

評価表現辞書の精錬による化粧品レビューの推薦精度向上

Improving recommendation adequacy of cosmetics reviews by refining evaluation expression dictionary

今江 柚実*
Yumi Imae

松下 光範
Mitsunori Matsushita

関西大学 総合情報学部
Faculty of informatics, Kansai University

Abstract: The aim of this study is to improve the recommendation adequacy of the system for providing cosmetic reviews. In recent years, to help a user's exploration of cosmetics, a system has been proposed that recommends cosmetics reviews close to the user's preference. The system, however, sometimes provide reviews with incorrect scores. The reason is that the value of the evaluation expression dictionary used by the system, which was created from the review corpus automatically, was incorrect, hence the system miscalculated the evaluation score. To solve this problem, this paper proposes a method to refine the evaluation expression dictionary; the method categorizes review sentences with the naive Bayes method and extracts sentences that mention the effect of the cosmetics product from the review corpus.

1 はじめに

化粧品の種類は日々増加している。これは、単なるブランド数の増加によるものではなく、肌質や使用感に焦点をあて、各ブランドが商品展開を広げているためである。利用者は、これらの多くの商品を比較し、自身の嗜好や条件に合致するものを選択する。化粧品は人体に直接塗布するという性質上、誤った化粧品を選択すると肌トラブルを引き起こす原因となるため、購入前に情報を収集することが重要である。近年では、利用者がこうした情報を集める手段として、化粧品レビューサイトが参照されている。レビューサイトの利用には様々な人の使用感や感想を取得できるというメリットがある。しかし、ひとつの商品に対し多数のレビュー文が投稿されているうえ、各レビュー文中には使用感や価格など様々な情報が混在しているため、それらのレビューを全て確認して必要な情報を得ることは困難である。これに加え、同じ価格や使用感であってもその捉え方には個人差があるため、利用者自身の嗜好や肌質に合致する化粧品の情報を効率的に取得することは容易ではない。

この問題を解決するため、松波らは利用者の好みや使用感に着目し、価値観の近い投稿者によるレビューを推薦するシステムを提案している [1]。このシステム

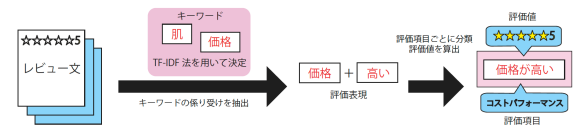


図 1: 評価表現辞書の構築の流れ (文献 [2] より抜粋)

では、化粧品の評価に関するキーワードとそれを修飾する特徴や程度、否定の有無で構成される評価表現に評価値を付与した辞書（以下、評価表現辞書と記す）を用いて化粧品の得点を算出しているが、この辞書の作成は手動で行われており大きなコストがかかっていた。この処理の軽減を目的として、酒井らは評価表現辞書の自動構築手法を提案した [2]。酒井らの辞書構築の流れを図 1 に示す。

酒井らが作成した辞書は、松波が手動で構築した辞書を参考に、@cosme¹のレビュー文や評価値のデータを用いて構築されていた。しかし酒井らの辞書には、辞書の評価表現の核となるキーワードに評価と関係のない単語が含まれていることや、「シミ」+「できる」のようなネガティブと判定すべき評価表現に高い評価値が付与されていたり、「シミ」+「なくなる」のようなポジティブと判定すべき評価表現に低い評価値が付与されていたりするなど、精度の点で問題があった。この問題の原因として、辞書の作成に用いるレビュー文

*連絡先：関西大学総合情報学部
〒569-1095 大阪府高槻市霊仙寺町 2-1-1
E-mail: k943749@kansai-u.ac.jp

¹<https://www.cosme.net/> (2020/3/3 確認)。

に本人の肌質に関する情報など商品の効果に関わらない文が含まれていることが挙げられる。こうした問題を解決するため、本研究では評価表現辞書に使用するレビュー文を商品の効果に関わる文に絞ることで評価表現辞書の精練を試みる。

2 関連研究

レビューサイトを分析することで、利用者の情報取得を支援したり、開発者の手助けとなる情報を抽出したりする研究が様々に行われている。本章では、レビュー情報を用いた研究とナイーブベイズ分類器を用いた研究について述べる。

開発者にとってレビューの情報を得ることはアイデアを得るきっかけになるなど重要であるが、レビュー文中には不必要な情報も多く含まれる。こうした観点から、Walidらは、App storeのレビュー文を分析し、バグの報告、機能へのリクエスト、ユーザの経験、評価への自動分類を可能にした[3]。開発者などにインタビューを行い、手法による精度を比較したところ、メタデータのみを基にした分類器は特に精度が悪く、テキストベースのナイーブベイズ分類器ではより精度の高い分類を確認した。

紀本らは、映画レビュー文から映画の内容や評価に関する情報を抽出し、利用者に提示する手法を提案した[6]。ある物事を評価する語(評価語)とし、これにより評価される語(評価属性)を特徴語として抽出した。紀本らによる特徴語抽出手法とTF-IDF手法により得られた単語と比較するため、評価やレビューの数である「評価数」と、映画の内容である「内容」、映画の中で肯定的か否定的に評価されている点である「評価点」、意見の傾向である「傾向」の4つの項目についてアンケート調査を実施した。その結果、紀本らの手法はこれらの項目のうち、「評価数」「評価点」「傾向」の評価を得る際に有用であることが確認された。

金子らは、購買意図によって商品の検索を行うシステムの階層的知識を自動獲得するため、商品レビューから抽出した購買意図の表現や、商品のカテゴリ情報を用いて購買意図の構造化を行った[5]。レビューをCaboCha²を用いて解析し、それに係る目的語を含む文節を抽出した。この抽出した表現に対しアノテーションを行い購買意図を表現している表現を判断した。この判断した表現が出現するカテゴリの集合の包含関係とcos類似度から購買意図の階層化を行った。構造化の観点からの分析や階層的知識の実用性の調査により「美肌を作る」などの具体的な意図よりも「痩せる」というような抽象的な意図の方が、出現カテゴリ数が多いという仮説が示された。また「セルフネイルをやってみる」

²<https://taku910.github.io/cabocha/> (2020/1/31 確認)

と「ジェルネイルデビューする」のように、意味が似ている表現間で、カテゴリ毎に頻度分布の形も似る傾向にあることが示唆された。

漆原らは、商品販売サイトのレビューにおいて商品の評価を一目で判断できるようにするため、評価表現の印象がポジティブかネガティブかを示す印象極性に色を付けて表示する手法を提案した[4]。この印象極性に対し、「悪く」の直後の「ない」のような否定表現を抽出し、その複合語の印象極性を反転し、文脈上に逆接表現が含まれる場合に直後の評価表現を修正した。人手で作成した正解データと比較し印象極性の推定実験を行い、辞書による判定ができなかったものが判定できるようになったことが確認された。

細田らは、青少年のインターネット・リテラシーレベルの向上のため、twitter³情報を用いて、ナイーブベイズ分類器を導入したリテラシーレベルを評価する問題解決環境を提案した[7]。この問題解決環境とは、ユーザが特別な知識を持たずともそれを評価できる環境のことである。ここではツイート内容が誹謗中傷であるか否かが対象であるため、時間などの評価に関係のない情報や、引用ツイート、URLは評価の対象から取り除いた。形態素解析したツイートに誹謗中傷のキーワードを含んでいるかを分類し、含まれていた場合、再度ナイーブベイズ分類器によりそのツイートが誹謗中傷かを分類した。これにより、従来手法であった誹謗中傷内容の単語によるパターンマッチングでの識別と比較して1.71倍高い結果が得られた。

本研究ではこれらの手法を参考に、評価表現辞書に用いるデータを商品の効果に関わる文に絞り、辞書の精練を行う。

3 評価表現辞書の精練手法

本研究ではレビュー文を商品の効果に関わる文であるかに注目し分類して、辞書構築に用いるレビュー文を商品の効果に関わる文に絞り、自動作成の際の精度の向上を図る。計算機による自動分類を実現するため、本研究では教師あり機械学習手法のひとつであるナイーブベイズ分類器を用いてレビュー文の分類を行い、その分類した文を用いて評価表現辞書を作成した。

3.1 レビュー文の収集

辞書構築に用いるレビューデータは、肌質による違いを評価表現辞書に反映するため、「普通肌」と「混合肌」の2つの肌質について、各々「化粧水」「口紅」「ファンデーション」の3つのカテゴリ(計6種類)を対象とした。この6種類について@cosmeの最新口コミラン

³<https://twitter.com/> (2020/1/31 確認)

キング上位 50 商品を選定し、1 つの商品につき 100 件の評価値とレビュー文を取得した。ここで、辞書の評価表現に対する評価値に、レビュー文に付与されている評価値を使用するため、評価値が「評価しない」と設定されているレビュー文は除去した。レビュー文の取得には、BeautifulSoup⁴を用いた。このうち、10 商品分の各 50 件を分類の学習に用いるデータとした。

3.2 ナイーブベイズ分類器を用いた商品効果文の抽出

前節で収集したレビュー文に対し、句点や、否定後の読点で文を区切り、これを人手で商品の効果に関わるものと関わらないものの 2 つに分類した。これらの文を単語ごとに分割し、その単語数をカウントさせナイーブベイズ分類器で学習した。

句点や、「ですが」のような否定形の後の読点である場合、「乾燥がひどく悩んでいましたが、これを使うとしっかりとりました」のように、読点の前後で記載されている情報が商品の効果に関する文と関わらない文に変化する可能性があるため、句点や否定後の読点で区切った。ナイーブベイズ分類器を用いて、これらの文を商品に関する効果が記載されている文と、レビューする商品の効果に関わらない文の 2 種類に分類した。この分類には機械学習ライブラリ scikit-learn⁵を用いた。

商品の効果に関する文のみを「yes」、本人の肌質や、他の商品の話題、商品の何が効果的であるのか記載されていない文などの、商品の効果に関しない文を「no」として人手で訓練データを作成した。

レビュー文が商品の効果に関する情報であるか否かという文章の分類に対しては肌質による違いはなかったため「化粧水」「口紅」「ファンデーション」のカテゴリごとに、普通肌のレビュー文の学習と混合肌のレビュー文の学習において同じ教師データを用いた。本研究では全ての品詞を対象にストップワードは設定せずに、それぞれ約 3000 文を訓練データとして扱うこととした。学習させるデータのサイズを統一させるため、形態素解析を行い文章を単語ごとに分割し 1 つの文章に含まれている単語数をカウントした。形態素解析には MeCab-0.996⁶を用いた。この際、「ネイル用品」が「ネイル」と「用品」に分割されるように 1 つの単語が複数の単語として認識されることを防ぐため、事前に MeCab に対し化粧品のカテゴリ名やブランド名などに対応したユーザ辞書を登録した。

このように作成したデータの 8 割を訓練データ、2 割をテストデータとして、学習数での精度を測った。訓練

データ数は、化粧水が 3191 文、ファンデーション 3192 文、口紅 3207 文であった。精度はいずれのカテゴリにおいても 78 %程度であった。

これらを学習させ、「yes」と判断された文のみを抽出し、本研究で評価表現辞書に用いるデータとした。

3.3 評価表現辞書の構築

前節で抽出した化粧品の効果に関する文集合を対象として、酒井らの手法を参考に評価表現辞書を構築した。評価表現は「ニキビ」や「乾燥」などのキーワードを修飾する語とした。キーワードの設定には、「化粧水」などの 1 つのカテゴリにおいて、どの商品のレビュー文にも用いることができる辞書を作成するため、化粧品ブランド名や商品名をストップワードとした。これに加え、「いつ」などの使用感を示さない単語がキーワードに入ることを防ぐため SlothLib⁷を用いてストップワードを拡張した。

以上の前処理を行った化粧品のレビュー文の中から MeCab を用いて名詞のみの単語を抽出した。さらに、TF-IDF 法を用いて抽出した単語ごとの重要度を算出した。この算出した重要度が上位 1 % の単語を評価表現辞書の評価表現の核であるキーワードとした。このように設定したキーワード数は、酒井らの手法を用いて作成した辞書では約 280 個、本研究の手法で作成した辞書では約 170 個であった。係り受け解析器 Cabocha を用いて設定したキーワードを修飾する形容詞、副詞、動詞、サ変接続の名詞、助動詞の「ない」を収集し、評価表現である「特徴」「程度」「否定」の抽出を行った。

評価値は、投稿者がその商品に対し付与した@cosme 上のクチコミ評価の星の数を用いた。また、設定した「キーワード」「特徴」「程度」「否定」「評価値」を辞書に登録した。

このようにして登録した辞書に@cosme に効果として付与されている「うるおい」や「アンチエイジング」などの単語を参考に評価項目を設定した。Word2Vec を用いて、評価表現である「キーワード」「特徴」と設定した「評価項目」の類似度を算出し、最も値が高いものをそれぞれの評価表現に対する評価項目として割り当てた。評価項目と評価表現の単語が一致する場合、その単語を、類似度を算出せず評価項目とした。

4 実験

4.1 辞書の精度比較

酒井らの評価表現辞書による精度と、本提案手法により作成した評価表現辞書の精度の比較実験を行った。

⁴<https://www.crummy.com/software/BeautifulSoup/bs4/doc/> (2020/1/31 確認)

⁵<https://scikit-learn.org/stable/> (2020/1/31 確認)

⁶<https://taku910.github.io/mecab/> (2020/1/31 確認)

⁷<https://ja.osdn.net/projects/slothlib/> (2020/3/2 確認)。

実験参加者は20から50代の女性15名で行った。実験には化粧水、口紅、ファンデーションの3カテゴリにおいて、普通肌と混合肌の2つの肌質を対象にした。この対象においてレビュー文をランダムにそれぞれ30件をあらかじめ用意した。評価値は、@cosmeの星の数と同様に設定しており、評価値が高いほど、評価の良いものとした。実験参加者は、レビュー文を読み、評価項目に対し0から7の評価値を付与することが指示された。レビュー文から判断不可能な評価項目がある場合は、その項目の欄に「×」を記入するように指示された。

評価表現辞書を用いた評価値として、レビュー文から取得した評価表現と同様のものを評価表現辞書の中から探し、評価項目に対する評価値を算出した。これを酒井らの手法の評価表現辞書と本手法による評価表現辞書の2つの辞書において行った。精度の比較方法としては、酒井らの提案した評価表現辞書を用いた評価値と、実験参加者に付与してもらった評価値の平均絶対誤差を算出した。これに加え、本研究の提案手法にて作成した辞書を用いた評価値と、実験参加者に付与してもらった評価値の平均絶対誤差を算出し、この2つの平均絶対誤差を比較した。

個人の化粧への関心や化粧品レビューの使用度や違いによって評価値に変動があるかを調査するため、事前アンケートを行った。実験参加者は化粧の経験の有無、化粧にかかる時間、化粧の頻度、購入方法、化粧品のレビューを普段見るか、見る人に対してはどのくらいの頻度であるかが尋ねられた。15人全ての人が化粧経験があり、週に6回以上化粧をしていることが確認できた。購入場所はネットと店頭に半分程度に分かれたが、化粧品のレビューを確認する頻度は、商品購入時に確認する程度であった。

実験の結果を図2、図3、図4に示す。化粧水（普通肌）では5項目に対し平均1.27、化粧水（混合肌）では3項目に対し平均0.65、口紅（普通肌）では7項目に対し平均0.73、口紅（混合肌）では7項目に対し平均0.41、ファンデーション（普通肌）では6項目に対し平均0.44、ファンデーション（混合肌）では7項目に対し平均0.57の平均絶対誤差が減少した。その他の15項目については、平均絶対誤差の減少は確認されなかった。

4.2 評価表現辞書のキーワードの比較

キーワードは評価の対象となる単語であるため、どの商品に対しても扱えるものでなくてはならないが、酒井らの手法による辞書には適当でないキーワードが含まれていた。化粧水におけるキーワードの例を表1に示す。酒井らの手法による辞書では「サンプル」や「キャ

表1: キーワードの比較

酒井らの辞書にのみ出現	本研究の辞書にのみ出現
コスメ	もちもち肌
肌質	トロツ
年齢	ノリ
評価	皮膚
クチコミ	つけ心地
大丈夫	ローズ
風呂上がり	アロマ
現品購入	テカリ

ンペーン」「大好き」「正直」など商品の効果に関わらないものがキーワードに含まれていたが、本研究の手法の辞書では除去されていた。代わりに「トロツ」や花の名前など効果を表すキーワードが追加されたことが確認できた。しかし、「シミ」や「年齢」のように評価を表すにあたり必要なキーワードにおいて、酒井らの手法により提案された辞書のみでしか確認できないものがあり、辞書に必要なキーワードも除去されていたことがわかった。

5 考察

前節で確認した実験結果を踏まえ、酒井らの提案した評価表現辞書と本研究で提案した評価表現辞書の違いを考察した。これに加え、実験を行った結果、平均絶対誤差の減少が確認されなかった問題点や改善すべき点について述べる。

5.1 評価値について

酒井らは、逆説の接続詞または接続助詞の前後で評価が反転すると指摘されていることから、接続詞の位置によって評価値の変動を与えていた。文中に逆説語を含む場合は前の評価値を、文頭に逆説語がある場合は文字数の少ない方の評価値を、7から与えられている評価値を引いた値としていた。しかし、本研究では1文ずつ否定の表現後の読点で区切って分類を行ったため、酒井らが行っていた否定による評価値の変化を与えることができなかった。否定的な表現に対してもそのままの評価値を当てはめており精度を下げてしまう原因となる。このため今後は1つずつの評価表現に対しポジティブかネガティブの判定を行うなど、他の評価値の算出方法を検討する必要がある。

5.2 辞書の精練について

分類をする上で否定形後の読点で文を区切ったため、読点後の文に主語がなくなる問題があった。例えば、「毛穴が隠れないと聞いて心配していましたが、私は気になりませんでした。」のように1度しか主語が現れない場合において「毛穴が隠れないと聞いて心配していましたが、」と「私は気になりませんでした。」を1文ずつ取得することになるため、1つ目の文は「レビュー文とは関係のない商品についての記述」「効果が具体的に示されていない記述」として、どちらの表現も商品の効果に関しない文と認識されていた。このため「毛穴が気にならない」という評価表現に対する評価値を算出することができなかった。このため今後は文を区切る前に文の主語を抽出する必要がある。

5.3 キーワードの設定について

前節で述べたように、本手法を用いて用いるデータを商品の効果に絞ることで評価表現辞書のキーワード中に「大丈夫」や「キャンペーン」など評価に関わらない単語が除外され、代わりに「トロッ」や花の名前など評価に関わる単語が加わった。しかし、「シミ」や「年齢」という評価に関わる単語も除外されていた。これは、TF-IDF 値の上位 1% の名詞をキーワードとしたため、キーワード数が酒井らの手法に比べ 100 個ほど少なくなったことが原因であると考えられる。そのため今後は、キーワードの取得数を検討する必要がある。

5.4 肌質による辞書の違いについて

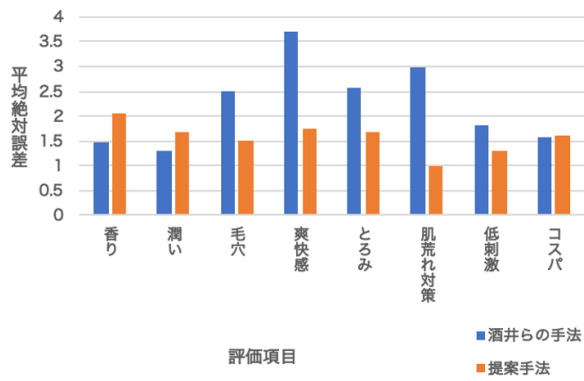
本研究では、肌質による評価値の変化を確認するため 2 つの肌質に対し評価表現辞書の構築を行った。これは、@cosme の「肌質」の項目に用いられている肌質から「普通肌」と「混合肌」を抽出し、それぞれの肌質のレビュー文を収集した。しかし、レビュー文を分類する上で「普通肌」として取得したレビュー文の中に、「私は乾燥肌なので」や「私は混合肌ですが」という表現が現れており「肌質」に設定している肌質とは異なる投稿者が多く確認された。より有用に肌質を考慮した辞書を構築するためには、文章中に出現する肌質も考慮する必要がある。今後は、1 つずつのレビュー文の中から「普通肌」「敏感肌」のように肌質名を抽出し、肌質を決定するべきである。ただし、「私は乾燥してしまいましたが脂性肌の人には良いと思います」のように投稿者の肌質でない文が含まれる。これを、投稿者の肌質と設定することは精度を下げる原因となるため、他の人に対して推薦する文や肌質を述べた文が過去形である時、肌質としない設定をする必要がある。

6 おわりに

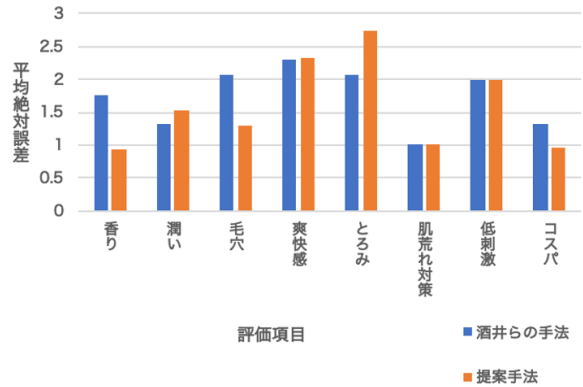
化粧品レビューサイトはレビュー数が多く、また使用感などの情報が混在しているため、利用者が求める情報を取得することは難しい。この問題を解決するため、個々の利用者に対し有用なレビュー文を推薦するシステムが提案されているが、そのシステムを構築する際に用いる評価表現辞書の精度が悪いという問題があった。本研究では化粧品の評価表現辞書に用いるレビュー文を商品の効果に関わる文に絞ることで自動構築における評価表現辞書の精練を行った。先行研究による手法と本提案手法の評価表現辞書の精度を比較した結果、35 項目の評価項目に対し、平均絶対誤差の減少が確認された。今後は、分類後に主語がなくなることへの対策、辞書構築における評価値の算出方法、キーワードの設定数、投稿者の肌質の判定方法、について検討し、より精度を向上させることを目指す。

参考文献

- [1] 松波友稀, 上田真由美, 中島伸介: コスメアイテムに対する評価項目別レビュー自動スコアリング方式の開発, 第 9 回データ工学と情報マネジメントに関するフォーラム, B5-3 (2017).
- [2] 酒井美春, 松下光範, 上田真由美: 化粧品の評価項目別スコア生成のための評価表現辞書の自動構築, 第 11 回データ工学と情報マネジメントに関するフォーラム, B6-2 (2019).
- [3] Walid, M., Zijad, K., Hadeer, N. and Christoph, S.: On the automatic classification of app reviews, *Requirements Engineering*, Vol. 21, No. 3, pp. 311-331 (2016).
- [4] 漆原駿, 町田翔, 延澤志保: 印象極性による商品レビューの視覚化, 情報処理学会第 80 回全国大会講演論文集 (第 2 分冊), pp. 311-312 (2018).
- [5] 金子貴美, 村上浩司, 石野亜耶: 商品レビューからの購買意図の抽出と自動階層化, 人工知能学会全国大会論文集, pp. 2P113in1(2016).
- [6] 紀本雅大, 伊藤淳子, 宗森純: 評価表現に着目した映画レビューからの評価情報抽出, 情報処理学会研究報告, Vol. 2019-GN-106, No. 43, pp. 1-8 (2019).
- [7] 細田尚志, 前田太陽, 井上聡, 石崎博基, 権藤俊彦: ナイーブベイズ分類器を用いたインターネット・リテラシーレベルを評価する問題解決環境の有効性, 日本計算工学会論文集, Vol. 2017, Paper No. 20170003 (2017).

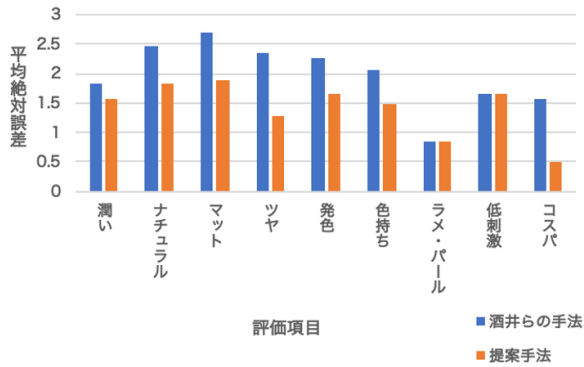


(a) 化粧水（普通肌）の平均絶対誤差

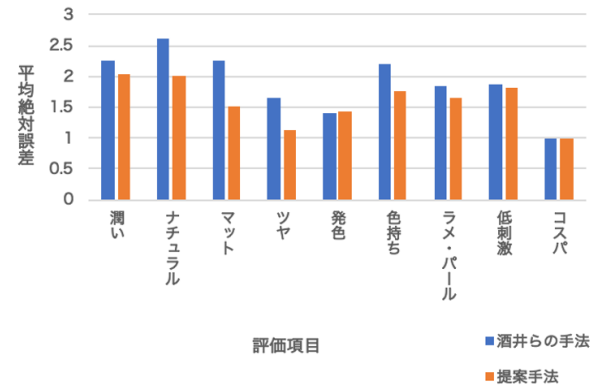


(b) 化粧水（混合肌）の平均絶対誤差

図 2: 化粧水の結果

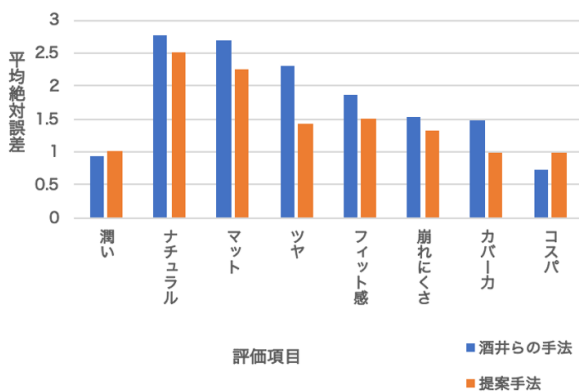


(a) 口紅（普通肌）の平均絶対誤差

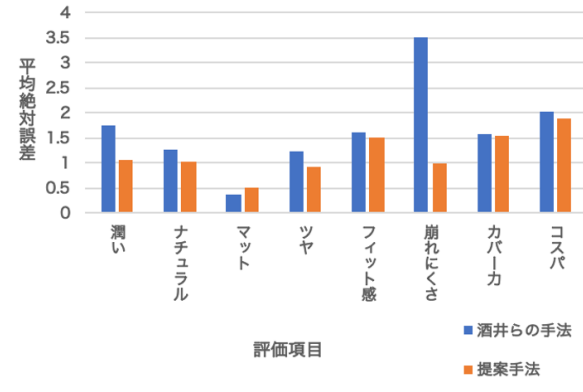


(b) 口紅（混合肌）の平均絶対誤差

図 3: 口紅の結果



(a) ファンデーション（普通肌）の平均絶対誤差



(b) ファンデーション（混合肌）の平均絶対誤差

図 4: ファンデーションの結果