

化粧動画の工程検索を指向した 化粧語彙セット構築の試み[†]

安尾 萌^{*1}・河野 雪乃^{*2}・山西 良典^{*2}・松下 光範^{*2}

本研究の目的は、化粧のプロセスを紹介する動画における特定の工程を検索可能にすることである。近年、化粧方法を紹介する動画が化粧を行う際の参考情報として利用されている。現在の化粧動画の検索は、通常の動画検索と同様に動画単位での情報収集が主であり、化粧動画の特定の工程を対象とした検索はできない。この課題を解決するには、化粧に関する専門用語を考慮した語彙セットを構築し、化粧特有の文脈を考慮した検索結果を獲得可能にする必要がある。本稿では化粧品のレビュー文と化粧動画の発話を併用した化粧語彙セットを構築し、各情報源における化粧語彙セットの特性を分析した。その結果、同一のラベルが付与されたセンテンスであっても、商品レビューからは化粧品の持ち運びに関する内容が抽出される一方、化粧方法を紹介する動画からは化粧品の使用による変化の情報が抽出されるなど、データソース間で取得可能な情報が異なることが確認された。

キーワード：語彙分析, 化粧, レビュー分析, 動画内発話

1. はじめに

化粧は現在、特に女性の日常的行為として広く行われている。化粧を行う動機は趣味嗜好、社会的要請、自己実現など多岐にわたり [1,2], それぞれの目的に合わせた化粧法が考え出されている。化粧は皮膚に直接塗布するその性質上、各商品と使用者の皮膚の色や肌質などとの相性が複雑であり、適切な商品や塗り方は個人によって異なる。この目的の多様性、および相性の複雑さが原因で、化粧は画一的な価値基準を設けることが難しく、化粧工程やプロセスにおける明確なマニュアルやルールが存在しない [3,4]。そのため、化粧行為のプロセスや目的達成の成否は、各人の目的やプロフィールに適した商品、および手法に基づいて化粧行為を行うことができたかに依存する。

化粧に関する情報の収集源としては、女性向けの雑誌や化粧品のレビューサイトなどが挙げられる。近年では、化粧の一連の工程を紹介する動画（以下、化粧動画と記す）が10~20代の女性を中心に人気を集めており、従来の情報媒体に匹敵する重要な情報源として利用されている¹。こうした現状から、化粧動画は単なるエンタテインメントコンテンツの一ジャンルとしてだけでなく、目的の印象や塗り方のテクニックなどを知るための情報源としての価値を内包していることが伺える。

参考情報の収集を目的として化粧動画を視聴するユーザは、必ずしも動画全体の情報を必要とするわけではない。現在の化粧動画の検索は、一般的な動画検索と同様に動画単位での情報

収集が主となる。そのため、「大人っぽい印象を与えられるようなアイメイクの方法を検索したい」といった、動画に含まれる特定の内容に焦点を当てた検索を行うことはできない。化粧辞書の構築に関する従来研究では、Web上の情報をデータソースとして化粧に指向した語彙を収集・分類し、それらの意味的類似性や文脈依存性を明らかにした上で知識ベースを構築するものがある [5,6]。これらの研究では、主に化粧品に関する販売サイトや口コミサイトに投稿されたレビュー文をデータソースとして語彙の抽出を行っている。そのため、化粧動画のような異なるコンテンツへの適用を検討する場合、データソースの性質による語彙特徴の偏りを考慮し、検索対象に適した語彙セットとして拡張する必要がある。本研究では、化粧に関する参考情報の収集における検索体験の向上を目的として、「化粧動画内における特定の工程」を検索可能にする枠組みの実現を目指す。この目的を実現するため、本研究では化粧動画内に含まれる発話を検索の手がかりとして特定の工程を抽出することを想定した。本稿では語彙セットの拡張に用いるデータソースとして、クチコミサイトに投稿されたレビュー文と化粧動画内の発話を併用することに着目した。従来研究で利用されてきたレビュー文は、商品単位での評価や感想が投稿されることが多い。一方、化粧動画では商品に関する技法や、その技法の適用結果として実現される仕上がりの印象に関する情報が提供される傾向にある。そのため、データソース間で獲得可能な情報の性質が異なることが予想される。

以上を踏まえ、本稿ではクチコミサイトに投稿されたレビュー文と化粧動画内の発話を併用した化粧語彙の抽出と語彙セットの構築を試みる。抽出したテキストデータから語彙を抽出し、単語分散表現を用いて類似するキーワードをクラスタリングすることで語彙セットを構築する。構築した語彙セットを

[†] Construction of Makeup Vocabulary Datasets for Scene Search on Makeup Movies

Megumi YASUO, Yukino KONO, Ryosuke YAMANISHI, and Mitsunori MATSUSHITA

*1 関西大学 大学院総合情報学研究科

Graduate School of Informatics, Kansai University

*2 関西大学 総合情報学部

Faculty of Informatics, Kansai University

1 ポーラ文化研究所化粧調査 2019: <https://www.cosmetic-culture-poholdings.co.jp/report/pdf/191212kitai.eps> [accessed Jan. 28, 2022]

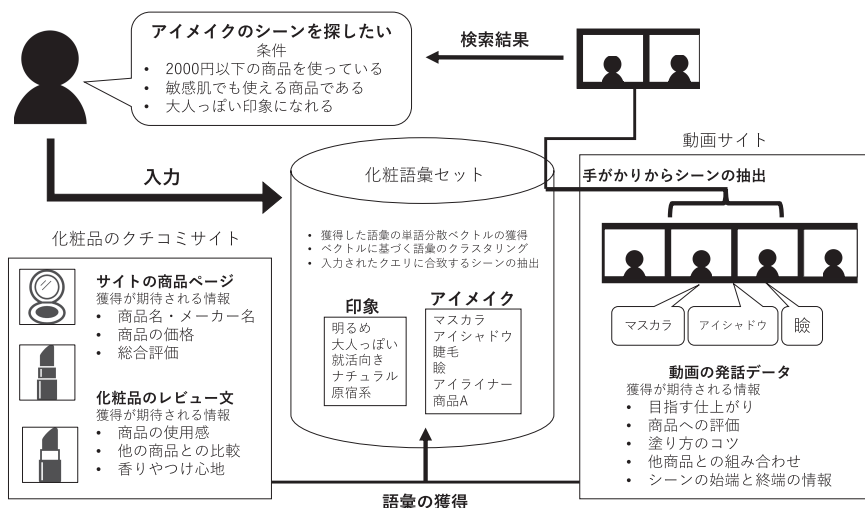


図1 化粧動画内の特定の工程を検索するシステム

利用して、利用される語彙の種類がデータソース間でどのように異なるかを分析し、提案手法の妥当性について検証する。

2. 化粧動画の特定の工程を検索可能にする枠組み

化粧行為を行う上で、使用者は「どの商品を選択するか」という情報と、その商品を「どのように使用するか」という情報を必要とする。前者の要求を満たす情報リソースとして、レビュー文が挙げられる。一般的な化粧品の口コミ投稿サイトでは、化粧品に関する情報が商品単位で分類されており、商品の価格や商品に対してつけられた総合評価などがクチコミとは別に付与されている。こうしたサイトに投稿される口コミは、投稿者のアカウント単位で表示されることが多く、投稿者の年齢や肌質、皮膚の色といった情報を保持していることもある。実際に投稿されるクチコミの内容としては、投稿される商品の内容を中心として、商品の香りやテクスチャーなどの使用感に関する情報、購入動機や利用シーンに関する情報、同ジャンルの他商品との比較に関する情報などが投稿されることが多い。レビュー文には商品ごとにどのようなものか、どのようなシーンで利用したかがテキストとして記述されており、ユーザが購入する商品を検討する際の参考情報となる。化粧品のレビュー文は化粧特有の語彙に対応した辞書構築のデータソースとして、従来研究においても利用されている [7]。

後者の要求を満たす情報リソースとして、化粧動画が挙げられる。化粧動画は一般的に、特定の印象やシチュエーションに沿った仕上がりを最終目的として示し、その仕上がりに至る過程を連続的に紹介する形式で提供される。化粧動画に含まれる参考情報は、使用している商品やその評価、商品間の組み合わせといった商品単位の情報や、目指す印象とそのための効果的な演出、塗り方のコツなど、視覚的に確認することを重視すべき情報が存在する。これらの情報を利用することで、使用者は所有している商品の使用方法やそのバリエーション、目指す印象の演出に関する情報を獲得することができる。ただし、化粧動画の情報は動画単位でパッケージングされており、特定の工

程を検索するためには動画内の情報を手がかりに目的のシーンを抽出する必要がある。

化粧動画の特定の工程を検索する試みは、動画のシーン検索の一種と位置付けることができる。一般的な動画のシーン検索では、映像の特定のフレームや音声を手がかりとして、クエリに合致するシーンを抽出するという手法が一般的である。本研究が対象とする化粧動画の場合、化粧に関する専門用語やユーザの属性に関する情報、また化粧動画が潜在的に持っている情報を考慮した検索を行う必要がある。この検索を実現するためには、化粧に関する専門知識を含んだ網羅性の高い知識ベースを構築する必要がある。専門知識を含んだ知識ベースの構築には、対象となるドメインにおける語彙を獲得する必要がある。この過程で獲得する語彙の網羅性が検索の成否に重要となる。以上を踏まえ、本研究では複数のデータソースを併用することで、化粧検索に対応可能な知識ベースの構築を目指す。

本研究が目指す化粧動画のシーン検索例を図1に示す。図1の例では、検索者が「2000円以下で敏感肌でも使える商品を利用した、大人っぽい印象になれるアイメイク」を探すためにクエリを入力し、シーン検索を行うことを想定している。クエリを入力された検索システムは、検索対象の動画から「動画内で使用されている商品が2000円以下で、敏感肌でも使用可能である」「大人っぽい印象を目的とした化粧行為である」「アイメイクのシーンである」といった情報を獲得し、動画内の情報を手がかりに該当するシーンを抽出する。このシーン抽出における課題のひとつとして、化粧特有の用語への対応が一般的な辞書では困難である点が挙げられる。化粧に関する専門用語は、略称や外来語の変形などが多く、化粧に関する情報を専門としないデータソースからは一般語彙として獲得することが難しい。例えば、デパートなどの商業施設で販売される高価格帯の化粧品を指す「デバコス」や、皮脂腺が多く化粧崩れの起こりやすい顔の部位を意味する「Tゾーン」などが該当する。また、一般的には季節を意味する「春」がパーソナルカラー²

2 皮膚色や目の色など、個人が持つ身体的な色調と調和する色のグループを指す化粧用語。

の一種として利用されるなど、通常時とは意味が異なるケースも存在する。こうした専門用語に対応しないまま検索が行われると、入力されたクエリが検索者の意図通りに解釈されず、また対象となる動画やシーンの切り出しにおいて目的とされないシーンが抽出されてしまう。これらの課題を解決するためには、目的とする動画とそのシーンを獲得するための知識ベースとして、検索対象に対応した専門的な語彙を含む辞書を適用することが前提となる。その上で、化粧動画のシーン検索に対応可能な語彙セットを構築するには、化粧というジャンルにおいて一般的に用いられる語彙に加え、特に化粧動画において用いられやすい語彙特徴を持つデータソースを対象として語彙の抽出を行う必要がある。

3. 関連研究

本研究は、「ユーザの目的に応じた参考情報をいかに抽出するか」に焦点を当てた情報検索に関する研究と位置付けることができる。この課題は嗜好抽出の課題と共通している。本章では、嗜好抽出に関する研究について述べた後、本稿の分析対象であるテキストのクラスタリングを利用した研究、および化粧に関する情報検索のための辞書構築に関する研究について述べる。

3.1 個人の嗜好を反映した情報検索に関する研究

情報検索において個人の嗜好を考慮する試みは多岐にわたる [8]、食事、音楽といった娯楽に関する研究を中心に組み込まれている。多くの場合、言語表現などを利用して印象を間接的に推定する手法や、類似検索と協調フィルタリングを利用する手法が採用される。前者の手法の例として、末吉ら [9] は、ニコニコ動画に投稿された VOCALOID 楽曲動画に付与されたタグと視聴者コメントなどのテキスト情報を利用し、ユーザの嗜好を考慮した楽曲クリエイターの推薦手法を提案している。この研究では、動画に付与されたタグを外部情報として扱い、動画に付けられたコメントから楽曲の印象を推定することで、クリエイターごとの楽曲動画の傾向を整理し、検索に反映している。後者の手法の例としては、段らの研究 [10] が挙げられる。この研究では、アパレル商品のような、デザインを重要視した商品のための推薦システムを提案している。深層学習で用いられる CNN の特徴量として商品の画像形状特徴を利用し、従来の協調フィルタリングのアルゴリズムに統合することで、商品数の増加に対応可能な新しいモデルを提案している。

3.2 テキストのクラスタリングを利用した情報抽出に関する研究

テキスト情報を利用して情報の抽出を行うための最も一般的な方法の一つに、単語分散表現を利用する方法がある。この手法は文書分類などに広く用いられるものだが [11]、現在では様々なデータに適用し、潜在的な情報の獲得に利用される。Magno らは twitter に投稿された 17 億のツイートデータを対象に、国別の文化的特徴や価値観を明らかにすることを試みている [12]。この試みは、「宗教」や「科学」など 22 の観点に基づいて、59 カ国の文化的特徴を分類している。この研究では、

国別の文化的特徴や価値観が明らかになり、いくつかの指標において実際の文化的特徴との相関が指摘されている。

前節で述べた情報推薦を指向した研究においても、単語分散表現は広く利用されている。Park らは、観光ツアーの経路設計に旅行サイトの口コミ情報を利用することで、旅行客のプロファイルに応じた観光経路の推薦を試みた [13]。この研究では、旅行に関する口コミサイトである tripadvisor に投稿されたレビューを対象に投稿者の分析を行い、投稿者のプロファイルによって特有の課題を抱えていることを明らかにしている。また Paul らは、EC サイトに投稿される大量のレビューの中から、より有益なレビューを単語分散表現を用いて抽出することを試みている [14]。この研究は、ある商品について多角的に言及している良質なレビュー文のサブセットと、レビュー文から得られる感情極性を紐づけることを目指している。一般的な EC サイトの中にはレビュー文自体の有用性を評価する機能が付与されているものも多いが、レビュー文から直接有用性のスコアを推定することで、新規に投稿されたレビュー文にも対応可能であることが利点として挙げられている。

単語分散表現を領域横断的に利用した例として、Han らによる歌詞推薦手法がある [15]。この研究は、観光地のレビューと楽曲の歌詞の単語分散表現を共有することで、観光地にいるユーザの環境に類似した歌詞を推薦する。従来の楽曲推薦では、ジャンルやアーティストなどのメタ情報が用いられてきたが、単語分散表現を横断的に利用することで、聴取環境に適した歌詞を推薦することに成功している。

3.3 化粧を対象とした辞書構築に関する研究

前節で述べたテキストのクラスタリングを利用した研究は、一般的な語彙を対象とした単語分散表現の活用が中心であった。一方、本研究で対象とする化粧には、造語や略称、一般語とは意味の異なる語など、多様な専門用語が多く含まれる。そのため、これらの情報を反映させた化粧専用の辞書の構築が求められる。本節では、化粧を対象とした辞書構築に関する従来研究について述べる。

Ueda ら [7] は、特定の商品と類似している化粧品のカテゴリと、類似していない化粧品のカテゴリに付与されたレビューに出現する評価表現の傾向を観察し、各化粧品のカテゴリに対応した評価表現辞書の構築を検討した。この研究では、化粧品のレビューから類出単語を抽出し、各カテゴリのレビュー文から抽出した単語の出現回数を用いて算出されたコサイン類似度の比較を行っている。結果として、化粧品のカテゴリごとに共有可能な評価項目があることが確認されている。酒井ら [6] は、化粧品につけられたレビュー文を利用して、化粧品を対象とした評価表現辞書を構築することを試みた。この研究は、レビュー内に記述された使用感や印象などの情報を抽出することで、ユーザの検索要求に沿ったレビュー文の推薦を可能にすることを目的としている。対象とするレビュー文から、「価格」や「使用感」といった項目に対して付与された「高い」「潤う」などの評価表現を抽出し、レビュー自体に付与された評価値をもとに、それぞれの項目の評価値を推定している。

3.4 本研究の立ち位置

情報抽出の技術的観点からは、本研究は嗜好抽出にテキストのクラスタリングを適用するアプローチとして捉えられる。嗜好抽出に関する従来研究では、テキストだけでなく画像情報や音響特徴を手がかりに目的の情報を推定する手法が検討されている。本研究が対象とする化粧の場合、香りやつけ心地、使用感など、主観的かつコンテンツドメインに深く依存するクエリを扱う必要があるため、そのクエリに応答するための手がかりを視覚的特徴に基づく手法でコンテンツ自体から得ることは難しい。そのため本研究では、テキスト情報を検索クエリとする手法を採ることとする。テキストのクラスタリングを活用した従来研究は、単語分散表現による潜在的知識の獲得が、テキストをベースとした情報検索や情報の推薦、知識の抽出などに有用であることを示している。本研究もこの手法に則り、単語のクラスタリングと語彙セットの構築を試みる。

化粧辞書の構築に関する従来研究においては、データソースとして化粧品のレビューサイトを対象としている。多くのレビューサイトにおけるレビュー文は、商品単位でレビューを記述する形式が採用されている。また、「イエベ春³」といった利用者のパーソナルカラーの情報が含まれていることも多い。そのため、商品単体の効果や特性に関する情報を詳細に収集することを企図する場合、レビュー文をデータソースとして利用することは合理的である。一方、化粧動画では一般的に、複数の商品を利用して化粧の過程を実践する形式をとることが多く、化粧品の組み合わせや利用順序に関わる情報を含むことが特徴として挙げられる。この点を考慮すると、化粧工程に関わる技法や他商品との組み合わせに関する情報を収集することを企図する場合、化粧動画内の発話を利用することは合理的である。このように、化粧動画と化粧レビューでは、化粧という同一のコンテンツに関連している言語情報である一方で内容には大きな違いが存在する。こうした異なる性質を持つデータソースを利用することで、化粧動画の検索へ適用しやすい辞書の構築が期待できるだけでなく、複数の商品間の相性とといった、従来の化粧辞書では得られにくい性質のデータの獲得が期待できる。

4. 化粧語彙セットの構築

本研究が目指す検索は、テキストをベースとした形式の検索を想定している。一般的なテキスト検索では、単語分散表現を用いて文脈的な特徴を獲得し、その特徴を手がかりに目的の情報を検索するケースが多い。しかし、単純な単語分散表現の適用を行うだけでは、多義的な語彙への対応や語彙の関連を考慮した検索が難しい。本研究で扱うような化粧における語彙では、本来季節を表す語である「春」がパーソナルカラーの一種として使用されるケースにあるように、一般語彙として利用される語とは異なる文脈で共通のキーワードが利用されることもある。このような場合、単語分散表現のみを用いた検索は検索者の意図に沿った検索結果を返すことができない。こうした課題に対応するには、特定のジャンルにおける類似した意味の語彙をまとめた単語集合を作成し、類似した概念の語彙を共通し

た概念として扱う必要がある。

そこで本稿では、化粧の特定の工程を検索可能にするための語彙セットを、概念アトリビュートの獲得 [16] に基づいて構築する。概念アトリビュートは、特定のドメインに属する語彙のうち、意味的な類似性の高い単語の集合を指す。概念アトリビュートは以下の手順で獲得される。

1. 化粧に関わるデータソースから語彙を収集する
2. 収集した語彙の単語分散表現ベクトルを獲得する
3. 単語間における類似度の高い語彙をクラスタとして獲得する
4. 各クラスタを説明する概念をラベルとして付与する

この手法は、意味的に類似した単語同士を同一の概念として扱うことで、より人の認知に近い感性で語彙を解釈することを企図している。パーソナルカラーとしての「春」を例にすると、化粧語彙というドメイン下における「春」というキーワードが、使用される文脈ごとに「季節」「色」に該当するクラスタにそれぞれ分類されることで、検索者が入力したクエリの文脈を考慮した検索結果を獲得可能になることが期待できる。

4.1 対象とするデータとデータの抽出

化粧語彙セットの構築にあたり、化粧品に対してつけられたレビュー文と、化粧動画における発話の書き起こしデータをそれぞれ収集した。化粧は複合工程であり、化粧動画内では複数の行為がひとまとまりのコンテンツとして提供される。化粧の工程は一般的に、化粧の下地となる「ベースメイク」、眉に施す「アイブロウ」、目の周りに施す「アイメイク」、頬に施す「チーク」、口に施す「口紅」の5工程が存在する。データソース間で獲得可能な語彙特徴を調査するには、特定の工程に絞った比較分析を行う必要がある。以上を踏まえ、今回は「口紅」および「アイメイク」に関する工程を抽出することを想定し、それぞれのジャンルに関わる化粧品につけられたレビュー文、および化粧動画内の発話の書き起こしを収集した。

レビュー文の収集は、化粧品レビューサイト @cosme⁴ に投稿されたレビュー文を対象に行った。2021年12月29日の時点における@cosmeのアイテムカテゴリで、商品一覧に記載されている化粧品(口紅・グロス・リップライナー 9,979点、アイシャドウ 6,994点、マスカラ 4,332点、アイライナー 3,937点)を商品の掲載順にソートし、スクレイピングを用いて商品ページのURLを取得した。収集対象となった商品は、商品カテゴリ「口紅」に登録された商品530点、商品カテゴリ「アイメイク」に登録された商品のうち、小カテゴリ「アイシャドウ」「マスカラ」「アイライナー」各110点分である。各商品1点に対し最大200件分のレビュー文を収集したところ、最終的なレビュー文数は、口紅45,319件、アイメイク(アイシャドウ、マスカラ、アイライナーの合計)43,385件であった。

化粧動画の発話文の収集は、YouTube⁵に投稿されている化粧動画を対象に行った。化粧動画に関しては、機械的なURLの収集が困難であり、化粧品の商品紹介動画や非日常的な場面

3 彩度や明度が高い暖色系の色が似合う肌の人を表現する化粧用語。

4 <https://www.cosme.net/> [accessed Jan. 28, 2022]

5 <https://www.youtube.com/> [accessed Jan. 28, 2022]

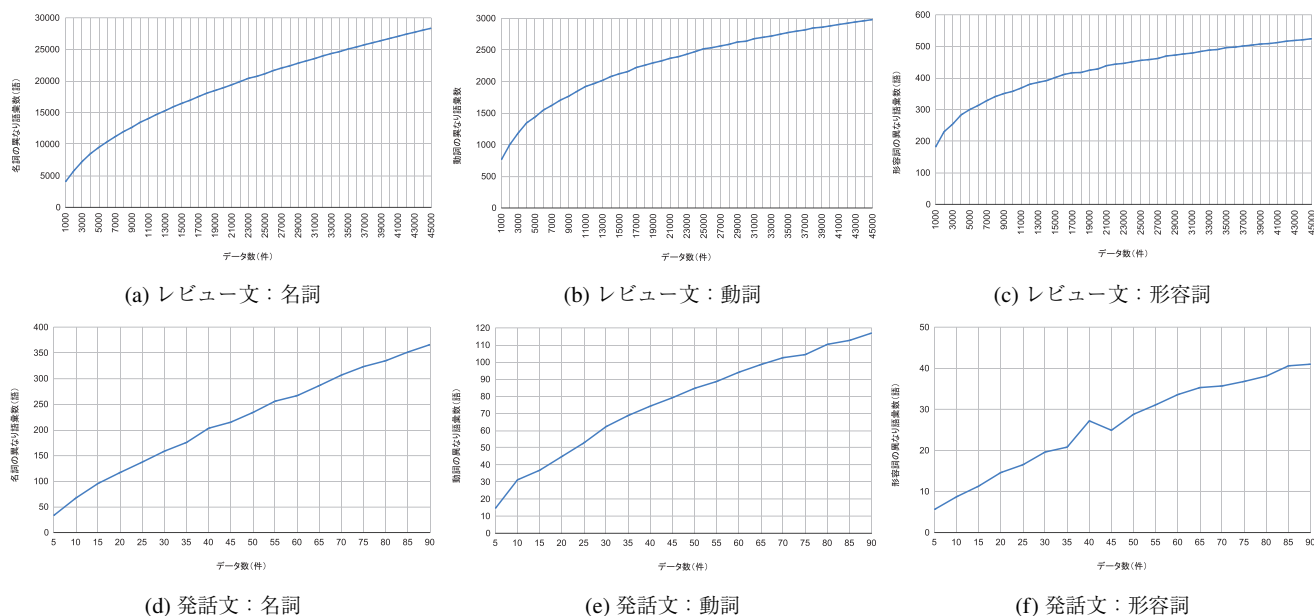


図2 各データソース：品詞別の異なり語彙数の推移

を対象としたメイクに関する動画などが収集されてしまう。そのため、「メイク」をクエリとしたキーワード検索の動画を関連度順で取得し、①取得した動画の内容が「目的に合わせた一連の化粧の工程」を紹介するものであること、②口紅、およびアイメイクの工程が省略されていないこと、③他者の顔に近づけるためのメイク（なりきりメイク）や非日常的な場面を対象としたメイク（e.g., ハロウィンメイク）に関する動画ではないこと、の3点を確認した上で収集し、該当する動画の関連動画から類似の内容の動画を収集するという工程で収集した。収集対象となった動画は150動画であった。発話の書き起こしデータはSavesubs⁶を用いて対象の動画から字幕を収集し、誤字脱字を手手で修正した。この取得した発話の書き起こしデータから、口紅とアイメイクに関する発話部分のみを抽出した。発話抽出の開始/終了の基準は、動画内の発話内容に基づいて行った。開始時点は化粧工程の開始を示す発言（「はい、次にアイメイクをやっていきます」）があるタイミングや、口紅、アイメイクに使用する化粧品が画面に映ったタイミングを基準とした。終了時点は化粧工程の終了を示す発言（「これでアイメイクの完成です」）があるタイミングや、口紅、アイメイク以外の化粧品が画面に映ったタイミングを基準とした。この一連の処理により、口紅に関する内容の発話120動画分、アイメイクに関する内容の発話30動画分を取得した。

4.2 化粧語彙の分散表現の獲得とクラスタ化

前節で収集した口紅とアイメイクのデータを用いて化粧語彙セットを構築する。構築する語彙セットの対象の品詞は、名詞、形容詞、動詞の3つを選定した。名詞からは「リップ」「マスカラ」などの道具名や具体的な商品名の語彙、形容詞からは「かわいい」「きれい」などの印象に関する語彙、動詞からは「塗る」「ほかす」などの技法に関する語彙を取得することを想

定した。

語彙セットの構築に先立って、適切なデータ数の決定のため、収集したデータソースに含まれる形容詞、名詞、動詞を対象として異なり語彙数を調査した。レビュー文、および化粧動画の発話文に対して形態素解析を行い、各品詞ごとに語を取得した。形態素解析にはMecab (Ver.0.996)を用い、辞書にはmecab-ipadic-NEologd (Ver.0.0.7)を用いた。調査対象は「口紅」のレビュー文45,319件、および化粧動画の発話文90件である。レビュー文については1,000件ごと、化粧動画の発話文については5件ごとにランダムサンプリングを行ってから異なり語彙数を算出し、10回の試行の平均値を異なり語彙数として決定した。各品詞ごとの異なり語彙数を図2に示す。分析対象のレビュー文に含まれる品詞のうち、形容詞、名詞の異なり語彙数がおよそ30,000件程度で収束する傾向にあることを確認した。この結果から、語彙セット構築のためのデータ量の基準として、レビュー件数の総量が30,000件を超える商品数を対象とすることを決定した。名詞に関しては、分析対象の範囲内の収束は確認されなかったが、新商品やメーカー名などの増加によって収束が見込みにくいことが想定されるため、データ数の決定基準としては採用しなかった。化粧動画の発話についても同様の分析を行ったところ、いずれの品詞においても収束する傾向が見られなかった。この結果は、化粧動画がエンタテインメントコンテンツの一つであり、投稿者のエピソードなど化粧と無関係な発話が含まれやすいことが原因として考えられる。

最終的な語彙セットの構築対象としたデータ数は、「口紅」「アイメイク」に関するレビュー文が各カテゴリごとに300商品分、化粧動画内の発話文90動画分であった。これらのデータソースから各品詞ごとに重複を除いて語彙の収集を行ったところ、レビュー文から27,889語、発話文からは1,162語の語彙が獲得された。

6 <https://savesubs.com/ja> [accessed Jan. 28, 2022]

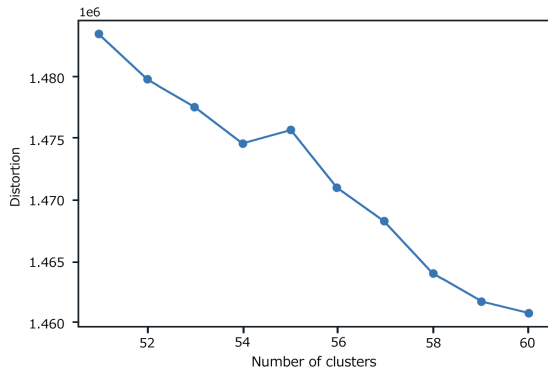


図3 エルボー法によるクラスタ数の決定

次に、得られた各語彙の意味的距離を算出し、類似するキーワードのクラスタリングを行った。本稿では、この処理に鈴木らが作成した日本語 Wikipedia エンティティベクトル⁷を用いた。これは日本語版 Wikipedia の本文全文を学習データとして構築（ベクトルの次元数が 200 次元、window サイズは 5 単語）されており、一般的な語彙の曖昧性に頑健であるという特徴を持つ [17]。この Wikipedia エンティティベクトルを用いて化粧語彙の分散表現を獲得し、これを k-means++ 法 [18] により 55 クラスタに分割することで語彙セットを作成した。クラスタの分割数の決定にはエルボー法を利用した [19]。エルボー法は、クラスタ内の誤差平方和（以下、SSE と記す）とクラスタ数が共に最小となる値をクラスタの分割数とする方法である。今回獲得した化粧語彙に対する SSE の値とクラスタ数を図 3 に示す。この方法に基づき、クラスタの分割数を 55 と決定した。

その後、作成したクラスタの内容に基づき、それぞれクラス名を付与した。クラス名の付与はオンライン調査を利用した。各クラスに分類された語彙からランダムに 30 語を抽出し、8 名の評価者に独立に提示した上で「この語群を意味する最も適切なグループ名をつけてください」という質問を行い、回答を収集した。収集した回答のうち、4 人以上が同一の内容を回答していた語について、他クラスとの重複を許可してラベル名として決定した。付与されたクラス名と各クラスに所属する語の一部を表 1 に示す。ラベル名が一意に定まらなかった項目についてはラベル名を付与せず、分析の対象外とした。最終的に 55 クラス中 48 クラスにラベル名が付与された。

5. 分析

本研究の提案は、化粧動画内における特定の工程を検索可能にするための化粧辞書を構築するために、レビュー文と動画内の発話をデータソースとして併用することで、レビュー文だけでは獲得の難しい性質のデータを獲得可能にすることを企図している。本提案の妥当性を検証するためには、異なるデータソースでそれぞれ得られる情報の性質が異なり、それらを併用することで単一のデータソースのみでは獲得できない情報を収集可能であることを実験的に示す必要があると考えた。そこ

表 1 付与したクラス名と各クラスに所属する単語（抜粋）

クラス名	所属する単語（抜粋）
価格	費用, 需要, 数量, 負担, 価格, 税込
トラブル	瞬時, 落ち, 割れ, 翌朝, 惨事, 崩れ
肌の状態	すみ, しわ, わか, つり, なみ, つや, くま
色	スカーレット, セブン, マット, ベーシック, ホホワイト
芸能	素顔, イベント, 編集, 舞台, 特集, ファン

表 2 付与したラベルと基準

ラベル名	付与した基準
商品	化粧品のブランド名, 品番に関する内容
商品の特性	商品の特性への言及
商品の価格	商品の値段や価格に関する内容
色味	化粧品自体の色味や化粧品を使用した後の色味
香り	化粧品の匂いに関する内容
塗り方	具体的な塗り方に関する説明
テクスチャー	化粧を実際に塗った使用感
印象	「優しい」や「可愛い」といった具体的な印象
評価	商品の評価を含む内容
効果	化粧前後の変化や塗り方による化粧効果の説明
他者の意見	「言われていて」など伝聞が含まれる内容
TPO	時, 場所, 場合に関する言及
リピート	再購入に関する表現が含まれる内容
比較	他の商品との相違点の説明
肌質や色	肌質 (脂性肌など) や皮膚の色 (地黒など)
その他	発話者の近況報告など化粧に関係のない内容

で、(1) レビュー文のみならず化粧動画内の発話が化粧表現の情報源として活用可能であるか、(2) 化粧特有の表現はデータソース間でどのような差があるか、という 2 つの観点を明らかにする。(1) は化粧語彙セット構築における提案法の妥当性に関する観点である。データソースごとにクラスの構成比を観察することで、化粧に関する語彙がどの程度の精度で分類可能かを観察する。(2) はデータソースの違いによって生じるセンテンスラベルごとの語彙特徴に関する観点である。データソースごとのクラスの構成比を観察した上で、バイナリ化されたクラスごとの数値をデータソース間で比較し、各データソースごとに特徴付けられていると推測可能なセンテンスを定性的に観察することで、発話とレビュー文からそれぞれ異なる特徴を持つ情報が抽出されたかどうかについて観察する。

5.1 テストデータの構築

分析にあたり、語彙セット構築に利用しなかったデータからテストデータを構築した。テストデータは、前章で収集したデータのうち、化粧語彙セット構築に利用しなかった発話文 30 動画分および口紅・アイメイク化粧品の各 30 件分のレビュー文を用いた。

テストデータは各データソースのいずれも句読点を基準として分割し、分割した各文にはテキスト内容に基づいて人手で分類ラベルを付与した（以下、センテンスラベルと記す）。作成したセンテンスラベルは、商品、商品の特性、商品の価格、色味、香り、塗り方、テクスチャー、印象、評価、効果、他者の意見、TPO、リピート、比較、肌質や色、その他の計 16 種類とした。ラベル付与の基準を表 2 に示す。これにより、口紅の発話

⁷ http://www.cl.ecei.tohoku.ac.jp/~m-suzuki/jawiki_vector

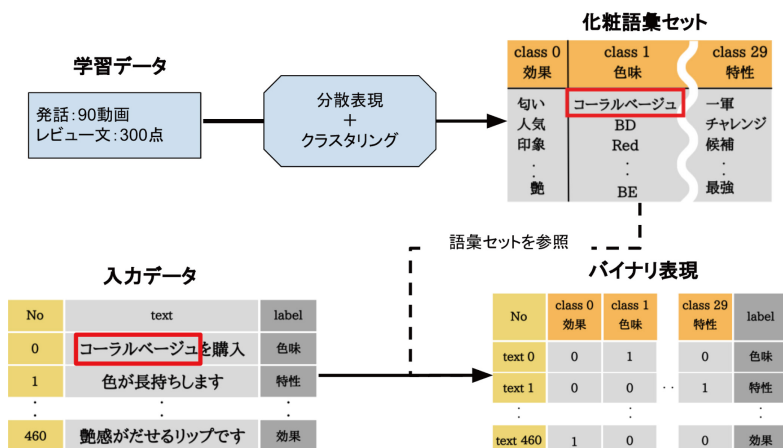


図4 語彙セットを用いたテストデータのバイナリ化

460文、レビュー310文、アイメイクの発話1532文、レビュー235文が得られた。このテストデータに対し、語彙セットと同様に Mecab (Ver.0.996) [20] および mecab-ipadic-NEologd を用いて形態素解析を行った後、名詞、形容詞、動詞の3つを抽出した。これにより得た単語のうち、重複を除くと、口紅の発話文の単語数が835単語、レビューデータの単語が887単語、アイメイクの発話文の単語数が1,818単語、レビューデータの単語が884単語であった。

これらの単語について、4章でクラスタリングされた語彙セットの各クラスの語彙とテストデータに含まれる語彙とを比較する。図4に示すように、テストデータに含まれる語彙が該当するクラス内に存在する場合は「1」、それ以外は「0」とバイナリ化した。以下では、この処理で得られた語彙のベクトル表現をセンテンスベクトルと記す。

5.2 センテンスラベルの構成比率の比較

付与したセンテンスラベルの構成比率をデータソース間で比較し、発話文から取得できる化粧表現がレビュー文から取得可能な化粧表現と差があるかどうかを明らかにする。口紅とアイメイクを合わせた発話文とレビュー文に付与したセンテンスラベルの構成比率を図5に示す。次に、各データソースにおいて多く含まれるセンテンスラベルの構成比率を確認するため、センテンスラベルごとの構成比のうち、上位5%を観察した。その結果、レビュー文に多く含まれるセンテンスラベルは「商品」「商品の特性」「色味」「評価」「効果」であった。発話文に多く含まれるセンテンスラベルは「商品」「商品の特性」「色味」「塗り方」「印象」であった。この結果から、データソース間でのセンテンスラベルの構成に差異が確認された。詳細については、6章で議論する。

次に、データソース間において大きな差が観察される項目を明らかにするため、データソース間のセンテンスラベルごとの絶対差を算出し、該当するセンテンス数と絶対差の比を観察した。分析の際、データ数の多寡による影響を考慮するため、絶対差の合計値と各センテンスラベルの絶対差の比を算出し、下位5%に該当するセンテンスラベルは分析の対象外とした。算出した割合の上位5%を観察したところ、「商品」「塗り方」「印

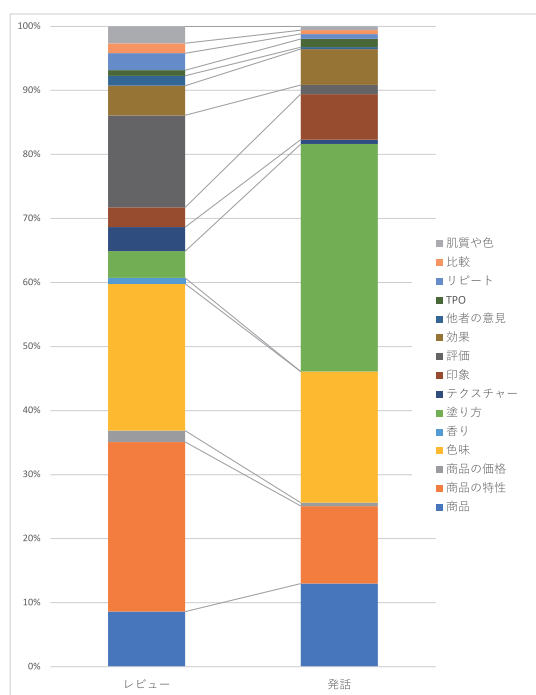


図5 データソース間におけるセンテンスラベルの構成比の比較

象」のセンテンスラベルが共通の特徴として観察された。またレビュー文からは「テクスチャー」「評価」が特徴的な項目であり、発話文からは「効果」「TPO」が特徴的な項目であることが明らかになった。

5.3 テストデータに関する語彙特徴の分析

データソース間による語彙特徴の差異を明らかにすることを目的として、センテンスラベルと各クラスの関係に着目した分析を行った。5.1節で構築したテストデータに対し、各クラスに該当する語彙の割合をデータソースごとに求め、該当する語彙数をセンテンス数で正規化した後にデータソース間の差分を算出した。算出結果は、値が大きい場合ほど該当クラスに属する語彙がレビュー文で用いられる頻度が高く、小さい場合ほど発話文で用いられる割合が高い。このデータに対し、各データ

表3 レビュー文と発話文の語彙特徴の比較 (一部抜粋)

ラベル名	色	数値	生活 シーン	芸能	時期	色
商品の価格	0.0001	4.2519	4.2519	4.2519	1.4009	1.6738
色味	3.9654	4.2541	4.2519	4.1949	3.227	4.2443
塗り方	3.8556	4.2527	4.254	4.2516	3.326	3.6619
印象	2.1118	4.2404	4.2519	4.2542	2.902	4.0776
TPO	0.4977	3.5654	4.2519	3.8658	0.0002	2.4132
比較	4.2519	4.2519	4.2519	4.2519	4.2519	9.1605

太字箇所はレビューと動画内発話の標準偏差から 3σ 区間に該当する項目を示す。

ソースにおいて顕著に特徴が確認される項目を定性的に分析するため、算出結果の標準偏差から 3σ 区間に該当する項目を確認した。該当する項目を表3に示す。この分析の結果、レビュー文では「印象—芸能」「塗り方—生活シーン」「色味—数値」、発話文では「価格—色」「比較—色」「TPO—時期」が抽出された。以下では、該当する項目が抽出されたセンテンスを各データソースごとに観察する。

「印象」ラベルが付与されたレビュー文のセンテンスは14件観察された。そのうち、「芸能」クラスの語彙が抽出されたセンテンスは1件であり、「口紅メインのメイクにすると可愛い!」という文章が確認された。この文章に含まれる「メイン」という語が「芸能」クラスに分類されていたことが影響したと考えられる。

「塗り方」ラベルが付与されたレビュー文のセンテンスは19件観察された。そのうち、「生活シーン」クラスの語彙が抽出されたセンテンスは0件であった。この結果は、データソース間の差分をとったことで、元となるセンテンス数の少ないデータが抽出されたと考えられる。

「色味」ラベルが付与されたレビュー文のセンテンスは104件観察された。そのうち、「数値」クラスの語彙が抽出されたセンテンスは7件であった。該当するセンテンスとして「506: かなり濃いめのピンクですがびっしり青ラメが入っていて角度によって色味が少し変わって見えます」といった文章が確認された。これは化粧品のカラーバリエーションを示す際に使用される番号と、その色に対する言及が影響したと考えられる。

「価格」ラベルが付与された発話文のセンテンスは7件観察された。そのうち、「色1」クラスの語彙が抽出されたセンテンスは3件であった。該当するセンテンスには「990円でゼンイレブンで売ってるやつなんですけど」といった文章が確認された。「色1」クラスに分類された語彙を確認したところ、「パープル」などの色を意味する語彙のうち、カタカナで表記される語彙が集中的に分類されており、「プチプラ」「セブン」などの色以外のカタカナ語が同一のものとして分類されていることが確認された。

「比較」ラベルが付与された発話文のセンテンスは8件観察された。そのうち、「色2」クラスの語彙が抽出されたセンテンスは5件であった。該当するセンテンスには「SUQQUと比べるとSUQQUの方が大人っぽいような色味かなと思うんで

すけど」に示されるような、類似商品との色の比較や、化粧工程の前後での変化を比較する言及が観察された。「比較」ラベルが付与されたレビュー文のセンテンスを同様に確認したところ、別商品との色味の比較は同様に観察されたが、化粧工程による変化についての言及は観察されず、持ち運びのしやすさやつけ心地に関する内容のセンテンスが中心であった。

「TPO」ラベルが付与された発話文のセンテンスは17件観察された。そのうち、「時期」クラスの語彙が抽出されたセンテンスは10件であった。該当するセンテンスには「なのでちょっと前にやった就活メイクと同じ感じなんですけど」といった、利用シーンや年中行事に関する言及が確認された。一方、「今日はねちょっとイエベ春感を前面に押し出したいので」など、パーソナルカラーに関する「春」が時期クラスの語彙として抽出される現象も確認された。

6. 議論

前章では、化粧動画内の発話から化粧表現が獲得可能であること、獲得可能な化粧特有の表現がデータソース間で異なっていることを明らかにするために、(1) レビュー文のみならず化粧動画内の発話が化粧表現の情報源として活用可能であるか、(2) 化粧特有の表現はデータソース間でどのような差があるか、という2つの観点から分析を行った。

(1) の結果として、データソース別に付与されたセンテンスラベルの構成比の比較から、対象とするデータソースによって取得できる内容に偏りがあるものの、化粧動画内の発話から化粧表現が取得できることが確認できた。これは、化粧に関する検索のためのデータソースとして、化粧動画を利用することの意義を示すものと言える。化粧動画コンテンツの検索機能の向上を図る上で、「どのような化粧品をどのように利用すれば所期の化粧が可能になるか」というユーザーの関心に応えるためには、利用される化粧品の特性に加えて、塗り方などの化粧品の利用方法を把握しておく必要がある。今回得られた知見は、その判断に活用する知識の構築にあたり、化粧動画の発話文とレビュー文とをその特徴の下で組み合わせるべきであることを示唆している。

(2) の結果として、データソース別に付与されたセンテンスラベルの構成比の比較から、レビュー文からは「評価」「効果」、発話文からは「塗り方」「印象」に関するデータが多く含まれることが明らかになった。またレビュー文からは「テクスチャー」「評価」が、発話文からは「効果」「TPO」に関する内容が、データソースごとの特徴的な情報として獲得できることがわかった。またセンテンスラベルと語彙クラスの関係の比較から、人手でセンテンスラベルを付与した発話文およびレビュー文と化粧語彙セットの各クラスの関係が明らかになった。「比較」ラベルにおけるデータソース間の比較の結果からは、同一のセンテンスラベルにおいてもデータソースによって獲得可能な化粧の知識が異なることが示唆された。この結果は、単一のデータソースに基づく従来の化粧辞書では得られなかったデータの獲得が可能であることを示しており、(1) において示されたデータソースの併用の効果を支持するものと考えられる。

本稿における限界として、化粧動画における語彙の獲得量、

および語彙のクラスタリングにおけるラベル名との不一致がある。語彙の獲得量に関しては、図2に示されるように異なり語彙数の収斂が確認されず、また獲得された語彙数もレビュー文と比較して少量であった。この原因として、化粧動画におけるデータの収集工程がレビュー文の収集工程と比較して複雑であることが考えられる。動画の字幕保存の自動化や分析対象の工程の範囲の制約を緩和することで、化粧動画の内容確認の工程の効率化を図るといった対応が必要である。語彙のクラスタリングにおけるラベル名の不一致に関しては、(2)の分析において示されたように、「色」ラベルにおけるカタカナ語の混同やパーソナルカラーを示す「春」が「時期」クラスに割り当てられるなどの現象が該当する。そのため、クラスタリングの精度の向上やラベルの精度の向上について今後の課題とする。また、化粧に関連する語彙のデータソースとしては、今回対象とした動画内の発話の他に、ショート動画や画像投稿サイトで利用されるハッシュタグなどが存在する。これらはユーザの検索時に利用されるものであり、語彙セットのデータソースとしての有用性が期待できる。これらのデータソースを対象とした語彙セットの拡張についても今後の検討事項とする。

7. おわりに

本研究は、化粧動画内における特定の工程を検索可能にする枠組みの実現を目指し、化粧辞書の拡張のための語彙分析を行った。本稿では、従来研究で用いられてきたレビュー文に加え、化粧に関する手法を紹介した動画の動画内発話をデータソースとして併用することを提案した。この提案の妥当性を検証するため、レビュー文と発話をデータソースとした語彙セットを構築した。化粧動画内の発話から化粧表現が獲得可能であることと、獲得可能な化粧特有の表現がデータソース間で異なっていることを明らかにすることを企図して、(1)レビュー文のみならず化粧動画内の発話が化粧表現の情報源として活用可能であるか、(2)化粧特有の表現はデータソース間でどのような差があるか、という2つの観点から分析を行った。

(1)の結果として、データソース別に付与されたセンテンスラベルの構成比の比較から、データソース間で得られる情報が異なることが明らかになった。また「比較」ラベルにおけるデータソース間での定性的な観察の結果から、共通するラベルが付与されたセンテンスであっても、データソース間で異なる意味を含む情報が観察された。この結果から、レビュー文と発話文のデータソースを併用することで、化粧動画の特定のシーンに合致する参考情報を提示することが可能になることが明らかになった。

(2)の結果として、データソース別に付与されたセンテンスラベルの構成比の比較から、レビュー文からは「評価」「効果」、発話文からは「塗り方」「印象」に関するデータが多く含まれることが明らかになった。センテンスラベルと語彙クラスの関係の比較からは、データソース間で異なるラベルクラス対が抽出された。「比較」ラベルにおけるデータソース間での定性的な観察からは、化粧工程の前後での変化を比較する言及など、データソース特有の言及が確認された。この結果は、各データソースから獲得される化粧表現が大きく異なること、またデー

タソース間で共通のラベルを持つ情報であっても獲得される化粧表現の内容が異なり、各データソース特有の情報が獲得可能であることが明らかになった。

謝辞

この研究の一部は2021年度国立情報学研究所公募型共同研究(21S0501)の助成によるものである。記して謝意を表す。

参考文献

- [1] 鈴森正幸: “人はなぜ化粧をするのか,” 日本化粧品学会誌, Vol.42, No.1, pp. 27-35, 2018.
- [2] 岡田明子, 芳住邦雄: “女子学生の化粧に対する意識と行動,” 繊維製品消費科学, Vol.47, No.11, pp. 652-660, 2006.
- [3] 末岡奈々, 村上千代: “顔の対象性からみたイメージと化粧に関する研究,” 日本家政学会研究発表要旨集, Vol.47, p. 116, 2017.
- [4] R. Russell: “Why Cosmetics Work,” in *The Science of Social Vision*, R. B. Adams, Jr., N. Ambady, K. Nakayama, and S. Shimojo eds., Oxford University Press, pp. 186-203, 2010.
- [5] 松波友稀, 上田真由美, 中島伸介: “各コスメアイテム分類に対する評価表現辞書構築方法に関する考察,” 第10回データ工学と情報マネジメントに関するフォーラム, pp. D1-3, 2018.
- [6] 酒井美春, 上田真由美, 松下光範: “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, pp. B6-2, 2019.
- [7] M. Ueda, Y. Taniguchi, D. Li, P. Siriaraya, and S. Nakajima: “A Research on Constructing Evaluative Expression Dictionaries for Cosmetics Based on Word2Vec,” *The 23rd Int. Conf. on Information Integration and Web Intelligence*, pp. 80-86, 2021.
- [8] 土方嘉徳: ソーシャルメディア論—行動データが解き明かす人間社会と心理—, サイエンス社, 2020.
- [9] 末吉優, 関洋平: “音楽のジャンルと印象を用いた VOCALOID クリエータの検索,” 人工知能学会論文誌, Vol.32, No.1, pp. WII-K_1-12, 2017.
- [10] 段玉鋒, 佐賀亮介: “画像形状特徴を用いたアパレル商品向け推薦システム,” 電気学会論文誌 C, Vol.140, No.1, pp. 93-99, 2020.
- [11] B. A. Rabut, A. C. Fajardo, and R. P. Medina: “Multi-Class Document Classification Using Improved Word Embeddings,” *Proc. of the 2nd Int. Conf. on Computing and Big Data*, pp. 42-46, 2019.
- [12] G. Magno and V. Almeida: “Measuring International Online Human Values with Word Embeddings,” *ACM Trans. on the Web*, Vol.16, No.2, pp. 1-38, 2021.
- [13] S.-T. Park and C. Liu: “A Study on Topic Models Using LDA and Word2Vec in Travel Route Recommendation: Focus on Convergence Travel and Tours Reviews,” *Personal and Ubiquitous Computing*, Vol.26, No.2, pp. 429-445, 2022.
- [14] D. Paul, S. Sarkar, M. Chelliah, C. Kalyan, and P. P. Sinai Nadkarni: “Recommendation of High Quality Representative Reviews in e-commerce,” *Proc. of the 11th ACM Conf. on Recommender Systems*, pp. 311-315, 2017.
- [15] Y. Han, R. Yamanishi, and Y. Nishihara: “Music Retrieval Focusing on Lyrics with Summary of Tourist-Spot Reviews Based on Shared Word-Vectors,” *2020 Int. Conf. on Technologies and Applications of Artificial Intelligence*, pp. 73-78, 2020.
- [16] 山西良典, 西原陽子, 松下光範: “語彙の標準化と量子化によるあらゆる特性表現に関する基礎検討,” 電子情報通信学会第7回コミック工学研究会, pp. 31-38, 2022.
- [17] 徐統非, 肥合智史, 嶋田和孝: “日本語 wikipedia エンティティベクトルを外部知識を利用した皮肉検出,” 電子情報通信学会技術報告, Vol.120, No.374, pp. 25-30, 2021.
- [18] M. Z. Hossain, M. N. Akhtar, R. B. Ahmad, and M. Rahman: “A dynamic k-means clustering for data mining,” *Indonesian J. of Electrical Engineering and Computer Science*, Vol.13, No.2, pp. 521-526, 2019.

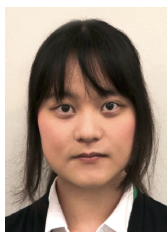
- [19] M. Cui: “Introduction to the K-Means Clustering Algorithm Based on the Elbow Method,” *Geoscience and Remote Sensing*, Vol.3, No.9, pp. 9-16, 2020.
- [20] T. Kudo, K. Yamamoto, and Y. Matsumoto: “Applying Conditional Random Fields to Japanese Morphological Analysis,” *Proc. of the 2004 Conf. on Empirical Methods in Natural Language Processing*, pp. 230-237, 2004.

(2022年8月29日 受付)
(2023年2月16日 採録)

[問い合わせ先]

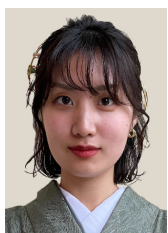
〒569-1095 大阪府高槻市霊仙寺町 2-1-1
関西大学 大学院総合情報学研究科
安尾 萌
TEL: 072-690-2437
E-mail: k290993@kansai-u.ac.jp

著者紹介



やすお めぐみ
安尾 萌 [非会員]

2016年関西大学総合情報学部卒業, 2018年関西大学大学院総合情報学研究科博士課程前記課程修了, 関西大学大学院総合情報学研究科博士課程後期課程入学. 協調的情報トリアージに関する研究に従事. 2017年 ACIS2017 Best Paper Award 受賞.



このの ゆきの
河野 雪乃 [非会員]

2022年関西大学総合情報学部卒業. 化粧コンテンツの特徴分析に関する研究に従事.



やまにし りょうすけ
山西 良典 [非会員]

2007年名古屋工業大学工学部知能情報システム学科卒業. 2009年同大学院工学研究科情報工学専攻博士前期課程修了. 2012年同博士後期課程修了. 博士(工学). 2012年立命館大学情報理工学部助手, 2013年同特任助教. 2014年同助教, 2018年同講師. この間, プリティッシュコロロンビア大学(カナダ)客員助教. 2020年関西大学総合情報学部准教授, 現在に至る. 関西大学先端科学技術推進機構兼務. 感性情報処理, Web インテリジェンス, マルチメディア情報処理, 音楽情報処理, 言語情報処理などの技術に興味を持ち, 音楽情報処理やコミック工学, 食メディア処理, ゲーム情報学などの文化・芸術コンテンツを対象とした計算機科学研究に従事. 電子情報通信学会, 情報処理学会, 人工知能学会, 日本感性工学会, 日本音響学会, 芸術科学会, ACM, ACL, 各会員.



まつした みつりの
松下 光範 [非会員]

1995年大阪大学大学院基礎工学研究科物理系専攻制御工学分野博士前期課程修了. 同年日本電信電話株式会社入社. 2008年関西大学総合情報学部准教授. 2010年同教授. 自然言語理解, インタラクションデザインに関する研究に従事. 博士(工学). 2003年情報処理学会論文賞, 2013年 LavalVirtual Award ほか各賞受賞. 電子情報通信学会, 情報処理学会, 人工知能学会, 芸術科学会, ACM 各会員.

Construction of Makeup Vocabulary Datasets for Scene Search on Makeup Movies

by

Megumi YASUO, Yukino KONO, Ryosuke YAMANISHI, and Mitsunori MATSUSHITA

Abstract:

This study aims to acquire the vocabulary of cosmetics and makeup techniques used for searching for an intended scene from makeup movies. When applying cosmetics, it is necessary to appropriately select a set of cosmetics by considering individual preferences, affective senses, and physical constitutions. Makeup movies are gaining popularity as reference information for applying makeup and are becoming as crucial as traditional information media. Currently, people retrieve makeup videos on a video-by-video basis, and methods for retrieving specific processes in makeup movies have not yet been developed. As the first step in developing such a system, this paper conducted a fundamental analysis of vocabulary collected from cosmetic reviews and speech in makeup movies. The characteristics of each modal were studied, focusing on the semantics of sentences and word categories. As a result, each data source contains different types of cosmetic vocabulary. This result implies that using each data source in a combining manner to build a vocabulary is a rational way to expand the vocabulary efficiently.

Keywords: vocabulary analysis, makeup, review analysis, speech in movies

Contact Address: **Megumi YASUO**
Graduate School of Informatics, Kansai University
2-1-1 Ryozenji-cho, Takatsuki-shi, Osaka 569-1095, Japan
TEL: +81-72-690-2437
E-mail: k290993@kansai-u.ac.jp