

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

濱野 花莉[†] 中村 聡史[†]

[†] 明治大学総合数理学部 〒164-8525 東京都中野区中野 4-21-1

E-mail: [†] hamano1meiji@gmail.com

あらまし 化粧品を購入する際の情報収集の手段として、インターネット上のクチコミを利用する人は多い。一方で、いいかげんな評価や誇張表現などを含むクチコミが存在するうえ、ステルスマーケティングが行われている可能性があることから、情報の信憑性に問題がある。ここで化粧品のクチコミは、実際の使用状況が判断できないことが多く、また期待する効果によってユーザの評価が異なるなど、信憑性判断が難しい。これまでも情報の信憑性に関する研究は行われてきたが、化粧品に特化したものはない。そこで本研究では、化粧品の特性を考慮した信憑性評価システムの実現を目指し、クチコミの文章のみに着目した調査を行った。その結果、同じクチコミを提示しても人によって信憑性の評価は異なることや、いくつかの評価基準があることが明らかになった。また、信憑性の評価基準についても検討した。

キーワード コスメ, レビュー, レビュー分析, クチコミ, 信憑性

1. はじめに

多くの女性が理想の自分の実現や、コンプレックスを隠す目的で化粧を日常的に施している。実際に、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]では、59.2%の人は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ことが明らかになった。特に 25~29 歳では、87.3%の人が化粧を行うと回答している。また、消費者の需要に応じて様々な化粧品メーカーは多種多様な商品を研究開発している。ファンデーションのみに着目しても、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。さらに近年、多くの化粧品メーカーが男性用の化粧品を発売しており、化粧品市場はますます拡大している。

一方で、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場合があるうえ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ユーザは多い。しかし、化粧品の選択は容易ではなく、株式会社 M2 が 2019 年に行った調査[2]では、「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。このような化粧品の選択における失敗を回避するために、購入前に製品に関する情報収集を行うことが重要である。

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やインターネット上のクチコミサイト、SNS 上でのクチコミを閲覧することが挙

げられる。特にインターネット上のクチコミは、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、近年で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も行われている。こうした理由から、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、クチコミが適切なものであったとしても、クチコミ通りの効果が得られるとは限らない。化粧品はこうした個人の持つ特性や、化粧品に期待する効果の違いが原因となり、ユーザによってクチコミの意見が分かれることがしばしばある。こうしたクチコミの特性は個人的なものであり、飲食店や家電製品のレビューとも似たものであるが、さらなる難しさは、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にある。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別することは不可能であるうえ、写真の撮り方やフィルタにより印象が大きく異なる。さらに、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を併せて使用することもある。そのため、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。また、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。ここで、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。

つまり、インターネット上の化粧品に対するクチコミの信憑性をユーザ自身が見極めるためには、クチコミの内容をよく読み、発信者や情報のソース、他のクチコミと一貫性があるかなど多くの情報を確認する必要がある。しかし、これらの情報を確認するためには多くの時間や手間がかかってしまい、手軽に閲覧できるというクチコミの利点が損なわれてしまう。そこで、こうした化粧品におけるクチコミの特性を踏まえ、信憑性の判断を支援するシステムが必要である。本研究ではインターネット上のクチコミが真実かどうかを完璧に判断することは不可能であると考えた。そこで、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す。

そこで本稿では、化粧をする人が化粧に対してどのような意識をもっているのか、どのように化粧品の情報を収集しているのか、どういった点でクチコミを信用するのかといった基本的な情報を明らかにするために、クラウドソーシングを用いて化粧品のクチコミや化粧自体に対する調査を行う。また、調査結果を踏まえて、化粧品に対するクチコミを収集し、そのクチコミの文章が、どの程度信頼できるか、商品を買いたいと思うか、使いたいと思うかといった複数の項目について回答してもらうことにより、化粧品のクチコミに関する信憑性データセットを構築する。また、調査結果を踏まえ、信憑性評価軸の候補について議論を行う。

2. 関連研究

化粧は日常生活との関わりが深いことから、これまでに多くの化粧支援に関する研究が行われている。高木ら[5]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[6]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、さらに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。

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[7]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。またMoranら[8]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を提案した。中島ら[9]は、ブロガーのマニア度を算出するマニア指向プログラミ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[10]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た。山本ら[11][12]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。

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。Mukherjeeら[3]は、Amazonのレビューアグループから偽のレビューア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、偽のレビューアグループを判別するための指標や偽のレビューア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アを検出することは、個々のレビューに注目して偽のレビューを検出するより容易であることも明らかにし

た。しかし、この研究は化粧品のカ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。

3. 化粧をする人を対象とした調査

3.1. 調査概要

本稿ではクチコミの信憑性について分析するにあたり、まず化粧をする人に対して事前調査を行った。

本調査は、「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を用いて実施した。アンケートでは性別や年齢、化粧の頻度や化粧が好きかどうか、化粧品についてのクチコミの利用状況やどういった点でクチコミを信用するかなどを質問した。なお、化粧の頻度に関してはCOVID-19の流行前の状況での頻度を回答するよう指示した。

3.2. 結果と分析

Yahoo!クラウドソーシングは不真面目な回答者が多く含まれるため、そうした回答者を除外する必要がある。この調査は、化粧をする人を対象としているため、化粧の頻度に関する質問で「ほとんどしない（週1日未満）」と回答した回答者426名と、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名の回答を不適切回答と判断し、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。

まず、回答者の多くは30～50代であり、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。

化粧品の情報を得る手段と、化粧品を買うときに参考にする情報についての結果を図1に示す。図1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する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情報を得てはいるが、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、新たに情報を得ていると考えられる。

ここで、化粧が好きかどうかと化粧品購入時に参考にする情報の回答について、クロス集計を行った結果を表1に示す。表中1列目の括弧内の数字は、化粧が好きかどうかの各項目を回答した人数を表す。表1で背景が灰色になっているセルは、化粧に対する好き嫌い度合い別に、化粧品を買うときに参考にする情報として回答者が選択した上位2つを表している。表1より、化粧が「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。

化粧品についてのクチコミで、信用できるものの特徴とできないものの特徴の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った。その結果を図2、3に示す。なお、描画はKH Coder[13]を用いて行っている。共起ネットワークでは、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の強さを表している。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2、3に抽出された特徴語を見ると、「良い」「悪い」「詳細」「説明」「使用感」といったクチコミ本文の内容に関するもの、「画像」「加工」といった画像に関するもの、「インフルエンサー」「近い」「世代」といったクチコミ投稿者に関するものが多いことがわかる。つまり、信憑性を判断する要因は文章、画像、クチコミ

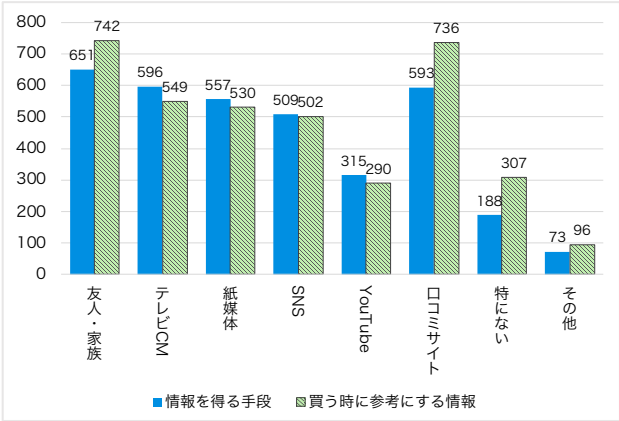


図1 化粧品の情報を得る手段

表1 化粧の好き嫌いと購入時参考にする情報

	友人・家族	テレビCM	紙媒体	SNS	YouTube	クチコミサイト	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

投稿者の大きく3つに分類できると考えられる。また、ユーザはこの3つの要因に含まれる様々な特徴を考慮して化粧品のク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。特に、信頼できないクチコミの特徴に関する回答の中で文章に着目したものに注目すると、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、人によって真逆の特徴をもったクチコミを信頼できると考えるということがわかるため、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た。そこで、この3つの要因について細かく調査するため、以降では、回答内で言及されることの多かった文章に着目する。

4. 化粧品のクチコミの文章に着目した分析

事前調査では、化粧品に対するクチコミについて、ユーザが文章、画像、クチコミ投稿者の3つの側面から信憑性を判断していることが明らかになった。また、ユーザによって信用できると感じる基準は異なっていたため、今回はまず、回答で特に多く言及されていた文章に着目しデータセット構築と分析を行う。

なお、この分析から得られるのはユーザ自身が信憑性を評価する軸である。化粧品のクチコミの信憑性を評価する際、このユーザによる評価軸のみでは正しく判断しきれない部分について、システムを用いて信憑性評価を補助することを目指す。本調査はシステムが支援すべき箇所を検討する前段階として行う。

4.1. 化粧品の選定とクチコミの収集

事前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。

データセット作成にあたり、まずクチコミを収集する化粧品を選定した。今回はファンデーション、化粧水、リップの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、他の化粧品と比較すると、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、価格が偏らないよう考慮し、@cosme内でランキングが上位でありクチコミ件数が多いものという観点から決定した。

各商品に対するクチコミを20件ずつ、計300件を@cosme、LIPS、Twitterから収集した。@cosmeやLIPSは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でありユーザ数も多いため、これらのサイトを選定した。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、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の記述の有無を付与した。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

収集したクチコミについて、人がどのように信憑性を判断するのか、また人による評価にブレがあるのか

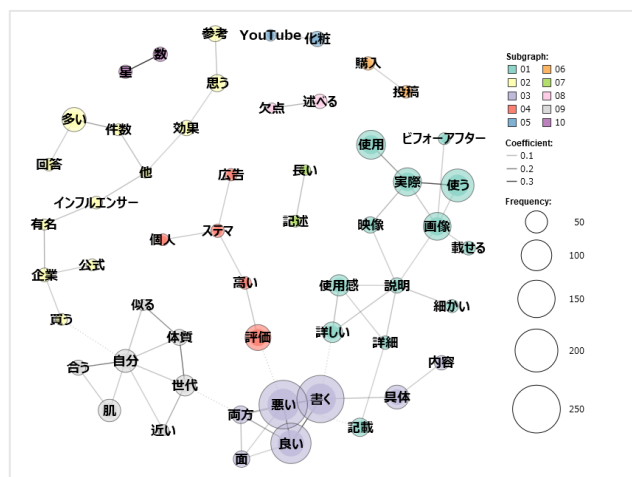


図2 信頼できるクチコミについての共起ネットワーク

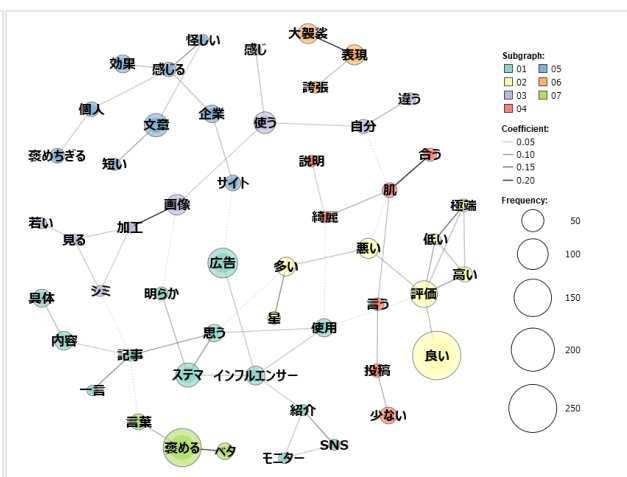


図3 信頼できないクチコミについての共起ネットワーク

を明らかにするため、各クチコミに対する信憑性の評価を行ってもらい、本調査における各クチコミへの回答項目を以下に示す。

- Q1. この口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

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。

なお、アノテーション付与システムはVue.jsを用いたWebアプリケーションを実験システムとして実装した。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、実験協力者自身が好きなタイミングで回答を中断できるよう設計した。本調査は女子大学生13名に依頼した。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

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。ここではこの9名のデータを用いて分析を行う。

まず、300件すべてに回答した9名（実験協力者A～I）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2に示す。ここではcos類似度が0.50未満の組み合わせの数値を太字で示している。この結果より、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ことがわかる。これは3章の調査と同様の結果である。ここで、実験協力者A,G,Iはcos類似度が特に低いことがわかる。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であったが、cos類似度が低かったA, G, Iの信憑性評価の平均はそれぞれ0.61, 0.47, 0.54であった。このことから、実験協力者A, G, Iは

クチコミの信憑性を低く見積もる傾向があると考えられる。また、分散は1.53, 1.19, 1.24であったことから、-2～+2の評価の幅を目一杯用いて評価を行う人物であると考えられる。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。また、実験協力者B, Cは表3内で太字のセルが一つもないことから、信憑性評価が平均的な回答者であると考えられる。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、使用時の具体的なクチコミが信憑性を高く評価されていた。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されていた。

実験協力者A, G, Iと他の実験協力者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に着目していた。つまり、実験協力者内でA, G, Iの3名は信憑性を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2で示した信憑性評価のcos類似度は低いいため、信憑性を低いと評価する基準は異なると考えられる。一方、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できなかった。

なお、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。この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、ただのクレーマーのように思え、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

表 2 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	

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。我々は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、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、分散が大きいクチコミに着目し、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、分散が 1.0 以上であった 45 件について、クチコミ収集時に付与した特徴 8 項目を用い、分散が大きいクチコミの特徴を分析した。

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。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、それぞれ共起ネットワークを描画した。描画結果を図4、5に示す。

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られるため、図 4、5 内で「信頼」と繋がりがある単語は、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い。

Subgraph:

- 01 05
- 02 06
- 03 07
- 04 08

Coefficient:

- 0.1
- 0.2
- 0.3
- 0.4

Frequency:

- 100
- 200
- 300
- 400
- 500

Subgraph:

- O1 (Red)
- O2 (Yellow)
- O3 (Green)
- O4 (Blue)
- O5 (Purple)
- O6 (Orange)
- O7 (Light Green)
- O8 (Pink)
- O9 (Dark Blue)

Coefficient:

- 0.2
- 0.3
- 0.4
- 0.5

Frequency:

- 25
- 50
- 75
- 100
- 125

	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
分散平均以上(122 件)	0.95	0.23	0.52	0.26	0.19	0.10	0.13	0.22
分散 1.0 以上(45 件)	0.93	0.18	0.58	0.33	0.20	0.07	0.13	0.31
変化量(全体-分散 1.0)	0.01	0.18	0.01	-0.02	0.00	0.06	0.06	-0.19

調査で用いた化粧品カテゴリについて、各カテゴリのクチコミに対する評価理由（Q4）に着目し、最小出現数を15とし、共起ネットワークを描画した。描画結果を図6～8に示す。図6より、「カバー」「毛穴」「パウダー」などはファンデーション特有の単語であり、ファンデーションのクチコミの信憑性を評価する際は、これらの単語に関する効果についての記述が重視されている可能性が示唆された。図7では、「コスパ」が含まれており、化粧水特有の単語であった。つまり、化粧水のクチコミの信憑性を評価する際はコストパフォーマンスの良し悪しについての記述からも信憑性を評価していることが示唆された。また、図6、7に共通して、「肌」「自分」「合う」が含まれている。このことから、ファンデーションや化粧水のような肌に直接塗る化粧品についてのクチコミでは、肌についての記述が重視され、特に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。図8より、「色」「発色」「落ちる」「荒れる」などがリップ特有の単語であった。そのため、リップについてのクチコミの信憑性を評価する際は、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8では「使い方」が含まれていることからリップについてのクチコミでは使用方法についての記述は重要であると示唆された。

4章のデータセット構築では, 13名に協力を依頼し, 9名の回答を集めることに成功した. ここで, 分析結果から, 9名の中には, A, G, Iのように信憑性を低く評価する傾向をもつ人, Dのように信憑性を高く評価しやすい人など様々な特徴をもつ人がいることがわかった. さらに, A, G, Iの3名は回答の平均や分散などの特徴は似通っているが, $\cos$ 類似度は低いことから, 回答のパターンが様々な存在すると考えられる. さらに, 信憑性の評価理由を分析すると, 同じクチコミに対して異なる観点から信憑性を評価していることが多くあった. したがって, データセット構築においては, さらに多くの実験協力者に回答してもらうことで, より多面的な信憑性評価軸の提案が可能になると期待される. また, 4章の調査より, 信憑性の評価軸としていくつかの候補を検討した. 本章では, この候補について述べる.

Subgraph:

- 01 (light blue)
- 02 (light green)
- 03 (light purple)
- 04 (light orange)
- 05 (light yellow)
- 06 (light pink)
- 07 (light blue)
- 08 (light green)
- 09 (light purple)
- 10 (light orange)
- 11 (light yellow)

Coefficient:

- 0.1
- 0.2
- 0.3
- 0.4
- 0.5
- 0.6
- 0.7

Frequency:

- 100
- 200

Subgraph:

Category	Color	Frequency Range
01	Light Blue	07 - 08
02	Yellow	09 - 10
03	Purple	11
04	Dark Blue	-
05	Orange	-
06	Green	-

Coefficient:

- 0.2
- 0.4
- 0.6

Frequency:

- 50
- 100
- 150
- 200
- 250

また、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頼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期待される。この単語の有無を提示することで、文

章が好みでないという理由で信憑性が低く評価されたクチコミについても、客観性のある信憑性評価が可能になると考える。文章が好みでないという評価理由がついたクチコミは、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憑性を評価されなかった可能性がある。正当な評価を行うことができれば、閲覧者は、さらに幅広いクチコミを閲覧することが可能になると期待される。

また、実験協力者 A, G, I がもつ他の協力者にはなかっ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。

これらの評価軸を利用し、実際にクチコミの信憑性を評価できるのかを今後検証する。今回検討した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。そこで、この評価軸について検証を行うことで、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ことができる。ユーザが正しく評価できていない部分についてさらに検討を重ねることで、ユーザが信憑性を評価する支援が可能になると考えている。

6. 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、まず、化粧をする人が化粧自体にもつ印象や化粧品の情報収集に用いる手段などを明らかにするための事前調査を行った。その結果、化粧品に対するクチコミの信憑性評価基準は人によって大きく異なり、文章、画像、投稿者など様々な観点から判断していることがわかった。

そこで、事前調査の回答で特に言及されることが多かった文章に着目した調査を行うために、実際の化粧品に対するクチコミを 300 件収集し、9 名に信憑性についてそれぞれ判断してもらい、データセットを構築した。また、このデータセットを用いた分析を行った。調査の結果、同じクチコミを提示しても人によって信憑性の評価は異なることや、いくつかの評価基準があり、独自の基準をもつ実験協力者もいたこと、クチコミがもつ特徴によっては実験協力者間で評価がブレやすくなることなどが示された。また、調査結果を踏まえ、信憑性評価軸の候補について議論を行い、信憑性の判断に影響を与えられられる単語を用いることや、独自の観点から信憑性を評価している実験協力者の評価基準を用いることなどを検討した。

今回は女子大学生 9 名の判断結果について主に分析を行ったが、9 名の間でも信憑性評価の基準に大きく差があることが明らかになった。今後はさらに実験協力者を増やし広い年齢層に依頼することで、より多面

的な評価軸の提案を目指す。また、5 章で述べた信憑性評価軸の候補についての有用性はまだ確認できていないため、この評価軸を用いて実際にクチコミの質の高さを判別できるのかを検証する評価実験を行う予定である。さらに、今回は文章のみに着目した調査を行ったが、今後は文章以外の画像やクチコミ投稿者といった側面からも調査・分析を行い、化粧品のクチコミの信憑性を様々な評価軸を用いて判断できるようなシステム設計をする必要があると考えている。

謝辞 本研究の一部は、JSTACCEL（グラント番号 JPAJAC1602）の支援を受けたものである。

参 考 文 献

- [1] ポーラ文化研究所, "化粧への期待と行動", <https://www.cosmetic-culture.holdings.co.jp/report/pdf/191212kitai.pdf>, (参照 2020-10-12)
- [2] Cocon de Soie, "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-breakage-warning-issued>, (参照 2020-10-05)
- [3] A. Mukherjee, B. Liu and N. Glance, "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 2020-10-06)
- [5] 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.
- [6] 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.
- [7] 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.
- [8] G. Moran and L. Muzellec, "eWOM credibility on social networking sites: A framework", Journal of Marketing Communications, 2014, vol.23, Issue 2, pp.149-161.
- [9] 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.
- [10] H. Huang, Y. Wen and H. Chen, "Detection of False Online Advertisements with DCNN", Proceedings of the 26th International Conference on World Wide Web (WWW'17), 2017, pp.795-796.
- [11] 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.
- [12] 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.
- [13] 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.