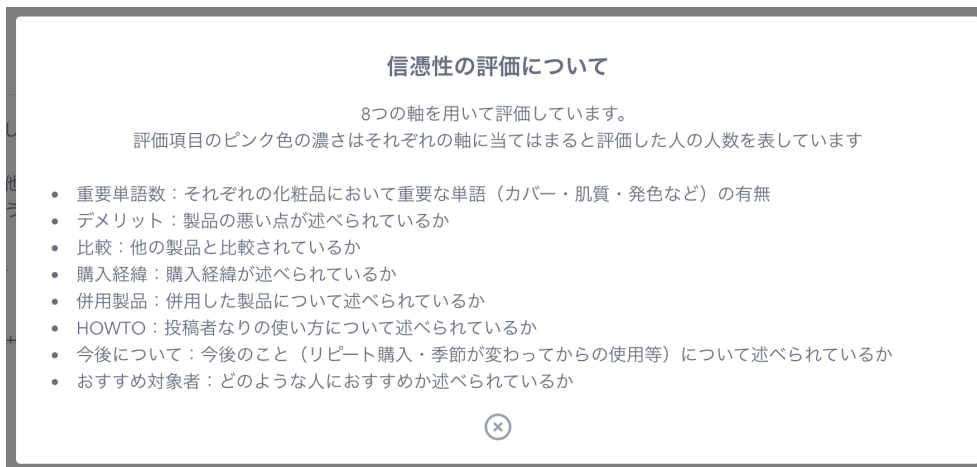


図 20 信憑性評価項目に関する説明の表示

6.2. クチコミに対する軸評価

本章では、4章で収集したファンデーション、化粧水、リップについてのクチコミ各100件ずつの計300件をシステムに使用する。そのため、これまで検討した信憑性評価軸をもとにこれらのクチコミに対して軸評価を行う。ここで、5章の評価軸の検証の際にはクチコミに対する同様の軸評価を13名に依頼していたが、13名のうちすべてのクチコミに対する軸評価を完了させていた8名のデータを利用し、1～8名に人数を減らして5.4節と同様に機械学習を行ったところ、判定精度が83～85%とほとんど変化しなかった（表16）。そのため、本章では著者を含む化粧をする大学生・大学院生3名が軸評価を行った。

軸評価は、クチコミに対してこれまで検討した評価軸の各項目に当てはまる場合1、当てはまらない場合0を入力してもらうものを300件分を行った。3名全員の回答が一致したクチコミの割合を表17に示す。一致率をもっとも高かったのは「併用した製品についての記述の有無」で94.7%であった。また、もっとも低かったのは「投稿者なりの使用方法についての記述の有無」であり、71.0%であった。この一致率に関して、5章でのデータセット構築における軸評価の際も一致率をもっとも高かったのは「併用した製品についての記述の有無」で87.3%、もっとも低かったのは「投稿者なりの使用方法についての記述の有無」で48.6%となっており、今回の軸評価では全体的に一致率が向上していることが確認された。なお、実際にシステムとして運用する際には、今回の軸評価と過去の研究の両方で一致率をもっとも低かった「投稿者なりの使用方法についての記述の有無」を評価者が容易に判断できるような判断基準を設ける必要があると考えられる。

表 16 1～8名による軸評価での判定精度

人数	精度
8C1 の各組み合わせの精度の平均	84.4%
8C2 の各組み合わせの精度の平均	84.0%
8C3 の各組み合わせの精度の平均	83.9%
8C4 の各組み合わせの精度の平均	84.0%
8C5 の各組み合わせの精度の平均	84.3%
8C6 の各組み合わせの精度の平均	84.3%
8C7 の各組み合わせの精度の平均	84.3%
8C8 の各組み合わせの精度の平均	83.9%

表 17 3名の回答の一致率

評価軸	一致率
製品の悪い点についての記述の有無	75.3%
他の製品との比較の有無	87.7%
購入経緯についての記述の有無	79.7%
併用した製品についての記述の有無	93.7%
投稿者なりの使用方法についての記述の有無	71.0%
今後のことについての記述の有無	72.0%
どのような人におすすめの記述の有無	86.7%

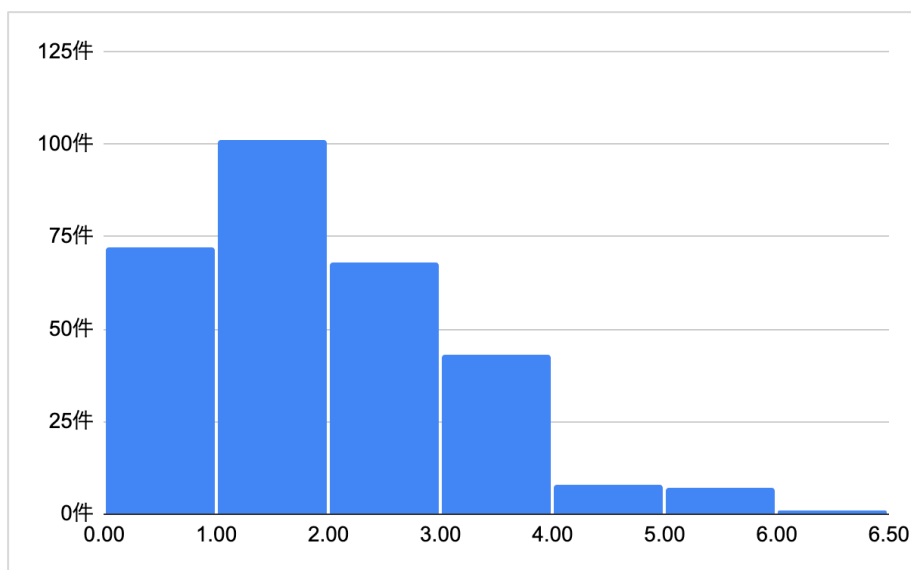


図 21 信憑性スコアの分布

6.3. クチコミの信憑性スコア算出

6.2節で作成した軸評価データを用いてクチコミの信憑性スコアを算出する。スコア算出では「それぞれの化粧品において重要な単語の有無」の評価軸に関して、一つのクチコミに含まれる重要単語数が8単語以上の場合1点、4~7単語の場合0.5点、3単語以下の場合0点とし、その他の7項目の評価軸に関しては3名の軸評価データの平均値をそのクチコミの評価値としたうえで、すべての評価値の和を信憑性スコアとした。スコアの分布を図21に示す。システムではこのスコアの分布をもとに、信憑性スコアが3.5点より大きい場合S、2~3.5点の場合A、1~2点の場合B、1点以下の場合Cとランク付けを行い、これをシステム上で提示する。

6.4. システム評価実験

6.4.1. 実験概要

信憑性評価軸やスコアを提示することの効果を一時的にシステムについてのフィードバックを得るため、評価実験を行った。本実験では化粧品をする大学生11名(女性10名、回答しない1名)にシステムを利用してもらい、信憑性評価の提示に関する意見やシステムの使いやすさについてアンケートを行った。

実験ではまず、システムの利用前に事前アンケートを実施した。その後「ファンデーションが切れてしまったので、新しく購入するファンデーションを探す」といった状況を説明し、その状況に沿いつつシステムを用いてどのようなものを購入するか決めてもらった。同様の試行を化粧水とリップについても行い、最後に事後アンケートを行った。なお、順序効果を防ぐため、ファンデーション、化粧水、リップの各試行はこちらがランダムに指定した順番で行ってもらった。アンケート内容を表18, 19に示す。

6.4.2. 事前アンケートの結果

事前アンケートQ1「化粧品のクチコミの閲覧頻度を教えてください」の結果を図22に示す。もっとも多かったのは時々閲覧する(2週間に1度程度)と回答した人であり、閲覧したことはないと回答した1名を除く10名は化粧品のクチコミを閲覧していることが明らかになった。

Q2「普段化粧品のクチコミを閲覧する際、どのようなものを参考にしますか」への回答では、自身と同じ肌質の人のクチコミや最新のクチコミなどが参考にするクチコミの特徴として挙げられるとともに、具体的なサービス名としてInstagram[9]やTwitter[8]といったSNSや、ECサイトのAmazon[51]、LIPS[7]や@cosme[6]などの化粧品専門のクチコミサイトが挙げられており、人によってどのようなものを参考にするかは異なることが確認された。

表 18 事前アンケートの内容

Q1	化粧品のクチコミの閲覧頻度を教えてください	5段階（非常によく閲覧する～閲覧したことはない）
Q2	普段化粧品のクチコミを閲覧する際、どのようなものを参考にしますか	自由記述
Q3	化粧品のクチコミの信憑性をどの程度気にしていますか	5段階（5：非常に気にする～1：全く気にしない）
Q4	クチコミの信憑性を気にする場合、どのような点で信憑性を判断していますか	自由記述

表 19 事後アンケートの内容

Q5	ファンデーション・化粧水・リップについて、どのようなものを購入したいと思いましたか	自由記述
Q6	化粧品のクチコミの信憑性をどの程度気にしてシステムを利用しましたか	5段階（5：非常に気にした～1：全く気にしなかった）
Q7	システムで表示される信憑性評価項目はどの程度役に立ちましたか	5段階（5：非常に役に立った～1：全く役に立たなかった）
Q8	システムで表示される信憑性ランクはどの程度役に立ちましたか	5段階（5：非常に役に立った～1：全く役に立たなかった）
Q9	システムで提示される信憑性についてどう感じましたか	自由記述
Q10	システムに対する意見・感想をお願いします	自由記述

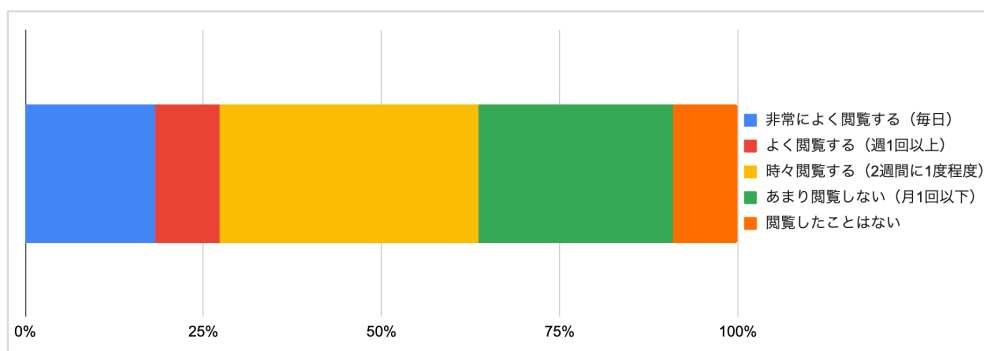


図 22 化粧品に関するクチコミの閲覧頻度 (Q1)

Q3「化粧品のクチコミの信憑性をどの程度気にしていますか」の結果を図23（左）に示す。この結果より11名中7名が5段階中4以上を選択しており、協力者の多くが信憑性を気にしていたが、3名は5段階中2以下を選択しており、信憑性をほとんど気にしていない人もいた。

Q4「クチコミの信憑性を気にする場合、どのような点で信憑性を判断していますか」の回答内容も人によって大きく異なっており、「実際の体験をもとに評価しているかどうかで判断している」「全てを褒めるのではなく、合わなかったものははっきり言っているという部分です」などクチコミ本文から判断している人もいれば、「評価の分布」「同じような意見を持っている人が複数人いるかどうか」など複数のクチコミから総合的に判断している人もいた。

6.4.3. 事後アンケートの結果

Q5の「ファンデーション・化粧水・リップについて、どのようなものを購入したいと思いましたか」についての自由記述では、具体的なクチコミ内容を引用した記述が多かったが、自身と同じ肌質かつ信憑性ランクがSのものを参考にしたという意見もあった。

Q6「化粧品のクチコミの信憑性をどの程度気にしてシステムを利用しましたか」の結果を図23（右）に示す。事前アンケートのQ3と比較すると、5段階中4または5を回答した人数は変化がなかったが、1を回答した人が事前アンケートでは2名いたのに対し、事後アンケートでは0名であった。この2名について回答を詳しくみると、2名とも事後アンケートでは5段階中4または5を回答しており、システムを利用することによって信憑性を意識するようになっていた。しかし、事前アンケートと事後アンケートの回答を比較すると事後アンケートの方が低い値を回答した実験協力者も2名存在していた。

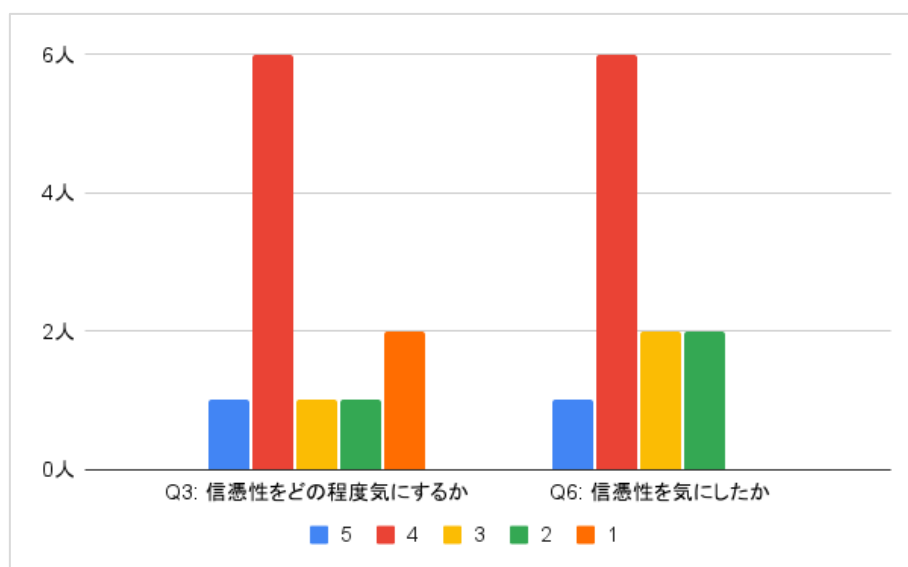


図23 信憑性に対する意識（事前アンケートQ3・事後アンケートQ6）

信憑性評価項目と信憑性ランクがどの程度役に立ったかを比較するため、Q7「システムで表示される信憑性評価項目はどの程度役に立ちましたか」とQ8「システムで表示される信憑性ランクはどの程度役に立ちましたか」の結果を図24に示す。Q7では11名中6名が4以上を選択していたが、4名は2を選択していた。Q8では、8名が4以上を選択しており、信憑性評価項目に比べ信憑性ランクを参考にする傾向があった。

Q9「システムで提示される信憑性についてどう感じましたか」では、「クチコミをととても信じてしまうため、信憑性について人間の偏見なくシステムが判定してくれるのはとても信用できると感じる」といったシステムに対するポジティブな意見のほか、「信憑性が低いと記されているものは見ずにA以上のものを参考にしたが、AかSかはあまり気にしなかった」「信憑性ランクが低いクチコミでも自分の気になる項目について書かれていれば十分参考になったので、信憑性ランクが低いからといって読み飛ばすということはない」など、ユーザ各々の信憑性提示に関する利用方法についての意見も得られた。また、「短文で重要単語数が0のレビューであっても信憑性評価に利用される8つの軸のうち複数個の内容が入ればランクが上がってしまう。基本的にレビューを読むときはランクが高いものを選択して読むのでそれが少々煩わしかった」といった意見もあり、信憑性スコアの算出方法については検討の必要性が示唆された。

Q10「システムに対する意見・感想」に対しては、「自分の重視する項目を選択して、自分に最適化された信憑性ランクが出てきたらいいと思った」「ポジティブな意見かネガティブな意見かが一目でわかるような欄があると良いなと思いました」「化粧品の種類によって多くの人が気にするポイント（例えばファンデなら「カバー力」「落ちやすさ」など）についての記述があるクチコミなのか、パッとみてわかると便利な気がした」などの意見が得られた。

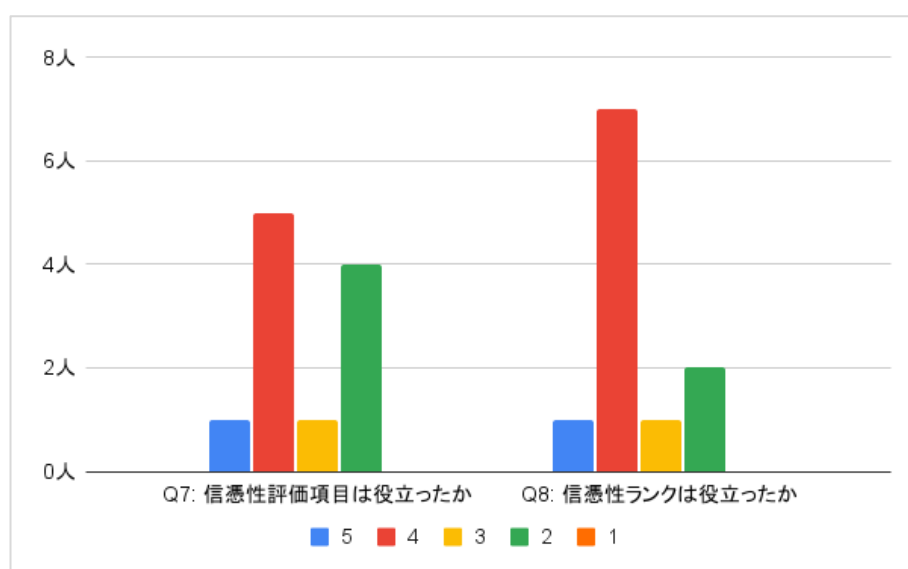


図 24 事後アンケート結果 (Q7・Q8)

6.5. 考察

事前アンケート Q3「化粧品のクチコミの信憑性をどの程度気にしていますか」と事後アンケート Q6「化粧品のクチコミの信憑性をどの程度気にしてシステムを利用しましたか」の比較から、普段信憑性を気にしていない人はシステムを利用することで信憑性を普段よりも意識できる可能性が示唆された。しかし、普段から信憑性を気にしている人の中にはシステムの信憑性評価提示によって信憑性を自身で疑わなくなってしまう人も存在する可能性がある。

また、多くの実験協力者が信憑性評価項目よりも信憑性ランクを参考にシステムを利用したことが明らかになった。しかし、事後アンケート Q7「システムで表示される信憑性評価項目はどの程度役に立ちましたか」と Q8「システムで表示される信憑性ランクはどの程度役に立ちましたか」の回答を個別に見ると、Q7を Q8より高い値で評価している協力者が2名確認できた。この2名について Q9「システムで提示される信憑性についてどう感じましたか」の回答をみると、信憑性評価項目の表示から自身が重視する記述が含まれるクチコミを探していることがわかった。さらに他の実験協力者の Q9「システムで提示される信憑性についてどう感じましたか」、Q10「システムに対する意見・感想をお願いします」への回答から、協力者ごとにクチコミの見方が異なり、信憑性ランクのみで閲覧するクチコミを決める人や、ランクが一定以上かつ自身が重視する信憑性評価項目を含むものを閲覧する人などが存在することが明らかになった。現在のシステムでは信憑性評価ランクの根拠として全信憑性評価項目を提示していたが、評価項目で検索する機能を実装することで様々なユーザの利用方法に対応可能なシステムにしていくことができる。また、評価項目やランクのソート機能を実装することでよりユーザが自身のニーズに合うクチコミを探しやすくなると考えられる。

また、Q10で得られた意見から、重要単語数に関して数字のみでなく詳細も表示することでより便利なシステムにできると考えられる。ここで、現在の重要単語数はクチコミに含まれる重要単語の個数を信憑性評価軸として用いているが、システムに用いたクチコミの中には Twitter[8]から収集した短いクチコミも存在しており、単語数のみでは文章量の差を考慮できていなかった。そのため、今後は文章量に対する重要単語の数にするなど文章量を考慮した軸を検討する必要がある。また、現在はスコア算出時のそれぞれの評価項目に対する配点を均等にしているが、信憑性判定精度を向上させるために重み付けを行うことも検討している。さらに、現状の評価軸は文章についてのもののみで構成されているため、投稿者や画像に関する軸などを検討することで、より高精度な推定が可能になると考えられる。

第7章 システムの改良と評価実験

7.1. システムの改良

6章で得られたフィードバックをもとにシステムの改良を行った。具体的な改良箇所は以下の4点である。

- 検索：任意のキーワードを含むクチコミを検索する
- 絞り込み：任意の信憑性評価項目を含むクチコミに絞り込む
- 重要単語指数：重要単語数について文章量を考慮して数値化
(クチコミに含まれる重要単語数÷クチコミの文字数×10)
- UIの改善：化粧品の種類の切り替えをラジオボタンではなくタブ切り替えに変更

なお、重要単語数を重要単語指数に変更したことで信憑性スコアの分布が変動したため、信憑性ランクの閾値についても変更を行った。具体的には、信憑性スコアが3点より大きい場合 S、2~3点の場合 A、1~2点の場合 B、1点以下の場合 C とした。システム画面と絞り込み選択画面を図 25、26 に示す。

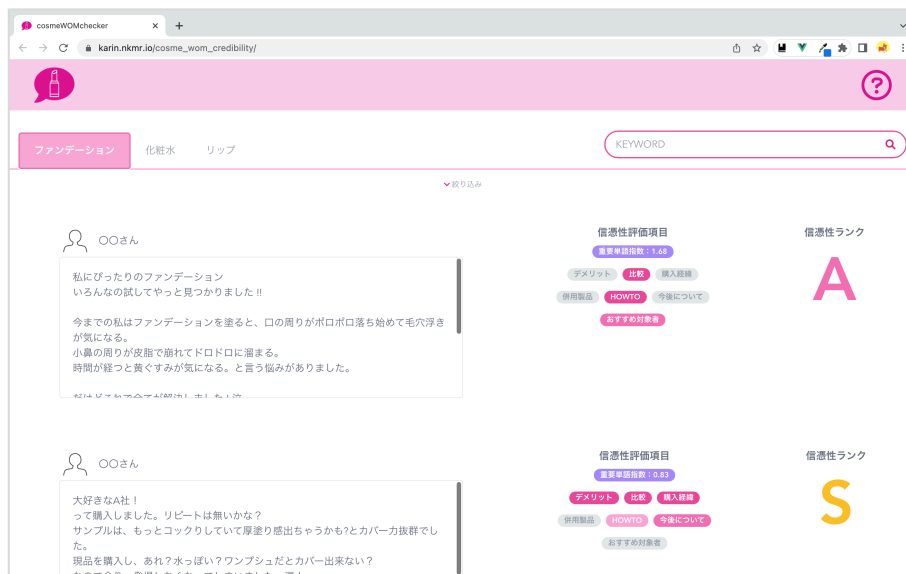


図 25 改良後のシステム画面

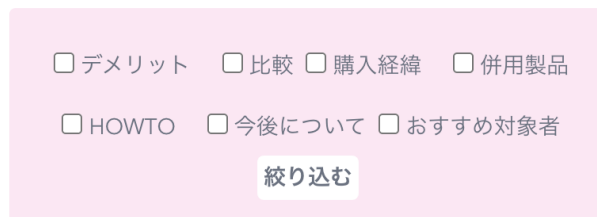


図 26 絞り込み選択画面

7.2. 実験概要

改良したシステムを用い、システム利用による信憑性への意識の変化を確認することやシステムの新しい機能を利用する様子を観察することを目的とし、6.4節と同様の手順で再度システム評価実験を実施する。今回の実験ではより詳細な分析を行うために画面収録を行い、ユーザがどのようにシステムを利用するのかを動画から観察する。

実験協力者は化粧をする女子大学生・大学院生 15 名である。6.4 節同様、まず事前アンケートで化粧品のクチコミの利用状況について回答してもらった後システムを 30 分程度利用してもらい、最後に事後アンケートでシステムの使用に関する質問に回答してもらった。事前アンケート・事後アンケートの質問項目は表 18, 19 と同様である。なお、順序効果を防ぐため、6.4 節と同様にファンデーション、化粧水、リップの各試行はこちらがランダムに指定した順番で行ってもらった。

7.3. 結果

7.3.1. 事前アンケートの結果

事前アンケートでは、6.4 節と同様に化粧品に関するクチコミの閲覧頻度や、どのようなものを参考にするのか、閲覧時に信憑性を気にするか、どのような点で信憑性を判断するかを回答してもらった。

Q1 の化粧品に関するクチコミの閲覧頻度を図 27 に示す。「非常によく閲覧する」と回答したのは 0 名であったが、「よく閲覧する」と回答した実験協力者は 6 名であり、6.4 節での実験より増加していた。

Q2 「普段化粧品のクチコミを閲覧する際、どのようなものを参考にしますか」への回答では、6.4 節と同様、自身と同じ肌質であることや最新の投稿であること、要点が端的に述べられていることなどが参考にするクチコミの特徴として挙げられるとともに、SNS や

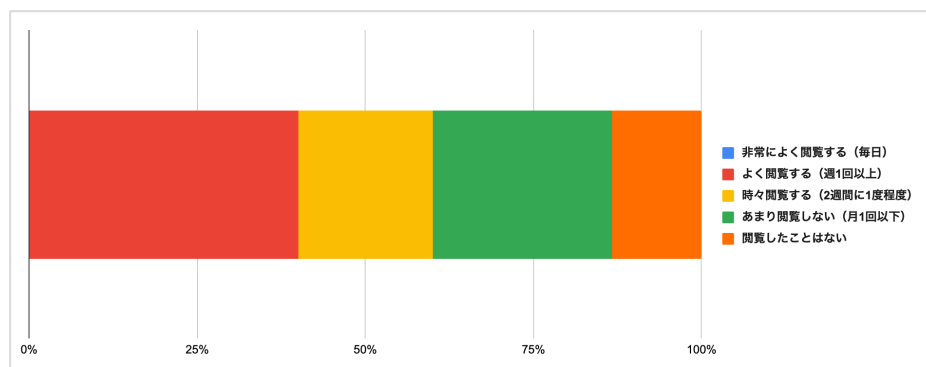


図 27 化粧品に関するクチコミの閲覧頻度

YouTube などのサービスが挙げられており、人によってどのようなものを参考にするかは異なることが確認された。

Q3「化粧品のクチコミの信憑性をどの程度気にしていますか」の結果を図 28（左）に示す。この結果より 15 名中 10 名が 5 段階中 4 以上を選択しており、信憑性を気にしていたが、3 名は 5 段階中 2 以下を選択しており、信憑性をほとんど気にしていない人もいた。

Q4「クチコミの信憑性を気にする場合、どのような点で信憑性を判断していますか」の回答内容も 6.4 節と同様人によって大きく異なっており、「良い点、悪い点など書かれているか、他社の製品と比べてどうか、文章量が多いか少ないか、化粧崩れが起きるまでのシチュエーションや時間など具体的に記述されているか」「自分の肌質を書いているレビューは気をつけて読むようにしている」などクチコミ本文から判断している人もいれば、「他の投稿がアフィリエイトに誘導されていないか、バズを狙いすぎているか（よくある構文使っていないか）」「同じことを書いている人がどれくらいいるか」など複数のクチコミから総合的に判断している人もいた。

7.3.2. 画面収録を用いた分析と結果

今回の実験では、新たに画面収録機能を利用し、動画を用いた分析を行う。まず、動画を用いて実験中に行われた検索・絞り込みなどのアクションをカウントした。実験全体と化粧品の種類別に分けたアクション回数を表 20 に示す。なお表中では「ファンデーション」を「ファンデ」と表記する。表より、実験全体を通してカウントすると検索がもっとも利用されており、アクションの過半数を検索が占めていることがわかる。また、実験協力者の中に

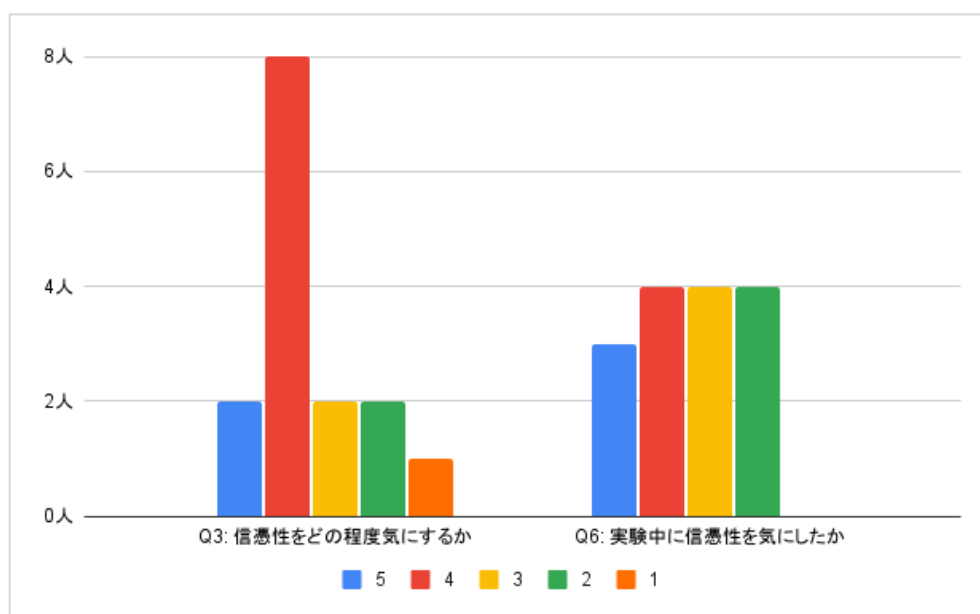


図 28 信憑性に対する意識
(事前アンケート Q3・事後アンケート Q6)

は検索のみ、または絞り込みのみを使用し、その他の機能を使用しなかった人も存在した。さらに、動画からどのようにクチコミを閲覧しているか観察すると、検索・絞り込みなどを行い、提示されたものを上から順に閲覧する人もいれば、ランクが低いものを飛ばして閲覧する人も確認できた。このように、システムの使用方法が人により異なる点は6章での実験と同様であった。

また、表20より化粧水では他の化粧品と異なり、絞り込み機能が多く使われた。ここで、絞り込みにおいて各項目が選択された回数を表21に示す。表より、化粧水では他の化粧品に比べて「比較」「おすすめ対象者」が多く選択されていることがわかる。また、「デメリット」はどの化粧品のときでも絞り込み時に多く選択されていた。

7.3.3. 事後アンケートの結果

Q5の「ファンデーション・化粧水・リップについて、どのようなものを購入したいと思いましたか」についての自由記述では、具体的なクチコミ内容を引用した記述のほか、検索した言葉や絞り込みで使った項目の記述もあった。

Q6「化粧品のクチコミの信憑性をどの程度気にしてシステムを利用しましたか」の結果を図28(右)に示す。事前アンケートのQ3では、5段階中1または2を回答した実験協力

表20 システム使用中に使用された機能と回数

アクション	全体	ファンデ	化粧水	リップ
検索・絞り込みなし	27	10	9	8
検索	100	34	21	36
絞り込み	76	26	28	23
検索+絞り込み	14	5	6	3
アクション合計	190	65	55	62

表21 絞り込み項目が選択された回数

項目	全体	ファンデ	化粧水	リップ
デメリット	40	15	11	14
比較	26	8	13	5
購入経緯	1	0	0	1
併用製品	3	2	1	0
HOWTO	17	8	3	6
今後について	13	3	5	5
おすすめ対象者	36	11	13	11

者は3名であったが、この3名のうち2名は事後アンケートにおいて5段階中3または4と回答しており、信憑性を普段よりも意識したと考えられる。しかし、事前アンケートと事後アンケートの回答を比較すると事後アンケートの方が低い値を回答した実験協力者も4名存在していた。

信憑性評価項目と信憑性ランクがどの程度役に立ったかを比較するため、Q7「システムで表示される信憑性評価項目はどの程度役に立ちましたか」とQ8「システムで表示される信憑性ランクはどの程度役に立ちましたか」の結果を図29に示す。Q7では15名中13名が4以上を選択していたが、2名は3以下を選択していた。Q8では、8名が4以上を選択しており、信憑性ランクに比べ信憑性評価項目が参考にされる傾向があった。これは6章と異なる結果であった。

Q9「システムで提示される信憑性についてどう感じましたか」では、「ランクが高いものほど、自分が求めているようにデメリットや今後の利用について触れていて、信頼できると感じました」「表現が好みで何となくこれは信用できるなと思ったクチコミでも、信憑性が低いと提示されると一度冷静になってそのクチコミを読むことができよかった」といったシステムに対するポジティブな意見のほか、「信憑性ランクの高い口コミを優先的に読むようになりました」「信憑性の高低に関わらず、自分の関心のある項目について書かれているクチコミには注目した」など、ユーザ各々の利用方法についての意見も得られた。また、「信憑性のランクや重要単語指数が何で決められているのか、あまり分からなかったため、あまり信用ならないなと感じてしまいました」という意見もあったため、信憑性ランクの決め方についてより詳細な説明をユーザに提示する必要性が示唆された。

Q10「システムに対する意見・感想」に対しては、「気になった口コミをお気に入り登録して見返せるとより便利だと思った」「SNSだとデメリットを知りたいときに『落ちやす

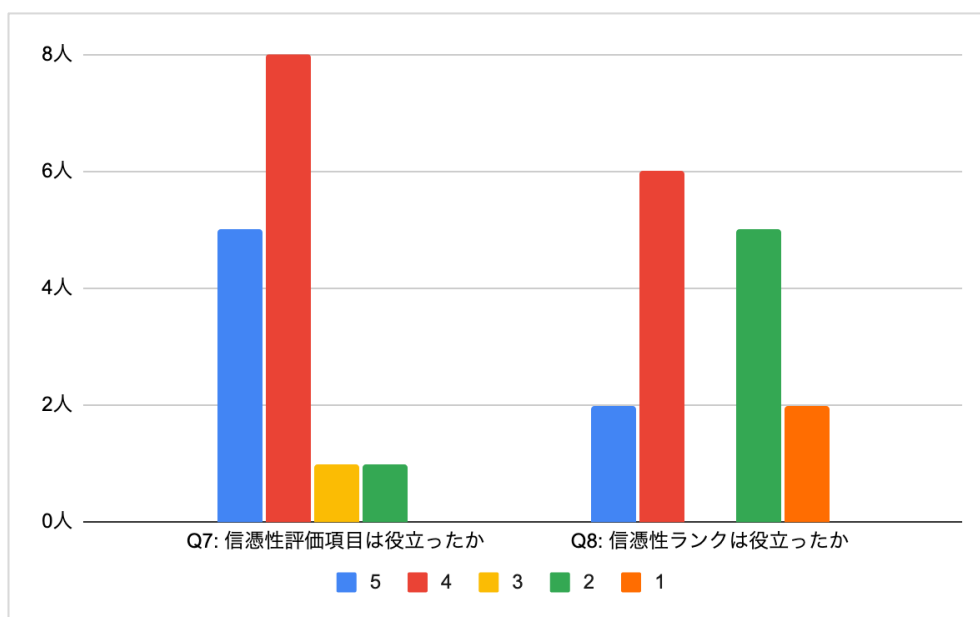


図 29 事後アンケート結果 (Q7・Q8)

い』など単語でしか絞れないので、自分の想定していないデメリットとかが存在する時にそれを知ることができないため、デメリット全般で検索できるこのシステムはすごく便利だなと思いました」「ランクでも絞り込みができるとさらに便利になるのではないかと思います」などの意見が得られた。

7.4. 考察

事前アンケート Q3 と事後アンケート Q6 の比較より、6.4 節の結果と比較すると意識の低下や変化しなかった実験協力者も多く存在した。これは今回の実験協力者が、元々信憑性を意識する人が多かったためではないかと考えられる。しかし、実験前は信憑性をあまり意識していなかった実験協力者3名のうち2名は実験中に信憑性を意識したと回答したため、6章の実験結果と同様に普段信憑性を意識しない人に対して本システムは有効であると考えられる。

絞り込みにおいて化粧水では他の化粧品に比べて「おすすめ対象者」が多く選択されていた。これは化粧水がファンデーションやリップといった他の化粧品に比べ、肌質や肌悩みによって合うもの・合わないものが顕著であり、自分に合うものかを重視してクチコミを閲覧したためであると考えられる。また、「比較」も他の化粧品の場合より多く選択されていた。化粧水で「比較」を含む絞り込みを行った人は、「乾燥」「ベタつき」などのワードで検索機能も併用しており、自分の肌悩みにより効くものを探索したと考えられる。

また、「デメリット」はどの化粧品のときでも絞り込み時に多く選択されており、多くの実験協力者が重視する項目であることが示唆された。事後アンケートでも、「普段 SNS でクチコミを探しているのですが、SNS だとデメリットを知りたいときに「落ちやすい」など単語でしか絞れないので、自分の想定していないデメリットとかが存在する時にそれを知ることができないため、デメリット全般で検索できるこのシステムはすごく便利だなと思いました」といった意見が得られたため、実験協力者にとって有用な項目であったと考えられる。

今回の事後アンケートでは、6章の結果とは異なり、信憑性評価項目が信憑性ランクよりも参考にされていることが明らかとなった。これは、信憑性評価項目の絞り込み機能を実装したことで使いやすさが向上したことが要因と考えられる。今後、信憑性ランクを用いた絞り込みやソートなどの機能を実装し、信憑性ランクについても使いやすさを向上させることでさらに有用性の高いシステムになることが期待される。

第8章 本システムの応用可能性の議論

本システム及び信憑性評価軸の化粧品以外のクチコミへの応用可能性および化粧品のクチコミの独自性について議論を行うため、アンケート調査を行った。最初に予備調査として、大学生・大学院生45名に8種類のクチコミについて利用方法に関するアンケート調査を実施した。この結果を踏まえ、クチコミの種類を4種類に絞り、同様の調査をYahoo!クラウドソーシングで実施した。

8.1. 大学生・大学院生を対象とした予備調査

8.1.1. 調査概要

予備調査は2022年10月20日～2022年11月26日に大学生・大学院生45名（男性22名・女性22名・回答しない1名）を対象に実施した。

本調査では、インターネット上のクチコミを閲覧するかと閲覧するクチコミのジャンルを回答してもらった後、下記の8種類のクチコミについて、閲覧時に利用するサービス、閲覧頻度、閲覧目的、参考にするクチコミの特徴を回答してもらった。

- 飲食店
- 化粧品
- 映像作品（映画、ドラマ、アニメ等）
- 病院
- 美容院・サロン
- 宿泊施設
- スポーツ用品
- 電子機器

8.1.2. 結果

結果を図30, 31に示す。図30より、回答者の8割以上がインターネット上のクチコミを閲覧するかという質問に対し、5（よく見る）～1（全く見ない）の5段階中4または5と回答した。また、図31よりクチコミを閲覧する人が多いジャンル上位3つは飲食店、宿泊施設、電子機器であった。ここで、ほぼ全員が化粧をすると考えられる女性の回答者のうち、化粧品のクチコミを閲覧しないのは3名のみであり、男性も2名から閲覧すると回答があった。

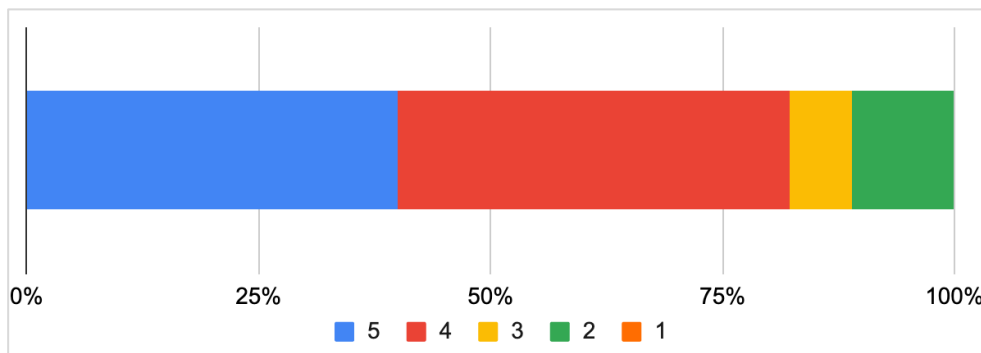


図 30 インターネット上のクチコミを閲覧するか

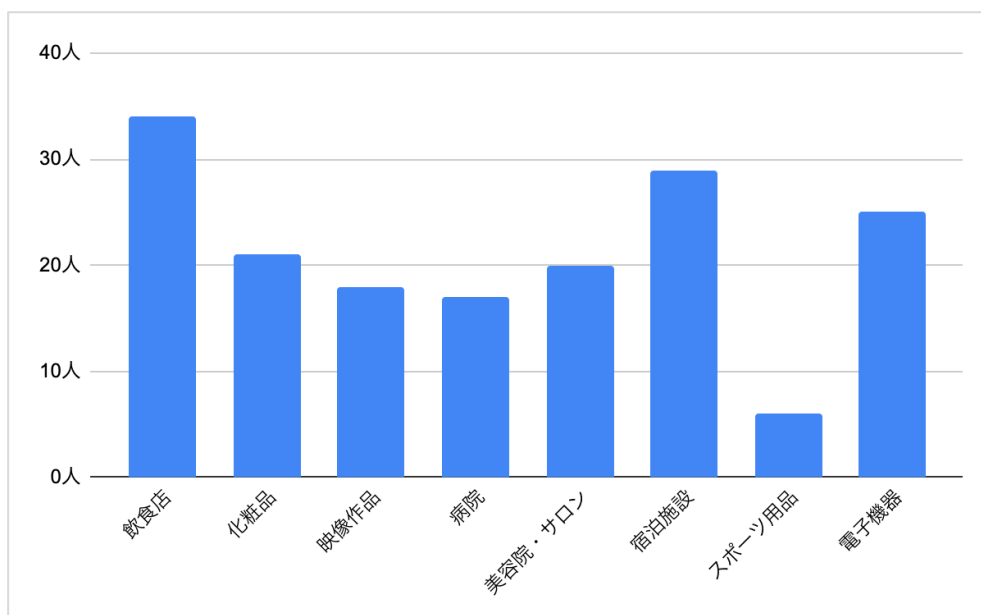


図 31 閲覧するクチコミのジャンル

図 32 にクチコミ閲覧時に利用するサービスの結果を示す。なお、これ以降図 35 までのグラフでは各ジャンルのクチコミを閲覧する人のみの回答の分布を示す。多くのジャンルでクチコミサイトが利用される傾向にあったが、化粧品はクチコミサイトと同程度の割合で SNS や動画サイトも利用されていた。これに近い傾向が電子機器のクチコミにも見られた。

図 33 にそれぞれのクチコミジャンルにおける閲覧頻度を示す。週 1 回以上見ると回答した人の割合は化粧品が最も高く、次いで電子機器や飲食店が高かった。

図 34 にクチコミを閲覧する目的を示す。化粧品のクチコミを閲覧する目的は主に「利用する商品を探すため」「気になる商品について情報を集めるため」「利用候補を比較するため」であり、飲食店、宿泊施設、電子機器にも類似の傾向が見られた。また化粧品特有の特徴として、「利用予定に関わらず何気なく閲覧する」が他のジャンルに比べ多く選択されていた。

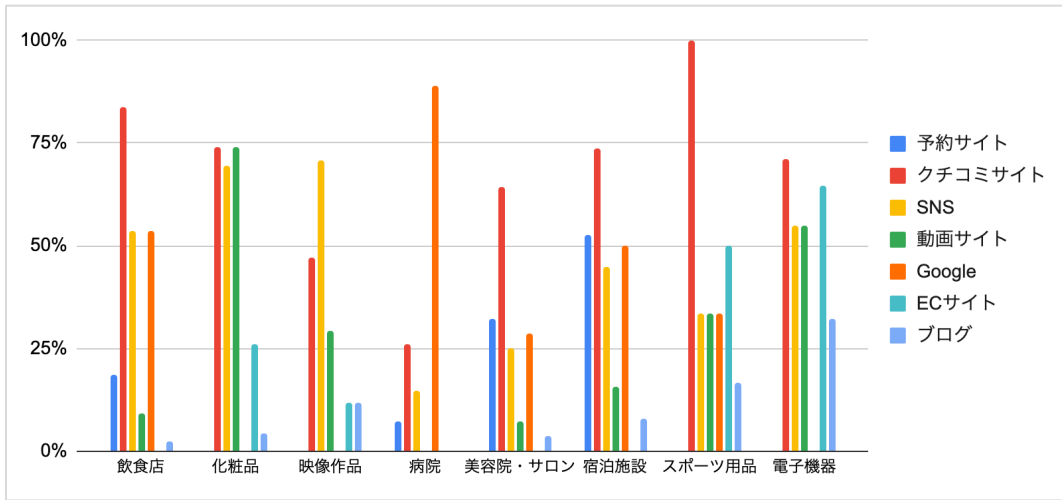


図 32 利用するサービス

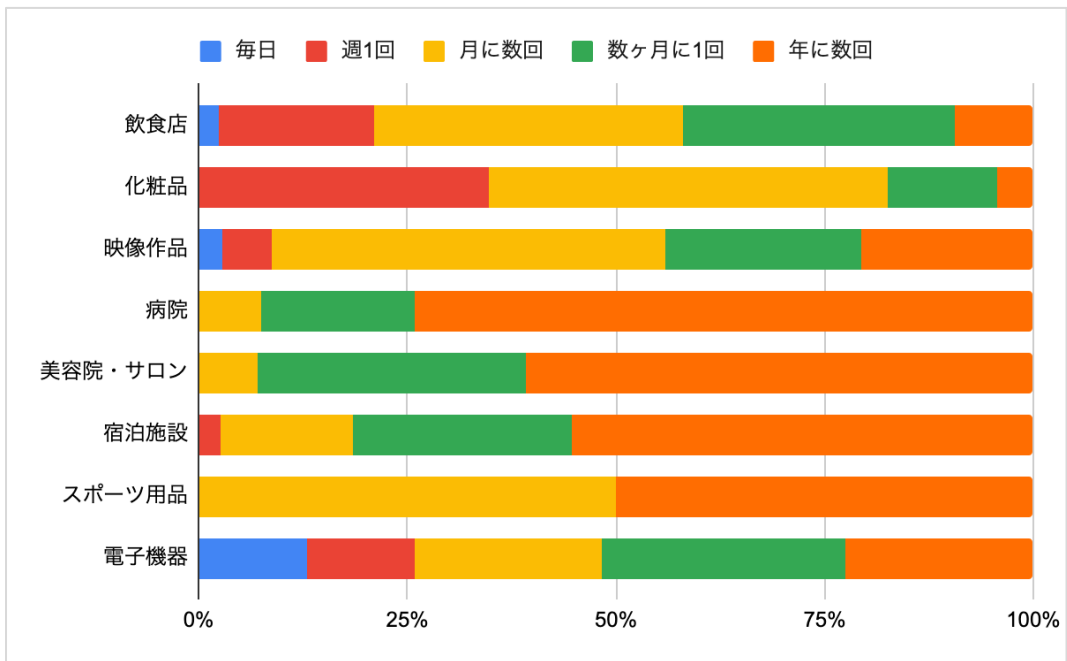


図 33 閲覧頻度

図 35 に参考にするクチコミの特徴を示す。化粧品のクチコミでは画像付き、具体的、商品の悪い点・良い点が書かれているなどの特徴をもつクチコミが参考にされやすいことが明らかになった。また、飲食店、宿泊施設、スポーツ用品、電子機器で類似した結果となった。また、他のジャンルと比較すると化粧品は投稿数やフォロワー数の多いクチコミが参考にされる割合が高かった。

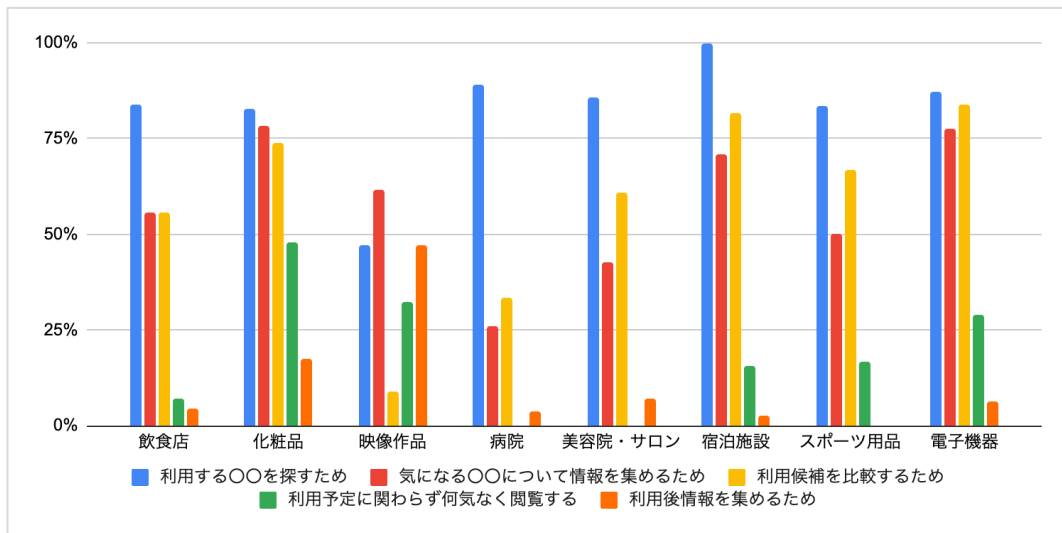


図 34 閲覧目的

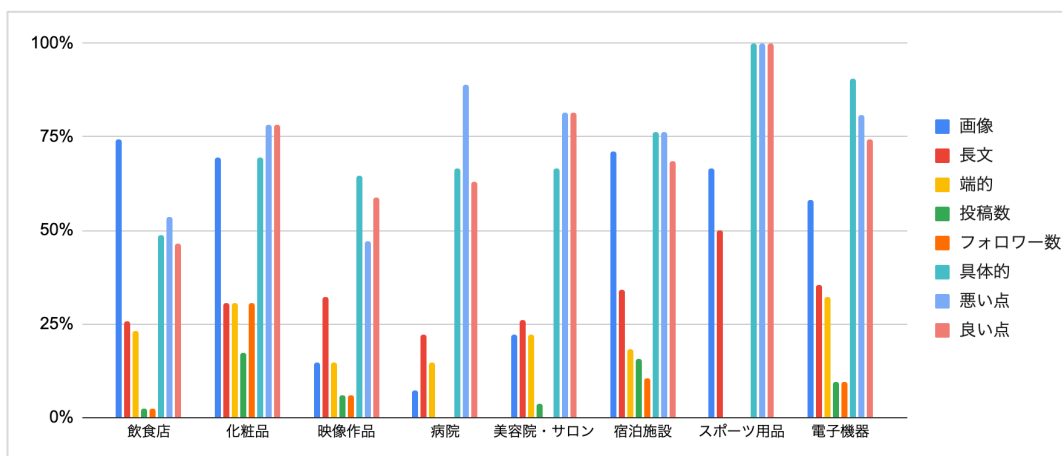


図 35 参考にするクチコミの特徴

8.2. Yahoo!クラウドソーシングを利用した調査

8.2.1. 調査概要

本調査は2022年11月24～25日にYahoo!クラウドソーシングにて実施した。広い年代の回答者を集めるために、20～30代向け・40代以上向けに分けて2回実施した。予備調査の結果をふまえて、本調査では化粧品と類似した特徴をもつことが多かった飲食店、化粧品、宿泊施設、電子機器の4種類のクチコミについて閲覧頻度や閲覧時に利用するサービス、閲覧目的、参考にするクチコミの特徴を回答してもらった。表22にアンケート項目を示す。

表 22 アンケート項目

Q1	年齢	4 択 (20~24 歳, 25~29 歳, 30~34 歳, 35~39 歳) または 7 択 (40~44 歳, 45~49 歳, 50~54 歳, 55~59 歳, 60~64 歳, 65~69 歳, 70 歳以上)
Q2	性別	3 択 (男性, 女性, 回答しない)
Q3	インターネット上のクチコミを閲覧しますか	5 段階 (毎日見る, 週に一度見る, 月に 1 度見る, 年に一度見る, 全く見ない)
Q4	飲食店のクチコミを閲覧しますか	Q3 と同様
Q5	飲食店のクチコミを閲覧するときは どういったサービスを利用しますか (複数選択可)	7 択 (クチコミサイト, 予約サイト, SNS (Twitter, Instagram 等), 動画サイト (YouTube, TikTok 等), Google のクチコミ, ブログ, 飲食店のクチコミを閲覧しない)
Q6	飲食店のクチコミを閲覧する目的として 当てはまるものを選択してください (複数選択可)	6 択 (利用する店を探すため, 気になる店について情報を集めるため, 利用候補の複数店舗を比較するため, 利用予定に関わらず何気なく閲覧する, 利用後その店に関する情報を集めるため, 飲食店のクチコミを閲覧しない)
Q7	飲食店のクチコミの中でどのようなものを 参考にしますか (複数選択可)	9 択 (画像付きのクチコミ, 文章が長いクチコミ, 文章が端的なクチコミ, クチコミの投稿数が多いユーザによるクチコミ, フォロワー数が多いユーザによるクチコミ, 具体的に書かれたクチコミ, 店の悪い点が書かれたクチコミ, 店の良い点が書かれたクチコミ, 飲食店のクチコミを閲覧しない)

表 22 アンケート項目

Q8	化粧品のカチコミを閲覧しますか	Q3 と同様
Q9	化粧品のカチコミを閲覧するときほど ういったサービスを利用しますか	6 択 (カチコミサイト, 公式サイト, SNS (Twitter、Instagram 等), 動画サイト (YouTube、TikTok 等), ブログ, 化粧品のカチコミを閲覧しない)
Q10	化粧品のカチコミを閲覧する目的として 当てはまるものを選択してください	6 択 (利用する商品を探すため, 気になる商品について情報を集めるため, 利用候補の複数商品を比較するため, 利用予定に関わらず何気なく閲覧する, 利用後その商品に関する情報を集めるため, 化粧品のカチコミを閲覧しない)
Q11	化粧品のカチコミの中で どのようなものを参考にしますか	9 択 (画像付きのカチコミ, 文章が長いカチコミ, 文章が端的なカチコミ, カチコミの投稿数が多いユーザによるカチコミ, フォロワー数が多いユーザによるカチコミ, 具体的に書かれたカチコミ, 商品の悪い点が書かれたカチコミ, 商品の良い点が書かれたカチコミ, 化粧品のカチコミを閲覧しない)
Q12	宿泊施設のカチコミを閲覧しますか	Q3 と同様
Q13	宿泊施設のカチコミを閲覧するときほど どういったサービスを利用しますか	7 択 (カチコミサイト, 予約サイト, SNS (Twitter、Instagram 等), 動画サイト (YouTube、TikTok 等), Google のカチコミ, ブログ, 宿泊施設のカチコミを閲覧しない)
Q14	宿泊施設のカチコミを閲覧する目的として 当てはまるものを選択してください	6 択 (利用する宿泊施設を探すため, 気になる宿泊施設について情報を集めるため, 利用候補の複数施設を比較するため, 利用予定に関わらず何気なく閲覧する, 利用後その宿泊施設に関する情報を集めるため, 宿泊施設のカチコミを閲覧しない)
Q15	宿泊施設のカチコミの中でどのような ものを参考にしますか	9 択 (画像付きのカチコミ, 文章が長いカチコミ, 文章が端的なカチコミ, カチコミの投稿数が多いユーザによるカチコミ, フォロワー数が多いユーザによるカチコミ, 具体的に書かれたカチコミ, 宿泊施設の悪い点が書かれたカチコミ, 宿泊施設の良い点が書かれたカチコミ, 宿泊施設のカチコミを閲覧しない)

表 22 アンケート項目

Q16	電子機器のクチコミを閲覧しますか	Q3と同様
Q17	電子機器のクチコミを閲覧するときはどういったサービスを利用しますか	7 択 (クチコミサイト, SNS (Twitter, Instagram 等), 動画サイト (YouTube, TikTok 等), EC サイトのレビュー (Amazon, 楽天等), Google のクチコミ, ブログ, 電子機器のクチコミを閲覧しない)
Q18	電子機器のクチコミを閲覧する目的として当てはまるものを選択してください	6 択 (利用する商品を探すため, 気になる商品について情報を集めるため, 利用候補の複数商品を比較するため, 利用予定に関わらず何気なく閲覧する, 利用後その商品に関する情報を集めるため, 電子機器のクチコミを閲覧しない)
Q19	電子機器のクチコミの中でどのようなものを参考にしますか	9 択 (画像付きのクチコミ, 文章が長いクチコミ, 文章が端的なクチコミ, クチコミの投稿数が多いユーザによるクチコミ, フォロワー数が多いユーザによるクチコミ, 具体的に書かれたクチコミ, 商品の悪い点が書かれたクチコミ, 商品の良い点が書かれたクチコミ, 電子機器のクチコミを閲覧しない)

8.2.2. 結果

回答者の中には、回答の辻褄が合っていない不真面目な回答者も含まれていたため、集まった 2000 名の回答者のうち 191 名を除外し、1809 名（男性 973 名・女性 813 名・回答しない 23 名）の回答を分析対象とした。

回答者の年齢分布 (Q1) を図 36 に示す。回答者内で最も多かった年齢層は 35～39 歳であったが、大きな偏りなく広い世代からの回答を収集できた。インターネット上のクチコミの閲覧頻度 (Q3) を図 37 に示す。7 割以上の回答者が毎日～週に 1 度程度クチコミを閲覧していることが明らかになった。

図 38 にそれぞれのクチコミの閲覧頻度 (Q4, Q8, Q12, Q16) を示す。なお、これ以降図 44 までのグラフでは 8.1.2 項同様クチコミを閲覧する人のみの回答の分布を示している。最も閲覧頻度が高いのは飲食店のクチコミであり、予備調査とは異なる結果となった。ここで、20～24 歳の各ジャンルの閲覧頻度を図 39 に示す。図より、化粧品が最も閲覧頻度が高く、次が飲食店となっており、予備実験と同様の結果となった。

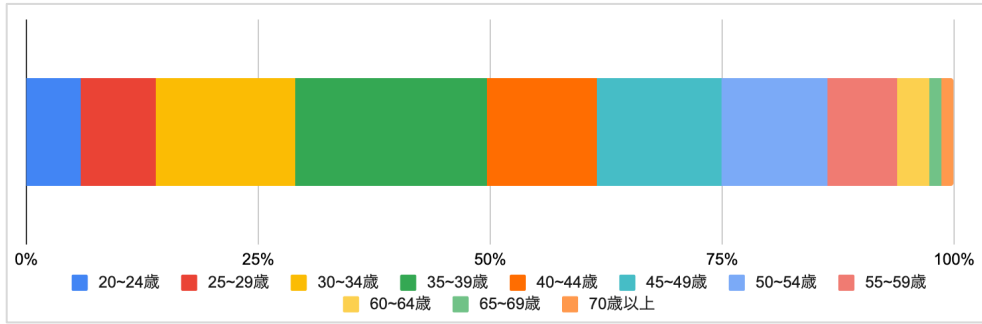


図 36 回答者の年齢分布

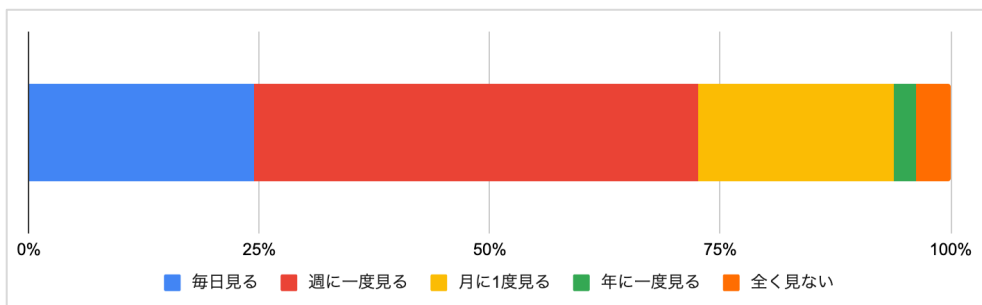


図 37 インターネット上のクチコミの閲覧頻度

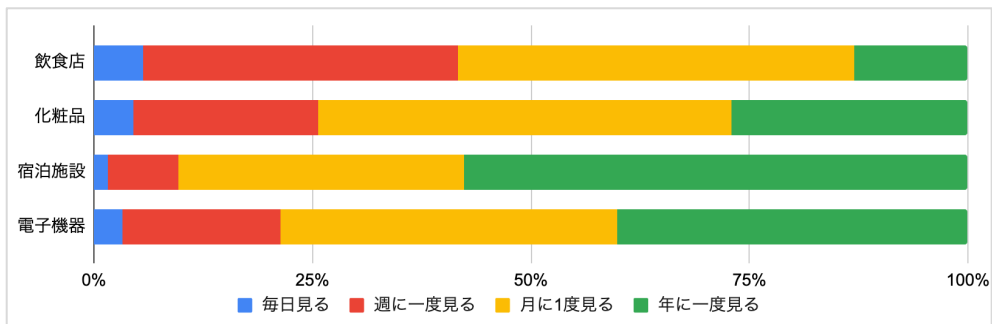


図 38 クチコミの閲覧頻度

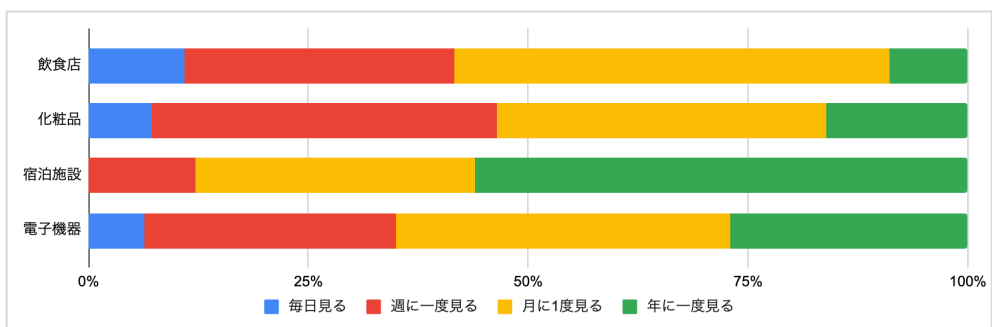


図 39 20~24歳のクチコミ閲覧頻度

図 40 にクチコミ閲覧時に利用するサービス (Q5, Q9, Q13, Q17) についての結果を示す。化粧品は予備調査同様クチコミサイトが多く利用されていたが、予備調査の結果と異なる点として SNS や動画サイトの利用率が下落した。しかし、利用サービスについても 20～24 歳に絞ると SNS の利用率が最も高くなった (図 41)。

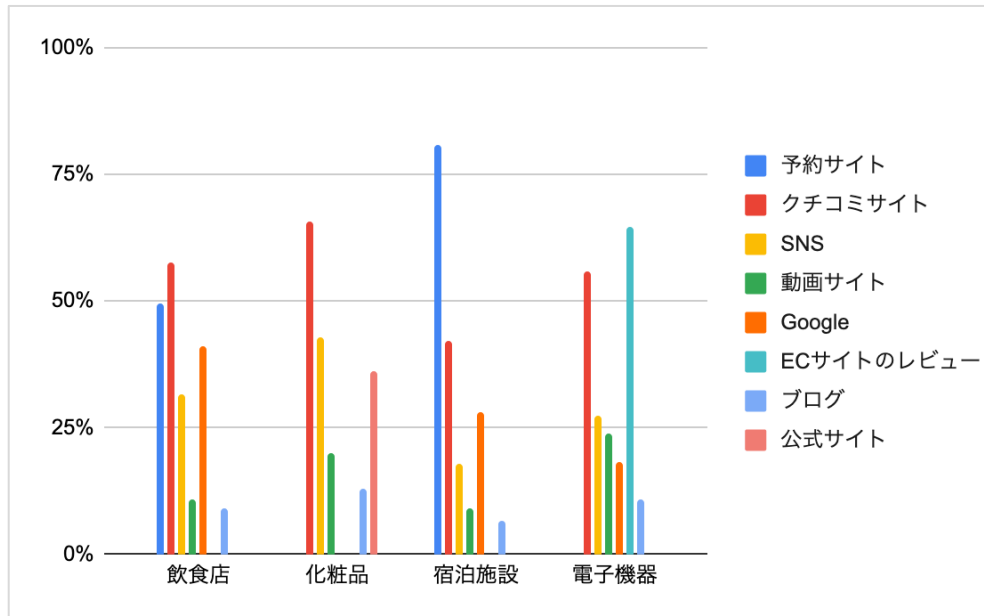


図 40 閲覧時に利用するサービス

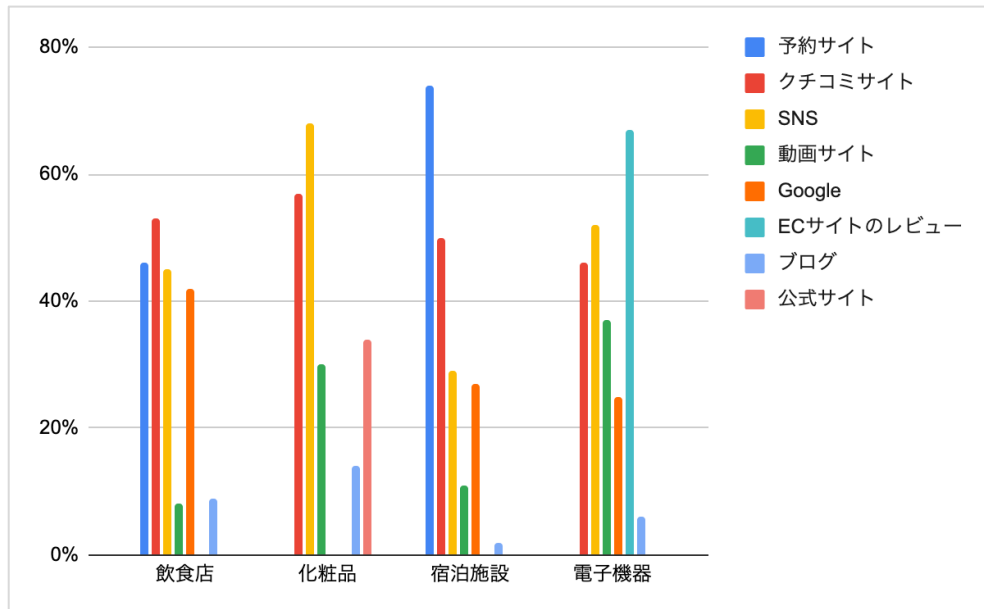


図 41 20～24 歳の利用サービス

図 42 に閲覧目的 (Q6, Q10, Q14, Q18) の分布を示す。どのジャンルのクチコミも利用する店・商品・施設を探す時や気になる店・商品・施設について調べるときに多く利用されることが明らかになった。予備調査では化粧品のクチコミを利用予定に関わらず閲覧する人が多くいたが、今回の調査では他のジャンルと大きな差はなかった。

図 43 に参考にするクチコミの特徴を示す。画像付きのクチコミ、具体的に書かれたクチコミはどのジャンルにおいても多くの人々が参考にすることが明らかとなった。また 20~24 歳が参考にするクチコミの特徴をみると、全ジャンルを通して文章が端的なクチコミを参

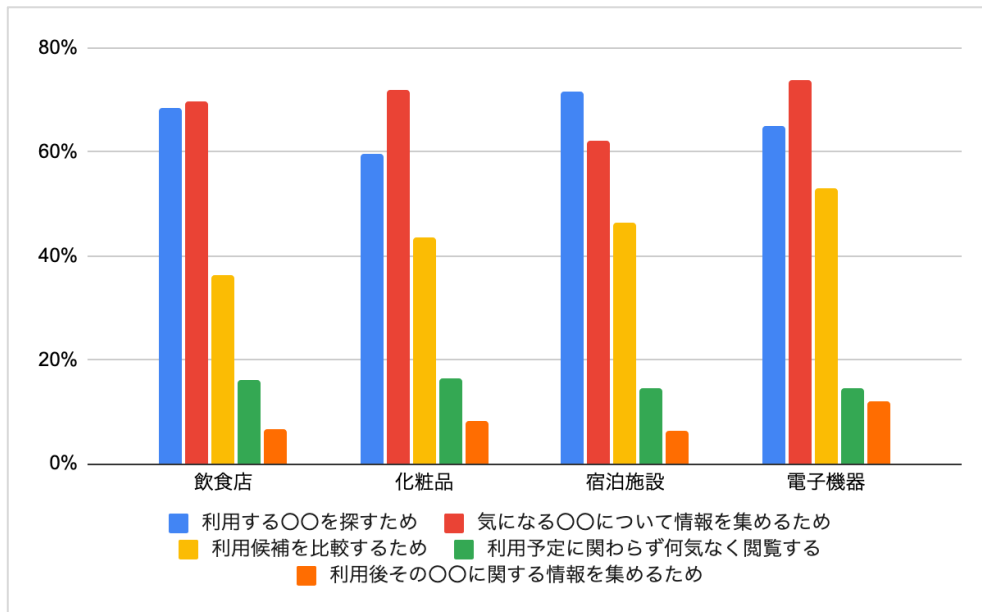


図 42 閲覧目的

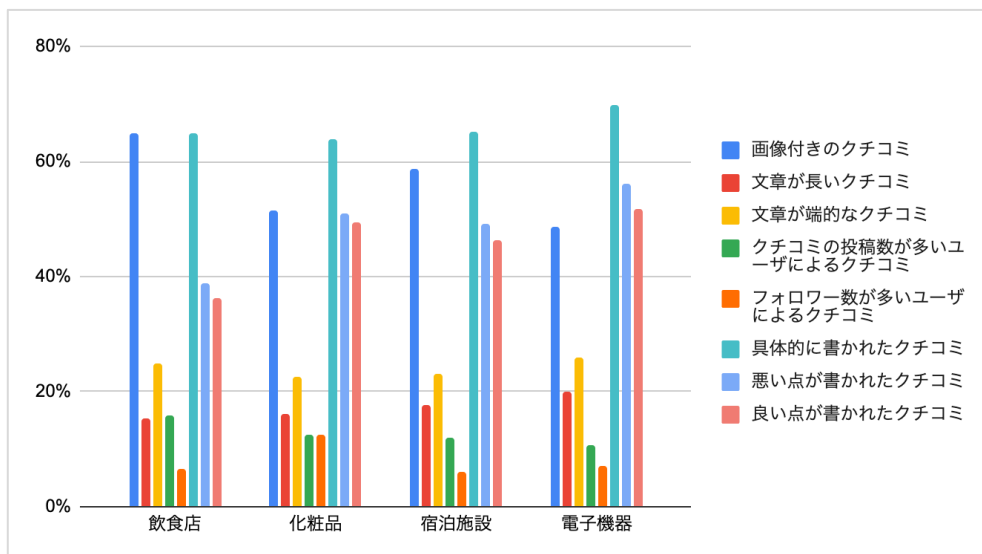


図 43 参考にするクチコミの特徴

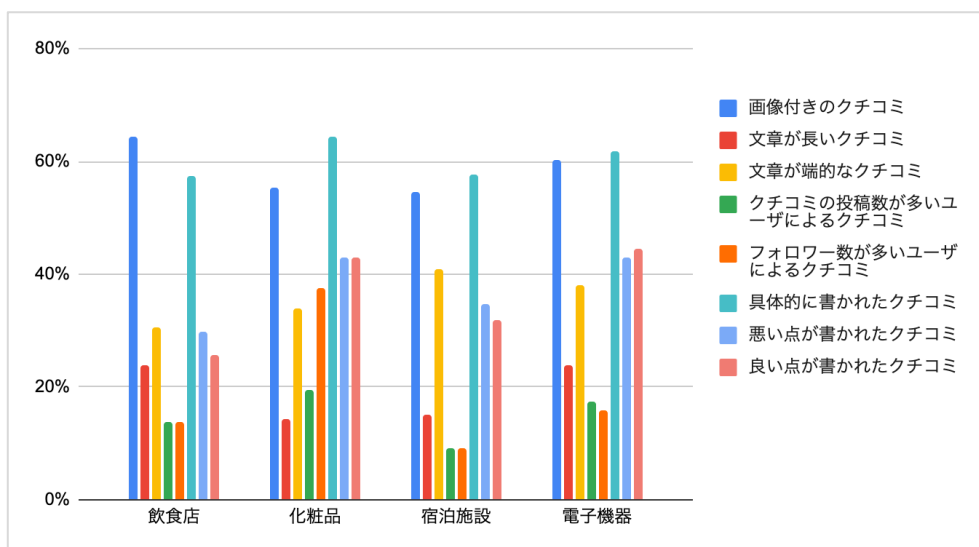


図 44 20~24 歳が参考にするクチコミの特徴

考にする人が増加したほか、化粧品のクチコミについてはフォロワー数が多いユーザーによるクチコミを参考にする人も増加した（図 44）。

8.2.3. 考察

予備調査と比較し、結果が異なる項目が複数存在した。まず化粧品のクチコミを週 1 度以上閲覧する回答者がクラウドソーシングの全世代分のデータでは減少しており、予備調査の回答者の世代とほぼ同じ 20~24 歳に絞ると週 1 度以上閲覧する割合が増加した。ここから、世代によって化粧品のクチコミへの関心の高さが異なり、特に若い世代は化粧品のクチコミを日常的に見る人が多いと考えられる。さらに、利用するサービスの結果から 20~24 歳の回答者は他の世代の回答者と比較し、SNS を多用することがわかる。すなわち、本研究で提案するシステムは、日常的に化粧品のクチコミを閲覧する人の多い 20 代前半がメインターゲットとなる可能性が高く、その世代が特によく利用する SNS でのクチコミに対応していく必要があることが示唆された。ここで、変化量の類似性を分析するため、以下の式でユークリッド距離を求めた。なお、数式内の A や B は飲食店、化粧品などのクチコミジャンルが当てはまる。

$$\text{ユークリッド距離} = \sqrt{\sum \left(\frac{A \text{ の } 20 \text{ 代前半の回答割合}}{A \text{ の全世代の回答割合}} - \frac{B \text{ の } 20 \text{ 代前半の回答割合}}{B \text{ の全世代の回答割合}} \right)^2}$$

閲覧頻度と参考にする特徴についてのユークリッド距離を表 23, 24 に示す。電子機器は化粧品とのユークリッド距離が小さく、傾向が類似していると考えられる。特に電子機器のクチコミについても若い世代で SNS のクチコミが多く利用されていたため、SNS に対応可能な信頼性評価システムを構築することで電子機器のクチコミにも応用可能性がある。ただ

表 23 閲覧頻度のユークリッド距離

頻度	飲食店	化粧品	宿泊施設	電子機器
飲食店		1.30	2.33	1.19
化粧品	1.30		2.04	1.01
宿泊施設	2.33	2.04		1.58
電子機器	1.19	1.01	1.58	

表 24 参考にする特徴のユークリッド距離

参考	飲食店	化粧品	宿泊施設	電子機器
飲食店		1.37	1.07	0.92
化粧品	1.37		1.73	0.82
宿泊施設	1.07	1.73		1.27
電子機器	0.92	0.82	1.27	

し、電子機器は全世代を通じて EC サイトのレビューが多く利用されており、他のジャンルに比較し SNS が多く利用されるのは化粧品のクチコミが独自にもつ特徴であると考えられる。

また、参考にするクチコミの特徴に関して、「画像付きのクチコミ」「具体的に書かれたクチコミ」が多く選択された。本研究が提案する信憑性評価軸は特に文章に着目しているが、画像に関しても評価軸の検討を行うことでより高精度の信憑性評価が可能になると考えられる。また、化粧品のクチコミでは内容の具体性について併用した製品や購入経緯、使用方法などの観点から評価を行っているが、これを他のジャンルにも置き換えることで様々なクチコミに応用できると考えられる。

そして、化粧品のクチコミを特に頻繁に閲覧する 20～24 歳においてフォロワー数が多いユーザによるクチコミを参考にすることが多いことが明らかになった。大きな影響力をもつインフルエンサによるクチコミは若い世代に参考にされることが多いが、投稿が話題になることを狙って誇張した表現を用いることや、ステルスマーケティングを行っている可能性など懸念点もある。SNS に対応可能にする場合、こうしたインフルエンサによるクチコミの扱いについても検討が必要である。

第9章 制約と展望

本章では、本研究の制約と今後の展望について述べる。

本研究ではクチコミの作成を大学生・大学院生が行ったが、より幅広い世代によるクチコミを収集することで、内容や文体の偏りをなくすことができると考える。幅広い世代からのクチコミを収集し、評価軸を用いた有用性の検証を行うことで、新たな評価軸の検討や精度向上に向けての議論を深められる可能性がある。

本システムは文章に特化した評価軸であるため、投稿者の情報や複数のクチコミの比較、画像を用いた信憑性評価を行っていない。しかし、実際にクチコミを閲覧するときは投稿者の肌質、年齢などの特性や、クチコミとともに投稿される画像、複数のクチコミとの比較などから複合的に信憑性を判断することが多い。そのため、投稿者や画像に関する評価軸の検討や同一商品に対する複数のクチコミからの比較機能を実装することで、より有用性を高めることができると考えられる。

また、化粧品のクチコミを閲覧する際に若い世代が多く利用する SNS に対応可能な信憑性評価の実現や、インフルエンサによる投稿の扱いに関する検討などを行うことでさらに実用性を高められることが期待される。

システムの機能に関して、事後アンケートで得られた意見をもとに信憑性ランクを用いたソート機能や絞り込み機能を実装することで、より使いやすいシステムにできると考えられる。また、実験協力者から「信憑性ランクについて、そのランクについて正しいと思うことも正しくないと思うこともあった」という意見も得られたため、信憑性ランクや評価軸などをユーザが納得しやすい提示方法にすることで安心して利用してもらえるシステムを目指す必要がある。また、信憑性スコアの算出について、現在は重み付けを行っていないが、重み付けを行うことで信憑性評価の精度の向上も見込める。

第10章 おわりに

本研究は、インターネット上の化粧品に関するクチコミの信憑性の判断が難しいという問題に対し、化粧品のクチコミに特化した信憑性評価システムを提案した。まずシステムの実現に向け、基礎調査とデータセット構築を行った。基礎調査では化粧品をする人がどのようなクチコミを信用するのかを明らかにした。また、データセット構築では300件のクチコミを収集し、どのような特徴をもつクチコミが信用されるのかをさらに具体的に明らかにするとともに、信憑性評価軸の検討を行った。次に、機械学習を用いて信憑性評価軸の有用性を検証し、81.7%の精度でクチコミ投稿者の化粧品使用状況を判別できることを確認した。この信憑性評価軸を用いたシステムを実装した後、システムを用いた実験を行い、実験前には信憑性をあまり気にしなかったユーザがシステム使用中に信憑性を意識することを示した。今後は、SNSのクチコミに対応可能な評価軸の検討のほか、画像や投稿者の情報、他の投稿との比較など複合的な信憑性評価を行うことでさらに信憑性判断の支援が可能であると考えられる。

本研究は化粧品のクチコミが抱える独自の問題に着目し、化粧品に特化した信憑性評価システムを構築した。しかし、ここまで検討した信憑性評価軸の内容を置き換えることで、他ジャンルのクチコミにも応用できる可能性がある。

化粧品は理想の自分を実現し、毎日を豊かにするものである。本研究によって、多くの人が化粧品を心から楽しみ、自分に自信をもって生活できることを強く望む。

参考文献

- [1] Langlois, H. J., Kalakanis, L., Rubenstein, J. A., Larson, A., Hallam, M. and Smoot, M.. Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*, 2000, vol. 126, no. 3, pp. 390-423.
- [2] 九島紀子, 齊藤勇. 化粧が対人印象に及ぼす影響 : 顔形態とメイクの差異による印象操作の実証的研究. *Japanese Journal of Applied Psychology*, 2015, vol. 41, no. 1, pp. 39-55.
- [3] “【シーン別】大人女子の TPO ぱっちりアイシャドウメイク♡”. <https://hair.cm/news/39624/>, (参照 2022-10-20)
- [4] “ウィズコロナ時代の美容行動「画面越しに見せたい顔」とは?”. <https://www.cosmetic-culture.po-holdings.co.jp/report/pdf/220623gamengoshi.pdf>, (参照 2022-10-20).
- [5] “化粧品購入行動に関する調査結果”. <https://research.nttcoms.com/database/data/002191/>, (参照 2022-10-20)
- [6] @cosme. <https://www.cosme.net/>, (参照 2022-10-20)
- [7] LIPS. <https://lipscosme.com/>, (参照 2022-10-20)
- [8] Twitter. <https://twitter.com/>, (参照 2022-10-20)
- [9] Instagram. <https://www.instagram.com/>, (参照 2022-12-09)
- [10] Mukherjee, A. Liu, B. and Glance, N.. Spotting fake reviewer groups in consumer reviews. *Proceedings of the 21st Annual Conference on World Wide Web*, 2012, pp. 191-200.
- [11] “サクラチェッカー”. <https://sakura-checker.jp/>, (参照 2022-10-20)
- [12] 加藤徹也, 青木滉一郎, 菅原徹, 村上智加, 宮崎正己. 目と眉の間隔の違いが顔の印象に及ぼす影響. *日本感性工学会論文誌*, 2015, vol. 14, no. 3, pp. 419-424.
- [13] 高木佐恵子, 波川千晶, 吉本富士市. メイクアップ技術上達のためのアドバイスシステム. *芸術科学会論文誌*, 2003, vol. 2, no. 4, pp. 156-164.
- [14] 神武里奈, 星野准一. 好みの顔画像の色に基づくメイクアップ支援システム. *日本感性工学会論文誌*, 2017, vol. 16, no. 3, pp. 299-306.
- [15] Chong, T., Maudet, N., Harima, K. and Igarashi, T.. Exploring a Makeup Support System for Transgender Passing based on Automatic Gender Recognition. *CHI'21*, 2021.
- [16] Kajita, M. and Nakamura, S.. Basic Research on How to Apply Foundation Makeup Evenly on Your Own Face. *20th IFIP TC14 International Conference on Entertainment Computing (IFIP ICEC 2021)*, 2021, pp. 402-410.
- [17] Jiyong, Y.. A Consumer Recommendation System based on Big Data. *International Journal of Smart Business and Technology*, 2019, vol. 7, pp. 25-30.
- [18] Trusov, M., Bucklin, E. R. and Pauwels, K.. Effects of Word-of-Mouth versus Traditional Marketing: Findings from an Internet Social Networking Site. *Journal of Marketing*, 2009, vol. 73, Issue 5, pp.

90-102.

- [19] Cheung, Y. M., Luo, C., Sia, L. C. and Chen, H.. Credibility of Electronic Word-of-Mouth: Informational and Normative Determinants of On-line Consumer Recommendations. *International Journal of Electronic Commerce*, 2009, vol. 13, Issue 4, pp. 9-38.
- [20] You, Y., Vadakkepatt, G. G. and Joshi, M. A.. A meta-analysis of electronic word-of-mouth elasticity. *Journal of Marketing*, 2015, vol. 79, Issue 2, pp. 19-39.
- [21] Heejae, S. and Dahana, D. W.. The Moderating Roles of Prior Attitude and Message Acceptance in Electronic Word of Mouth. *International Journal of Business and Information*, 2017, vol. 12. no. 2, pp. 183-207.
- [22] Brandes, L. and Dover, Y.. Offline Context Affects Online Reviews: The Effect of Post-Consumption Weather. *Journal of Consumer Research*, 2022, vol. 49, Issue 4, pp. 595-615.
- [23] Sokolova, K. and Kefi, H.. Instagram and YouTube bloggers promote it, why should I buy? How credibility and parasocial interaction influence purchase intentions. *Journal of Retailing and Consumer Services*, 2019, vol. 53. 101742.
- [24] Castillo, R. A., Jaramillo, C. Z. and Sy, L.. The Effectiveness of Social Media Influencers in the Cosmetic and Skincare Industry to the Purchase Intention of the Generation Z Filipinos. *Journal of Business and Management Studies*, 2022, vol. 4, no. 2, pp. 180-191.
- [25] Lee, S. and Kim, E.. Influencer marketing on Instagram: How sponsorship disclosure, influencer credibility, and brand credibility impact the effectiveness of Instagram promotional post. *Journal of Global Fashion Marketing*, 2020, vol. 11, no. 3, pp. 232-249.
- [26] Oh, Y.. Determinants of Online Review Helpfulness for Korean Skincare Products in Online Retailing. *Journal of Distribution Science*, 2020, vol. 18, no. 10, pp. 65-75.
- [27] Rofianto, W., Kornelys, D., and Rifkhansyah, M.. Visual eWOM Usefulness and Credibility: The Antecedents and Consequences. *Jurnal Ilmu Manajemen & Ekonomika*, 2017, vol. 9, no. 2, pp. 103-108.
- [28] Min, X., Rang, W. and Sylvia, Chan-Olmsted.. Factors affecting YouTube influencer marketing credibility: a heuristic-systematic model. *Journal of Media Business Studies*, 2018, vol. 15, no. 4, pp. 1-26.
- [29] Laer, T. van, Escalas, E. J., Ludwig, S., and Hende, E. A. van den. What Happens in Vegas Stays on TripAdvisor? A Theory and Technique to Understand Narrativity in Consumer Reviews. *Journal of Consumer Research*, 2019, vol. 46, Issue 2, pp. 267-285.
- [30] Sumita, U. and Kim, A. H.. Analysis of e-Word-of-Mouth Information for Cosmetic Products in Japan. *International Journal of Electronic Commerce Studies*, 2012, vol. 3, no. 1, pp. 61-82.
- [31] Pugsee, P., Sombatsri, P. and Juntiwakul, R.. Satisfactory Analysis for Cosmetic Product Review Comments. *International Conference on Data Mining, Communications and Information Technology (DMCIT '17)*, 2017, Article 13, pp. 1-6.

- [32] Matsunami, Y., Ueda, M., Nakajima, S., Hashikami, T., O'Donovan, J. and Kang, B.. Mining Attribute-Specific Ratings from Reviews of Cosmetic Products. *International MultiConference of Engineers and Computer Scientists*, 2017, pp. 101-114.
- [33] Matsunami, Y., Ueda, M. and Nakajima, S.. Tag recommendation method for a cosmetics review recommender system. *the 19th International Conference on Information Integration and Web-based Applications & Services (iiWAS '17)*, 2017 pp. 166-170.
- [34] Thinnukool, O., Charoenkwan, P., Khuwuthyakorn, P. and Tinamat, P.. Word Cloud Analysis of Customer Satisfaction in Cosmetic Products in Thailand. *the 2020 the 4th International Conference on Compute and Data Analysis (ICCD A 2020)*, 2020, pp. 179-182.
- [35] Wathen, C. N. and Burkell, J.. Believe it or not: Factors influencing credibility on the Web. *Journal of the American Society for Information Science and Technology*, 2002, vol. 53, no. 2, pp. 134-144.
- [36] 藤本雄紀, 今井龍一, 中村健二, 田中成典, 有馬伸広, 荒川貴之. 信憑性の高い記事を投稿するマイクロブログユーザの特定手法に関する研究. *土木学会論文集*, 2017, vol. 73, no. 2, pp. I_259-I_267.
- [37] Johnson, J. T. and Kaye, K. B.. Credibility of Social Network Sites for Political Information Among Politically Interested Internet Users. *Journal of Computer-Mediated Communication*, 2014, vol. 19, Issue 4, pp. 957-974.
- [38] Moran, G. and Muzellec, L.. eWOM credibility on social networking sites: A framework. *Journal of Marketing Communications*, 2014, vol.23, Issue 2, pp.149-161.
- [39] 池田圭佑, 榎剛史, 鳥海不二夫, 風間一洋, 野田五十樹, 諏訪博彦, 篠田孝祐, 栗原聡. マルチエージェント型情報拡散モデルの提案. *人工知能学会論文誌*, 2016, vol. 31, no. 1, pp. NFC-C_1-13.
- [40] 中島伸介, 稲垣陽一, 草野奉章. 高信頼性情報の提示を目指した熟知度に基づくプログラムキング方式の提案. *日本データベース学会論文誌*, 2008, vol. 7, no. 1, pp. 1-6.
- [41] Huang, H. Wen, Y. and Chen, H.. Detection of False Online Advertisements with DCNN. *Proceedings of the 26th International Conference on World Wide Web (WWW'17)*, 2017, pp. 795-796.
- [42] Nakamura, S., Konishi, S., Jatowt, A., Ohshima, H., Kondo, H., Tezuka, T., Oyama, S. and Tanaka, K.. Trustworthiness Analysis of Web Search Results. *the 11th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2007)*, 2007, pp. 38-49.
- [43] 山本祐輔, 田中克己. データ対問のサポート関係分析に基づく Web 情報の信憑性評価. *情報処理学会論文誌*, 2010, vol. 3, no. 2, pp. 61-79.
- [44] 山本祐輔, 山本岳洋, 大島裕明, 川上浩司. ウェブアクセスリテラシー尺度の開発. *情報処理学会論文誌*, 2019, vol. 12, no. 1, pp. 24-37.
- [45] Iwamoto, E., Matsubara, M., Ota, C., Nakamura, S., Terada, T., Kitagawa, H. and Morishima, A.. Passerby Crowdsourcing: Workers' Behavior and Data Quality Management. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2018, vol. 2, no. 4, pp. 1-20.

- [46] 三輪聡哉, 中村聡史. マイクロタスク埋め込み型音楽ゲームの提案. 情報処理学会 研究報告エンタテインメントコンピューティング (EC) , 2014, vol. 2014-EC-34, Issue 2, pp. 1-6.
- [47] 佐々木孝輔, 平田章, 井上智雄. マイクロタスクによる線画イラスト生成手法. 情報処理学会論文誌, 2016, vol. 57, no. 1, pp. 260-269.
- [48] 三原鉄也, 石川夏樹, 豊田将平, 永森光晴, 杉本重雄. 画像認識とマイクロタスク型クラウドソーシングを組み合わせたマンガのコマ領域の判定. 人工知能学会全国大会論文集, 2018, JSAI2018 巻, 第 32 回, pp. 4M105.
- [49] Yahoo!クラウドソーシング. <https://crowdsourcing.yahoo.co.jp/>, (参照 2022-10-20)
- [50] 樋口耕一. テキスト型データの計量的分析 —2つのアプローチの峻別と統合—. 理論と方法, 2004, vol. 19, no. 1, pp. 101-115
- [51] Amazon. <https://www.amazon.co.jp/>, (参照 2022-12-16)

研究業績

- [1] 伊藤理紗, 濱野花莉, 野中滉介, 菅野一平, 中村聡史, 掛晃幸, 石丸築. 手書きとフォントの文字形状の違いによる記憶効果の比較. 電子情報通信学会 ヒューマンコミュニケーション基礎研究会 (HCS) , 2019, HCS-24, vol. 119, no. 38, pp. 119-124.
- [2] 濱野花莉, 細谷美月, 佐々木美香子, 山浦祐明, 中村聡史. フォントと味の印象が商品選択行動に及ぼす影響. ヒューマンインターフェースシンポジウム 2019, 2019, No. 5A-2.
- [3] Ito, R., Hamano, K., Nonaka, K., Sugano, I., Nakamura, S., Kake, A. and Ishimaru, K.. Comparison of the Remembering Ability by the Difference Between Handwriting and Typeface. International Conference on Human-Computer Interaction (HCII 2020), 2020, CCIS, vol. 1224, pp. 526-534.
- [4] 山崎郁未, 澤佳達, 伊藤理紗, 濱野花莉, 中村聡史, 掛晃幸, 石丸築. 文字の見た目が記憶に及ぼす影響. 情報処理学会 研究報告ヒューマンコンピュータインタラクション (HCI) , 2020, vol. 2020-HCI-189, no. 16, pp. 1-7.
- [5] 山崎郁未, 伊藤理紗, 濱野花莉, 中村聡史, 掛晃幸, 石丸築. 記憶対象の文字の太さの違いが記憶容易性に及ぼす影響. 情報処理学会 研究報告ヒューマンコンピュータインタラクション (HCI) , 2020, vol. 2020-HCI-190, no. 22, pp. 1-8.
- [6] 濱野花莉, 中村聡史. 化粧品に対するクチコミの信憑性判定に向けたクチコミ文章の調査. 第13回データ工学と情報マネジメントに関するフォーラム (DEIM2021) , 2021, no. C14-3, pp. 1-8.
- [7] 高野沙也香, 山崎郁未, 伊藤理紗, 濱野花莉, 菅野一平, 中村聡史, 掛晃幸, 石丸築. 筆跡の自筆との類似性が記憶容易性に及ぼす影響の検証. 情報処理学会 研究報告ヒューマンコンピュータインタラクション (HCI) , 2022, vol. 2022-HCI-196, no. 2, pp. 1-8.
- [8] 濱野花莉, 伊藤理紗, 中村聡史. 文章に着目した化粧品クチコミの信憑性評価軸の検討. 情報処理学会 研究報告グループウェアとネットワークサービス (GN) , 2022, vol. 2022-GN-115, no. 32, pp. 1-8.
- [9] 濱野花莉, 梶田美帆, 中村聡史. 化粧品クチコミに特化した信憑性評価システムの実現とその評価. 情報処理学会 研究報告ヒューマンコンピュータインタラクション (HCI) , 2022, vol. 2022-HCI-199, no. 30, pp. 1-8.
- [10] 高野沙也香, 梶田美帆, 濱野花莉, 中村聡史. Make-up FLOW: 個人差・状況差の大きい化粧工程の構造化と忘れやすさに関する調査. 情報処理学会 研究報告ヒューマンコンピュータインタラクション (HCI) , 2022, vol. 2022-HCI-200, no. 18, pp. 1-8.