

文章に着目した化粧品クチコミの信憑性評価軸の検討

濱野花莉¹ 伊藤理紗¹ 中村聡史¹

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、化粧品は多くの消費者の需要に応じて多種多様な商品が開発されている。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実際の利用を確認できないなど信憑性において様々な問題を抱えている。そこで本研究では化粧品に特化したクチコミの信憑性判断を支援するシステムの実現を目指し、これまでに検討した信憑性の評価軸について有用性の検証を行った結果、80.2%の精度で信憑性評価が可能であった。また、新たな軸の検討や、化粧品の種類別に分析を行った。その結果、化粧品の種類により効果的な評価軸が異なる可能性が示唆された。

キーワード: 化粧品, クチコミ, コスメ, レビュー, レビュー分析, 信憑性

1. はじめに

ポーラ文化研究所が2019年に行った女性を対象とする調査[1]によると、59.2%の人が日常的に化粧を行うと回答しており、ときどき行うと回答した人も含めると78.9%の人が化粧を行っている。また、特に25~29歳では、87.3%の人が化粧を行うと回答していた。消費者の需要に応じて各化粧品メーカーは多種多様な商品を研究開発しており、ファンデーションに絞っても、シミや毛穴を隠すカバー力を重視したもの、時間が経過しても崩れにくいことを重視したもの、自然由来や肌に優しい成分にこだわったものなど、消費者の使用目的に応じた様々な商品が存在する。また、男性用の化粧品の発売する化粧品メーカーも増えており、化粧品市場はますます拡大している。

化粧品は継続的に直接肌に塗るものであるため、自身の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、化粧品の購入に失敗したくないと考える消費者は多い。しかし、株式会社M2が2019年に行った調査[2]によると、「ファンデーション選びで失敗したことはありますか?」という質問に対して、65.9%の人が「はい」と回答しているなど、化粧品の選択は容易ではないことがわかる。

ここで化粧品に関する情報を得る方法として、家族や友人、化粧品専門店のスタッフであるビューティアドバイザーの話や、化粧品メーカーの公式サイトに記載されている情報、クチコミサイトやSNS上でのクチコミを閲覧することが挙げられる。特にクチコミサイトやSNSでは、手軽に多くの消費者の意見を閲覧できることから、情報収集の手段として利用する人は多い。しかし、誰でも書き込むことができるため、質の低いクチコミやサクラによるクチコミが多く存在することが問題点として挙げられる。質の低いクチコミの具体例としては、専門知識のないユーザによる参考にならない意見や正しい知識に基づかない意見、

SNSで拡散されたいと考えるユーザによる誇張表現などが挙げられる。また、サクラによるクチコミの具体例としては、企業がユーザに依頼して、広告であることを明記せずに良いクチコミを書かせるステルスマーケティングなどが挙げられる。これらの背景から、化粧品の購入に失敗しないために、クチコミサイトやSNS上に書き込まれているクチコミの信憑性をユーザ自身が見極める必要がある。

こうしたクチコミの問題に着目し、クチコミの信憑性についての研究[3]やECサイトのクチコミの信憑性を確かめるためのシステム開発[4]も行われており、様々な業界でクチコミが重要視されていることがわかる。ここで化粧品のクチコミは、飲食店や家電製品といった他のカテゴリのクチコミとは異なる特性をもっている。まず、化粧品はユーザの肌質や好みの差が大きいため、ユーザによってクチコミの意見が分かれることがある。また、こうしたクチコミは飲食店や家電製品のクチコミとは異なり、化粧品のクチコミを投稿した人が実際にその商品を使っているのか、どの程度使い込んでいるのかなど、クチコミの内容だけでは判断できないこともある。商品使用時の肌の写真を載せた投稿であっても、その写真のみからその商品を利用しているか判断することは不可能であるうえ、写真の撮り方やフィルタ等の加工により印象が大きく異なってしまう。また、化粧下地の上にファンデーションを塗る、アイシャドウベースの上にアイシャドウを塗るなど、一つの部位に対して複数の化粧品の併せて使用することもあるため、クチコミの内容や写真がその化粧品のみによる効果なのか、併せて使用した化粧品による効果なのかを判断することは容易ではない。さらに、化粧水や美容液といった基礎化粧品に関しては、ある一定の期間使用し続けることで効果が現れるものが多いため、短期間しか使用していないユーザのクチコミは参考にならない可能性が高い。そのうえ、化粧品は試供品が配布されていることも多く、その試供品を一度利用しただけでクチコミを書いていることも珍しくない。

¹ 明治大学
Meiji University

このように、化粧品のクチコミには特有の特性があるため、信憑性の判断を支援するシステムが必要である。しかしクチコミサイトや SNS 上のクチコミが真実かどうかを完璧に判断することは現実的ではない。こうした問題をふまえて、我々は、システムによる客観的な信憑性評価に加え、信憑性評価が得意なユーザの評価基準を共有することで、信憑性の評価が苦手なユーザを支援するシステムの実現を目指し、化粧品のクチコミの信憑性評価軸の検討を行ってきた[5]。しかし、この評価軸を用いてどこまで信憑性を正しく判断できるのかについては明らかにできていなかった。

そこで本稿では、検討した信憑性評価軸における有用性を検証する。まず、検証に使用するための化粧品のクチコミデータセットを使用状況別に作成する。ここで、使用状況は、「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）」の3種類とした。また、それぞれのクチコミについて、評価軸をもとにデータセットの構築を行った。このデータセットをもとに評価軸の検証を行い、80.2%の精度でクチコミの信憑性を評価できることを明らかにした。さらに、新しい評価軸についての検討や化粧品の種類別の評価も行い、化粧品の種類によって軸の向き不向きがあることが示唆された。

2. 関連研究

2.1 化粧に関する研究

化粧は日常生活との関わりが深いことから、これまで多くの化粧支援に関する研究が行われている。

高木ら[6]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[7]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。

このように化粧の支援に関する研究がこれまで多く行われてきた。本研究では化粧品使用時の支援ではなく、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目的としている。

2.2 情報の信憑性に関する研究

また、インターネットの普及により、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。

藤本ら[8]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 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 Moran ら[9]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。中島ら[10]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[11]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。山本ら[12][13]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。

本研究では、このようなウェブ情報の中でも、化粧品のクチコミに特化して信憑性の判定をしていくものであるが、こうした手法と組み合わせることでより良くすることができると考えている。

2.3 クチコミに関する研究

一方、クチコミはマーケティング戦略において重要視されている。

Lee ら[14]は、Instagram におけるインフルエンサー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、インフルエンサー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 ら[15]はネット通販のレビューについて、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。さらに、Mukherjee ら[3]は、Amazon のレビューグループから偽のレビューグルー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、偽のレビューグループを判別するための指標や偽のレビューグルー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

一を検出するより容易であることも明らかにしている。

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。

2.4 マイクロタスクに関する研究

また、我々が目指すシステムではマイクロタスク的にクチコミ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[16]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。佐々木ら[17]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[18]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。また、三輪ら[19]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。

我々は、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沿った質問を提示し、データを集約・蓄積することで信憑性評価を行うことを検討している。

3. 化粧品クチコミの信憑性評価軸

我々は、化粧品のクチコミに特化した信憑性評価システムの実現を目指している。我々が想定しているクチコミの信憑性評価システムは、比較がなされているか、悪い点が書かれているか、長期的に利用さ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。ここでクチコミの各種評価軸について、完全に機械により自動評価できることが望ましいが、人が投稿するクチコミは口語的で不完全な文章も多く、容易ではない。また、機械による自動評価のための学習のためには膨大な情報が必要である。そこで、クチコミに対する軸評価をマイクロタスク化し、膨大なユーザの手でアノテーション付与したうえで、データを集約・蓄積し、信憑性評価を行う予定である（図

1）。また、将来的に構築されたアノテーションデータセットをもとに自動推定を目指す。

我々はこれまでの研究[5]において、化粧品のクチコミに特化した信憑性評価システムの実現にむけ、特に文章に着目した信憑性評価軸を作成した。評価軸を作成するために、化粧品のクチコミの文章について信憑性データセットを構築し、分析を行った。まず、このデータセットを構築するにあたり、ファンデーション・化粧水・リップのクチコミをインターネット上から300件収集した。さらに、各クチコミについて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項目について、記述の有無を付与した。その後、各クチコミの信憑性評価を女子大学生9名に依頼した。

このデータセットを分析し、信憑性評価軸の検討を行った結果、信憑性評価軸として、「商品の悪い点が含まれるクチコミ」「PRであっても率直な意見や他の製品との比較が含まれるクチコミ」「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」「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」が挙げられた。

これらの評価軸はあくまでユーザがどのような観点でクチコミを信用するかという基準が主となっている。そのため、この評価軸について検証を行い、ユーザがどの程度正しくクチコミを評価できているのかを明らかにする必要がある。また、ユーザが正しく評価できていない部分に対応できる評価軸を検討することで、ユーザが信憑性を評価する際の支援が可能になると考えている。

4. 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。有用性検証にあたり、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

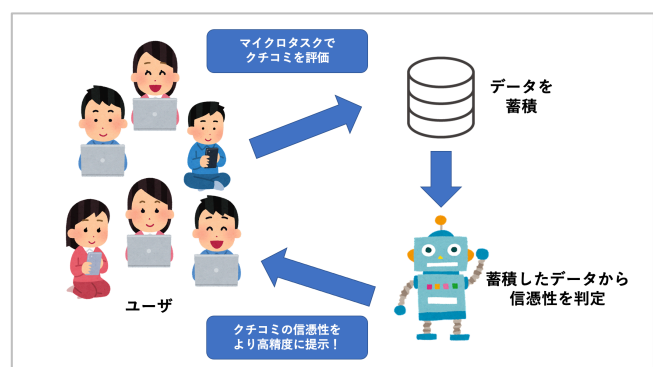


図1：システムでの評価軸利用イメージ図

セットの構築が必要となる。これまでの研究で用い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。そこで本稿では、使用状況別のクチコミを作成した。

本稿ではこれまでの研究と同様、ファンデーション・化粧品水・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[20]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、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、男性にもクチコミの作成を依頼したのは、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男性が投稿している可能性があるためである。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1に示す。

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

表1：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	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。
	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。
一度だけ使用	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。
	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ありません。毛穴が消えてびっくりしました。
未使用	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。
	少し前に友人から貰ってからずっと愛用しています！私を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

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め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. クチコミに対する軸評価データセット構築

5.1 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。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2に示す。本システムでは、クチコミが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3章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あ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、3章で作成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2に示す。ここに示した単語は、3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。

軸評価データセットの構築は18~25歳の17名(男性8名、女性9名)に依頼した。

5.2 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人について、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(mean±1.5SD)となる

図2：軸評価データセット構築システムの画面

表2：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

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3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。

また、各軸における使用状況別のスコア平均を表4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4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ず、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4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ら[20]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。

5.3 評価軸の有用性の検証

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5のように求められた。表5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

表3：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5：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4：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

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、大きな変化はなかった。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 6 にまとめる。表 6 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。

6. 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な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[3]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る。

我々は、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

表 6：化粧品の種類別の精度と各評価軸の重要度

評価軸	ファンデーション	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

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。

7. まとめ

本研究は、化粧品のクチコミがもつ独自の性質を踏まえたクチコミの信憑性評価基準の提案および信憑性評価システムの開発を目指し、クチコミを使用状況別に作成し、これまでに検討した信憑性評価軸の有用性を検証した。検証の結果、精度は 80.2%となり、もっとも重要度の高い評価軸は「クチコミ投稿者なりの使用方法が書かれている」であった。さらに、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から、新しい評価軸について検討を行った。また、化粧品の種類別に分析を行うことで、それぞれの化粧品に適した評価軸が存在することが示唆され、ファンデーションのみに絞ると精度が 87.7%まで上昇することが明らかとなった。

今後は、他の化粧品に対しても有効な評価軸を探索する。ここで、本稿ではクチコミ作成を大学生、大学院生に依頼したが、今後は幅広い世代にクチコミを作成してもらうことで、実際のクチコミサイトに近い状態においても評価軸が有用であるか検証を行う。これまでは文章のみに着目し、評価軸を検討・検証していたが、今後は投稿者や画像といった要素にも着目し、評価軸の追加を行う予定である。さらに、これらの評価軸を用いた、化粧品に特化したクチコミの信憑性評価システムの実現を目指す。

参考文献

- [1] “化粧への期待と行動”. <https://www.cosmetic-culture.po-holdings.co.jp/report/pdf/191212kitai.pdf>, (参照 2021-12-14)
- [2] “メイク崩れ注意報発令！7割近くの女性が「ファンデーション選びで失敗したことがある」と回答。気になる失敗エピソードとは？”. <https://www.cocondesoie.com/post/make-up-brakeage-warning-issued>, (参照 2021-12-14)
- [3] Mukherjee, A. Liu, B. and Glance, N.. Spotting fake reviewer groups in consumer reviews. Proceedings of the 21st Annual Conference on World Wide Web (WWW'12). 2012, pp.191-200.
- [4] “サクラチェッカー”. <https://sakura-checker.jp/>, (参照 2021-12-14)
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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. 土木学会論文誌. 2017, vol.73, no.2, p.p.1_259-1.267.
- [9] 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.
- [10] 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ブログランキング方式の提案. 日本データベース学会論文誌. 2008, vol.7, no.1, pp.1-6.
- [11] 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.
- [12] 山本祐輔, 田中克己. データ対間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.
- [13] 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.
- [14] 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, p.p.232-249.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] 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, p.4M105.
- [19] 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.
- [20] Miho Kajita, Satoshi Nakamura. 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, 2021.