

# コスメ向けに学習した Word2Vec に基づく評価表現辞書構築手法

谷口 祐奈<sup>†</sup> 奥田 麻美<sup>††</sup> 上田 真由美<sup>†††</sup> Panote Siriaraya<sup>††††</sup> 中島 伸介<sup>†††††</sup>

<sup>†</sup> 京都産業大学 コンピュータ理工学部 〒603-8555 京都府京都市北区上賀茂本山

<sup>††</sup> 京都産業大学大学院 先端情報学研究科 〒603-8555 京都府京都市北区上賀茂本山

<sup>†††</sup> 流通科学大学 経済学部 〒651-2188 兵庫県神戸市西区学園西町 3-1

<sup>††††</sup> 京都工芸繊維大学 〒606-8585 京都府京都市左京区松ヶ崎橋上町

<sup>†††††</sup> 京都産業大学 情報理工学部 〒603-8555 京都府京都市 北区上賀茂本山

E-mail: <sup>†</sup>g1644763@cc.kyoto-su.ac.jp, <sup>††</sup>{i1888033,i1788223}@cc.kyoto-su.ac.jp,

<sup>†††</sup>Mayumi\_Ueda@red.umds.ac.jp, <sup>††††</sup>spanote@gmail.com, <sup>†††††</sup>nakajima@cc.kyoto-su.ac.jp

あらまし 近年インターネット上には様々な口コミ・レビューサイトが存在している。オンラインショッピングにおいて、口コミ・レビューサイトは消費者の購入意思決定に強く影響を与えるため、欠かせないものである。我々は、コスメレビューに着目し、個々の利用者の肌質や使用感を考慮した情報推薦を実現するため、各コスメアイテム分類毎の評価表現辞書を用いた自動スコアリングシステムによるレビュー推薦手法の開発に取り組んでいる。コスメアイテム分類は細かく分類されるため、全てのコスメアイテムに対応した自動スコアリングを行うには、全アイテム分類毎の評価表現辞書が必要である、そのため、評価表現辞書の構築手法は効率的で、自動化できることが望ましい。本稿では、コスメ向けに学習した Word2Vec を用いた評価表現辞書の構築手法を用いて、評価表現辞書の精度の向上と効率的な開発手法の検討を行う。

キーワード コスメアイテム, レビュー分析, クチコミ分析, レビュー推薦, 評価表現辞書

## 1. はじめに

近年、日本国内のインターネット市場、スマホ EC 市場は拡大傾向にある。商品の実物を見て購入を決定できないそれらの市場では、消費者の購入意思決定において、レビューの影響力は大きい[1]。特にコスメアイテムは、直接肌に使用するのであり、自分の肌に合わない商品を使用すると肌トラブルを起こすことがあるため、レビューサイトを利用することの意義は大きいといえる。しかし、既存のコスメアイテムに対するレビューサイトでは、各コスメアイテムに対する評価は総合的な評価値が付与されているのみで、嗜好や使用感に関する詳細な評価を確認するためにはレビュー文を確認する必要があり、価値観を共有しうるか否かを直感的に判断することは困難である。さらに個人によって肌質や嗜好が異なることから、参考となるレビューは異なり、膨大なレビューの中から個々のユーザーにあったレビューを見つけ出すことは容易ではない。そこで、我々は、ユーザーの好みや価値観に合ったレビューを優先的に提示することが可能なシステムの構築を目指している[3][4]。まず第一段階として[3]では、既存のコスメに関するレビューサイトで用いられているコスメアイテムの効果を表現する項目を参考にして、“化粧水”に対する評価項目を設定し、レビュー文から“うるおい効果”、“美白効果”、“毛穴・角質ケア効果”といった各評価項目に対する評価値を算出する自動スコアリング方式の開発に取り組んできた。また、コスメアイテムには、化粧水や口紅など、アイテムごとに用途が違い、それらの効果を表現するための語はアイテムの分類ごとに異なる。そのため、全てのアイテムに対応した自動スコアリングシステムを開発す

るためには、アイテムの分類ごとに評価表現辞書が必要になる。しかし、コスメアイテムの分類<sup>(注1)</sup>は表1のように、スキンケア・メイクアップ・ベースメイクに関する分類内に限定しても、17種類もの分類が存在することから[5]、すべての評価表現辞書を手動で構築することは手間がかかる。我々は、全てのコスメアイテム分類に対応可能なレビュー自動スコアリングシステムの構築を最終的な目標としている。そのため、構築手法は人手ではなく、自動化できることが、コストにおいても、評価表現辞書を用いた自動スコアリングの精度においても望ましい。

これまで、化粧水など既に構築している評価表現辞書の再利用や、辞書に登録する評価表現の半自動抽出などを行うことで、評価表現辞書の効率的な構築方法の検討を行ってきた。本稿では、さらなる評価表現辞書の構築手法の効率化、自動化を目指し、コスメレビューを学習した Word2Vec を用いた単語間の類似度を、評価表現辞書の構築に用いた、構築手法を提案する。

表1 コスメアイテム分類

| スキンケアアイテム    | メイクアップアイテム     | ベースメイクアイテム  |
|--------------|----------------|-------------|
| 洗顔料          | アイブロウ          | ファンデーション    |
| 乳液・美容液       | 口紅・グロス・リップライナー | 化粧下地・コンシーラー |
| その他スキンケア     | アイライナー         | フェイスパウダー    |
| クレンジング       | アイシャドウ         |             |
| パック・フェイスマスク  | チーク            |             |
| 化粧水          | マスカラ           |             |
| 目元・口元スペシャルケア | ネイル・ネイルケア      |             |
|              | ベースメイクアイテム     |             |

(注1)：コスメ・美容に関する代表的なサイトである@cosme[2]で用いられている分類

## 2. 関連研究

商品レビューを扱うサイトとしては、Amazon.com [6] や価格.com [7]、食べログなどが有名である。また、コスメアイテムに関しては、@cosme が有名である。@cosme の運営会社によると、2019 年 3 月時点で月間 3.1 億ページビュー、メンバー数が 540 万人、総レビュー数が 1500 万件と報告されており [8]、多くのユーザがコスメ・美容に関する情報をやり取りしていることがわかる。ユーザは、レビュー投稿時に総合評価として 7 段階で評価しており、また、会員登録の際に年齢と肌質を記録するカルテ機能があり、自分の年齢や肌質が同じユーザによるレビューを検索するなど、目的に応じてレビューを検索することが可能である。

レビュー情報とユーザの意思決定に関する研究がいくつか行われている。中邨らの研究では、化粧品サイト@cosme と、コスメ・化粧品の通販サイト COSME.COM を対象とし口コミがどの程度売り上げに影響を与えるのかを分析している [12]。この研究から、口コミが多い場合は売り上げ先行型、値段が高いものは口コミ先行型の傾向にあるという結果を得ている。山田らの研究では、口コミに含まれる感情情報を抽出し可視化した口コミ感情可視化システムを提案している [13]。このシステムにより、ユーザが従来のマップよりもユーザが訪問したいと思えるスポットを選択できることを示している。

また、レビュー分析を行う研究が数多く行われている。アンケートの自由記述文などの自由回答欄で言及されている膨大なテキストの中から評判に関する発言を抽出して、書き手の意図を扱う研究 [14] や、商品レビューから語彙構文パターンを使って購買意図の抽出を行う研究 [15] が取り組まれている。入江らはレビュー情報をもとにレストラン利用者の評価観点に焦点を当て、共通の評価観点でレビューされることが多い店舗を、違いの類似競合店舗として推薦する推薦方法を提案している [19]。吉田らは Paragraph Vector モデルをユーザレビューに用い、生成した特徴ベクトルとユーザの入力した評価キーワードを用いた主観的特徴を考慮した観光スポット検索システムの提案をしている [20]。Pham らは、least square method を用いてホテルのレビューを対象とし、評価視点毎にスコアを推測する手法と最も重要な評価視点を推測する手法を提案している [22]。この手法では、レビュー内に評価視点に関する内容がない場合でも全ての評価視点に対する評価点を常に推測している。

以上の通り、コスメおよびその他のアイテムに対するレビュー分析を行う研究が数多く行われているが、本研究にて実施している、コスメアイテムに特化した評価表現辞書の構築手法に関する研究は行われていない。

## 3. コスメアイテムに関するレビュー自動スコアリング方式

### 3.1 評価項目別自動スコアリング方式の概要

従来のコスメレビューサイトでは、各アイテムに対するスコア（星の数）が付与されていることが多いが、基本的には対象アイテムに対する総合評価であり、レビュー投稿者の使用感を

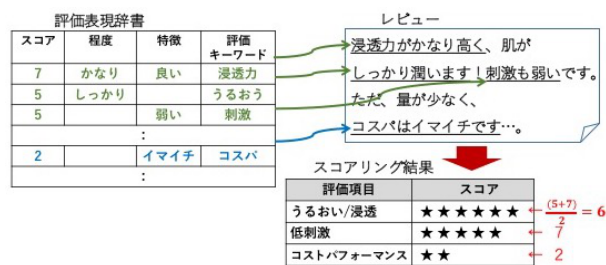


図1 未知データに対する自動スコアリング処理の手順

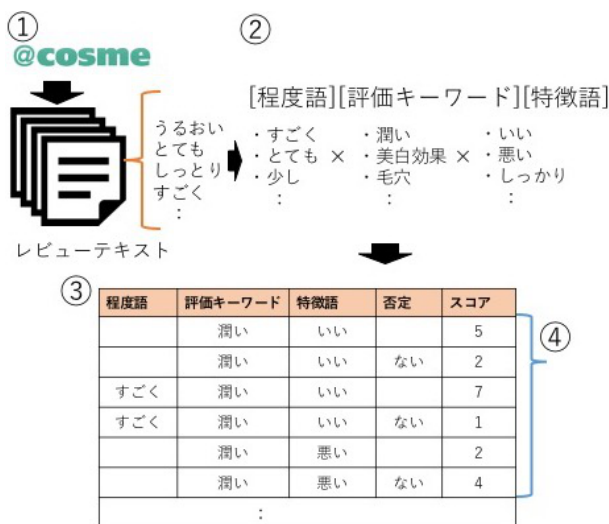


図2 評価表現辞書の構築手順

判断することは難しい。例えば、「化粧水」に対する評価項目としては、「うるおい効果」、「美白効果」、「毛穴・角質ケア効果」、「低刺激」、「エイジングケア効果」等が考えられる。したがって、レビュー投稿者の使用感を分析するために、これらのような評価項目別のスコアを算出する手法を開発した。

コスメレビューに対する評価項目別自動スコアリングは、図1に示すように、自然言語で記述されたレビューテキスト内に含まれる評価表現に対し、これに合致する評価表現辞書を用いてスコアリングすることで実現している。手順としては、レビューテキストに形態素解析をかけて単語に分割し、各単語が評価表現辞書に含まれるかどうかを調べる。その上で各コスメアイテム分類の評価項目に関連の深い評価キーワード、特徴語、程度語の組み合わせから、評価表現辞書に示されるスコアを付与する。図1の例では、「かなり」「良い」「浸透力」と「しっかり」「うるおう」が同じ文に含まれることから、評価項目「うるおい/浸透」に対して「7」および「5」のスコアが付与されるが、これらの平均である「6」がこのレビューの評価項目「うるおい/浸透」に対するスコアとなる。この評価項目別自動スコアリングでは、各コスメアイテム分類に対する評価表現辞書における表現の登録数および設定スコアの精度が、非常に重要である。

### 3.2 各コスメアイテム分類に対する評価表現辞書構築手法

質の高い評価表現辞書の構築を実現するため、我々は実際にレビュー投稿サイトに投稿されたデータから、その効果を表す

特徴的な表現を収集し、スコアを付与することによって、評価表現辞書を構築している。

一般的に、アイテムの効果を表すフレーズを辞書に登録すると、「香りがすごくいい」と「すごく香りがいい」といった類似した表現を辞書に登録することとなり、冗長である。そこで我々は、評価キーワードの共起に基づく評価表現辞書を構築することで、冗長性を除去し、「香りがすごくいい」と「すごく香りがいい」といった表現を同一のものとして扱うことを実現した。

評価表現辞書の構築手順（図2参照）は、以下の通りである。

(1) 既存のレビューサイトからレビューテキストを収集し、アイテムの効果を表すフレーズを抽出。(2)(1)で抽出したフレーズから、評価キーワード・特徴語・程度語のリストを作成。(3)(2)で作成した各リストを用いて、特徴語・程度語・評価キーワード・否定の4つの語の組み合わせを評価表現として作成。(4)作成した評価表現に対して人手によりスコアを付与。上記の通り、先行研究[3]にて「化粧水」に対する評価表現辞書および、これを用いた自動スコアリングのプロトタイプシステムを既に開発している。ここで構築した「化粧水」に対する評価表現辞書には、1332種類の評価表現を格納しており、約81%の精度で自動スコアリングを行うことができています。このように、ある程度の精度を達成しているが、これは「化粧水」に関する評価表現辞書を人手で丁寧にメンテナンスしたことによるものである。今後は「化粧水」以外のコスメアイテム分類に対する評価表現辞書の構築が必要であるが、コスメアイテム分類の数は多く、全ての辞書を人手により構築するのはコストが大きく現実的とはいえない。したがって、化粧水など既に構築しているコスメアイテム分類に対する評価表現辞書を可能な限り再利用するなどして、他のコスメアイテム分類に対する評価表現辞書を効率的に構築する手法を検討することの意義は大きい。

3.3 評価表現辞書構築手法の効率化

従来では、評価表現辞書は3.2節の手法により、全て人手で構築していたが、全てのコスメアイテム分類分の評価表現辞書を手動で構築することは、手間がかかりすぎる点と、辞書構築者の主観的な判定やスコア付けによる辞書になってしまい、評価表現辞書を用いた自動スコアリングの精度の面で望ましくない。そのため、化粧水など既に構築している評価表現辞書の再利用や、辞書に登録する評価表現の半自動抽出などを行うことで、他のコスメアイテム分類に対する評価表現辞書の効率的な構築方法の検討を行ってきた。効率化した手法を交えた、構築手法（図3参照）は以下の通りである。(1) コスメレビュー文を形態素解析し、名詞のtf-idfの上位単語を評価キーワード候補とし、その中から評価キーワードを抽出する。(2)(1)で抽出した評価キーワードが該当する評価項目を判定し、登録する。その際評価キーワードを含む評価表現から、特徴語、程度語、否定の有無を登録する。(3) 作成した評価表現に対して人手でスコアを付与する。このようにtf-idfを用いて評価キーワードを抽出する方法は人手で評価表現を見つける手法に比べコスト縮小につながる。また、従来では、一つの評価表現に対して一つの評価項目のみ登録していたが、一つの評価表現に、該当す



図 3 評価表現辞書構築手法の効率化

るだけ評価項目を登録することで効率的に辞書の登録評価表現数を増やせるようになり、評価率向上に繋がった。この手法を用いて、既存の化粧水の辞書に加えて、リップ、乳液、ファンデーション、チークの辞書を構築した。しかし、まだ人手で構築している部分が多いため、本研究ではさらなる効率化、半自動化した評価表現辞書構築手法を検討している。

4. Word2Vecを用いた評価キーワードが属する評価項目の自動判定手法の検討

4.1 提案手法の概要

本節では提案手法である Word2Vec を用いた評価キーワードが属する評価項目の自動判定手法について説明する。従来はアイテム分類から特徴的な評価キーワード群を抽出し、それらが該当する評価項目を人手で判定し、登録していた。例えば、評価キーワード「ハリ」が評価項目「潤い」に該当すると判断した場合は、「潤い」に登録し、「エイジングケア」に該当すると判断した場合は「エイジングケア」に登録する、このように人手で判定した場合、辞書作成者の主観に偏った辞書となってしまい、評価表現を用いた自動スコアリングの際の精度が低くなる可能性がある。また全ての評価キーワードに対し、人手で判定するのは手間がかかる。そこで、本手法では Word2Vec を用いた単語間の類似度を用いて評価キーワードが各評価項目に該当するかどうかの判定を半自動的に行う。評価項目と評価キーワードの類似度が高ければ、関連が高いと判定し、その評価項目に振り分ける。例えば、評価キーワード「乾燥」がどの評価項目に該当するか Word2Vec を用いて判定したとする。表2の示すように、「乾燥」と各評価項目ごとに類似度を算出する。この場合、「乾燥」と類似度の高い、「潤い」と「肌荒れ対策」を適切な評価項目だと判定する。この手順を全評価キーワードに対

表 2 評価キーワードが属する評価項目の自動判定手法

| 評価キーワード | 評価項目    | 評価キーワードと評価項目の類似度 |
|---------|---------|------------------|
| 乾燥      | 肌荒れ対策   | 0.79             |
| 乾燥      | 潤い      | 0.52             |
| 乾燥      | 香り      | 0.1              |
| 乾燥      | 低刺激     | 0.2              |
| 乾燥      | エイジングケア | 0.15             |
| 乾燥      | コスバ     | -0.1             |

して行う。提案手法により評価表現辞書構築手法のコストの縮小、半自動化による辞書精度の向上が望める。



| wiki |      | cosmeレビュー2000件 |      | cosmeレビュー10000件 |      |
|------|------|----------------|------|-----------------|------|
|      | 類似度  |                | 類似度  |                 | 類似度  |
| 安らぎ  | 0.59 | 向い             | 0.99 | しっとり感           | 0.76 |
| 活力   | 0.59 | 乾燥肌            | 0.99 | 潤っ              | 0.73 |
| 手厚い  | 0.52 | 物足りない          | 0.99 | うるおい            | 0.72 |
| 安心感  | 0.52 | とろみ            | 0.98 | 潤う              | 0.7  |
| 清潔   | 0.5  | 低              | 0.99 | べたつき            | 0.68 |
| 富    | 0.49 | でも             | 0.98 | ベタ              | 0.67 |
| 豊か   | 0.49 | 強め             | 0.98 | 保湿効果            | 0.66 |
| 活気   | 0.49 | 保湿効果           | 0.98 | 保湿力             | 0.66 |
| 住環境  | 0.48 | 嫌い             | 0.98 | 保湿感             | 0.66 |

図 4 “潤い” と類似度の高い上位単語の比較

## 4.2 Word2Vec の学習に用いるコーパスにコスメレビューを用いることの有効性検証

4.1 節の提案手法では、コスメレビューを学習した Word2Vec を用いて類似度を算出する。学習に用いるテキストコーパスを一般的な wikipedia の本文を用いた場合とコスメレビューを用いた場合で比較し、提案手法の Word2Vec の学習にコスメレビューを適用することの有効性を検証する。学習に用いたコーパスは wikipedia の本文 2GB 以上、コスメレビュー 2000 件、コスメレビュー 10000 件の 3 つである。各テキストコーパスを Word2Vec に学習させ、a. 各手法における評価項目“潤い”に対する類似度上位単語の比較、b. 各手法における評価項目と評価キーワードの類似判定結果の比較、以上 2 つの比較方法にて検証を行う。

### 4.2.1 各手法における評価項目“潤い”に対する類似度上位単語の比較

4.2 節の a の比較では、評価項目“潤い”と類似度の高い単語を取得し比較した。比較結果は図 4 の示すようになり、青字はコスメ特有単語である。コスメ特有単語とはコスメを評価するにあたり、頻出する単語である。コスメレビューを学習した Word2Vec では“潤い”という単語に対し、コスメ特有単語を多く取得できた。提案手法において、コスメ特有単語が多く抽出できているという結果は大変良い結果であるといえる。また、学習するレビュー数を増やすほど、“潤い”と類似度の高い単語のコスメ特有単語の占める割合が大きくなった。今後学習するレビュー数を増やす意義は大きいといえる。

### 4.2.2 各手法における評価項目と評価キーワードの類似判定結果の比較

4.2 節の b の比較では、評価項目と評価キーワード間の関連性の高い低い、類似度の大きい、小さいに現れるかを検証する。図 5 の示す通り、“潤い”と“香り”という評価項目に関して、各評価項目と関係性が高い評価キーワードと関連性の低い評価キーワードとの類似度を算出し、各コーパス間で比較した。赤字で書かれた評価項目と評価キーワード間で類似度が大きくなり、青字で書かれた評価項目と評価キーワード間で類似度が小さくなるのが望ましい。比較結果は、コスメレビューを 10000 件学習したモデルのみが関連性の高い評価項目と評価キーワード間で類似度が大きくなり、関連性の低い評価項目と評価キーワード間で類似度が小さくなった。4.2.1 節、本節の比較より、提案手法の Word2Vec の学習コーパスにコスメレビューを適用することの有効であることがわかる。

検証結果

|              | wiki  | cosmeレビュー2000件 | cosmeレビュー10000件 |
|--------------|-------|----------------|-----------------|
| “潤い”と“保湿”    | 0.23  | 0.94           | 0.49            |
| “潤い”と“乾燥”    | 0.12  | 0.55           | 0.32            |
| “潤い”と“もちもち”  | 0.16  | 0.77           | 0.48            |
| “潤い”と“コスパ”   | -0.04 | 0.83           | 0.16            |
| “潤い”と“香り”    | 0.28  | 0.79           | 0.05            |
| “香り”と“アルコール” | 0.48  | 0.93           | 0.58            |
| “香り”と“潤い”    | 0.28  | 0.79           | 0.059           |
| “香り”と“肌荒れ”   | 0.28  | 0.21           | -0.148          |

赤字: 関連性の高いペア 青字: 関連性の低いペア

図 5 各手法における評価項目と評価キーワードの類似判定結果の比較

## 4.3 提案手法の評価

本節では、本稿の提案手法であるコスメ向けに学習した Word2Vec を用いた評価キーワードが属する評価項目の自動判定手法を用いて振り分けた各評価項目に対する評価キーワードの組合せと、既存の乳液辞書の評価キーワードを用いて実験を行い作成した正解となる人手による評価項目と評価キーワードの組合せを比較して提案手法の有効性を評価する。

### 4.3.1 人手による各評価項目と評価キーワードの組合せの作成と評価方法

評価に用いた正解となる人手による各評価項目に対する評価キーワードの組合せの作成方法について説明する。既存の乳液辞書から各評価項目の評価キーワードを最大 5 個ずつ、なければあるだけ抽出し、それら計 43 個の評価キーワードをランダムに並べ語群とし、被験者 5 名に語群の評価キーワードを関連性の高いと判断した評価項目に人手で振り分けてもらった。その際、被験者が 2 つ以上の評価項目に関連すると判断した場合は、該当する全ての評価項目に評価キーワードを振り分けてもらった。被験者 5 名による人手による評価キーワード振り分けの結果を集計し、5 人中 5 人または 5 人中 4 人が選択した評価項目に対する評価キーワードのペアを正解のペアとした。評価キーワードの中には 5 人で結果が分かれ、5 人中 3 人が選択したペアが最大になる場合があるが、その場合は 3 人が選択した評価項目と評価キーワードのペアを正解のペアとした。この手順で正解となる評価項目と評価キーワード組合せを作成した(図 6)。

評価方法は作成した正解となる評価項目と評価キーワードの組合せ(図 6)を用い、実験で用いた評価キーワード 43 個と各評価項目に対する類似度を Word2Vec で求め、提案手法で適切な評価項目と判定された評価項目と評価キーワードの組合せが正解となる人手による評価項目と評価キーワードの組合せと、どの程度一致するか評価した。例えば、図 6 の人手による評価項目と評価キーワードの組合せの“乾燥”という評価キーワードが「潤い」と「肌荒れ対策」という評価項目に属しているため、各評価項目に対する“乾燥”との Word2Vec による類似度の上位 2 つが、「潤い」と「肌荒れ対策」だった場合、提案手法は有効だと分かる。

本来の正解ペアである人手による評価項目と評価キーワードの組合せと、提案手法にて適切だと判定された評価項目と評価キーワードの組合せの一致率を算出し提案手法の有効性を検証する。

| 評価項目    | 香り   | 潤い                 | 美白/UV                | 毛穴・角質ケア/高保湿/クレンジング | 爽快感/顔のテカリ防止 | サラサラな肌トロッロ           | 低刺激              | 肌荒れ対策                     | エイジングケア     | コストパフォーマンス  |
|---------|------|--------------------|----------------------|--------------------|-------------|----------------------|------------------|---------------------------|-------------|-------------|
| 評価キーワード | 香り匂い | 乾燥保湿度心地もちり潤いハリツヤ水分 | シミUV紫外線トーンアップ美白日焼け止め | 吹き出物油分角質テカリ毛穴      | 油分ベタ        | とろっベタ心地伸び油分角質使用感使い心地 | 成分自然心地刺激敏感肌ビリッ敏感 | 乾燥肌荒れニキビセラミドスキンケア敏感肌ビリッ敏感 | シミハリツヤエイジング | コスパブチブラ値段価格 |

図 6 人手による評価項目と評価キーワードの組合せ（本来の正解ペア）

| 評価キーワード<br>評価項目 | 乾燥    | 保湿度   | 心地    | もちり   | 潤い    | ハリ    | ツヤ    |
|-----------------|-------|-------|-------|-------|-------|-------|-------|
| 香り              | -0.03 | 0.16  | 0.40  | 0.14  | 0.10  | -0.16 | -0.09 |
| 潤い              | 0.34  | 0.64  | 0.39  | 0.56  | 1.00  | 0.28  | 0.15  |
| 美白              | 0.10  | 0.19  | -0.22 | -0.19 | 0.24  | 0.36  | 0.14  |
| 毛穴              | 0.25  | -0.20 | -0.06 | 0.11  | -0.03 | 0.70  | 0.70  |
| 爽快              | 0.09  | 0.21  | 0.58  | 0.50  | 0.31  | 0.21  | 0.27  |
| サラサラ            | 0.09  | 0.38  | 0.64  | 0.73  | 0.45  | 0.03  | 0.12  |
| 刺激              | 0.20  | 0.44  | 0.06  | -0.07 | 0.33  | -0.06 | -0.17 |
| 肌荒れ             | 0.50  | 0.07  | -0.27 | -0.31 | -0.05 | 0.05  | -0.04 |
| エイジングケア         | 0.09  | 0.17  | -0.18 | -0.12 | 0.06  | 0.38  | 0.24  |
| コスパ             | -0.18 | 0.00  | -0.06 | 0.00  | -0.05 | 0.07  | 0.08  |

赤いセル：本来の正解ペア

赤字：提案手法による正解ペア（単語間類似度による）

数値：評価項目と評価キーワード

赤いセル：本来の正解ペア 赤字：提案手法による正解ペア（単語間類似度による） 数値：評価項目と評価キーワード

図 7 評価キーワードが属する評価項目判定（提案手法）の精度検証

#### 4.3.2 評価キーワードが属する評価項目判定の精度検証と考察

各評価項目と 43 個の評価キーワードの類似度を Word2Vec により算出した結果の一部は図 7 の通りである。色がついているセルは本来の正解ペアで、図 6 の人手で作成した評価項目と評価キーワードの組合せである。赤字のペアは提案手法にて適切だと判定された評価項目と評価キーワードの組合せのことで、評価項目に対し、2 つ本来の正解となる評価項目がある場合は類似度上位 2 位までを適切な評価項目であると判定している。正解である人手による評価項目と評価キーワードの組合せと提案手法にて適切だと判定された評価項目と評価キーワードの組合せがどの程度一致したのか検証したの結果、67.2 % で一致したため、ある程度の有効性を確認できた。正しく判定できなかったペアに関しては、提案手法による判定が間違っていると感じるペアも、実験により作成したペアが間違っていると感じるペアもどちらもあった。正解データが主観による人手での判定、且つ 5 人という少人数での判定の集計のため、必ずしも本来の正解のペアが正しいとはいえず、提案手法によるリストが完全一致する必要はないと判断し、本節の評価で 7 割程度の判定精度を得られたため、提案手法は有効であると考え。

しかし、今後提案手法を用い構築していないアイテム分類に対する評価表現辞書を構築する場合は、アンケートをとって正解となる評価項目と評価キーワードの組合せを一から構築することはできないため、評価キーワードが評価項目に属するかの判定基準となる閾値を決める必要がある。

そこで、4.4 節にて、類似判定の閾値の検討方法を述べる。

#### 4.4 閾値を用いた評価キーワードが属する評価項目判定（提案手法）の精度検証

本稿での提案手法では、各評価項目に評価キーワードを振り分ける可否かの判定を、閾値を用いて判定するが、完全に自動化することは難しい。そこで、判定基準となる閾値を決め、類

| 閾値  | 平均+0.3           | 平均+0.2           | 平均+0.1           |
|-----|------------------|------------------|------------------|
| 一致率 | 58.1%<br>(32/55) | 72.7%<br>(40/55) | 80.0%<br>(44/55) |

図 8 閾値を用いた場合の評価項目キーワードが属する評価項目の判定（提案手法）精度検証

似度がその値を超えた評価項目を適切な評価項目としその中から実際に登録する評価キーワードと評価項目の組み合わせを決める。これにより完全に自動化するよりはコストがかかるが、従来の方法である、評価項目と評価キーワードだけを見て人手で判定するより効率的な評価項目への判定が可能になると考える。

本手法における閾値は、評価キーワードが評価項目に属するかの判定基準となるもので、評価キーワードに対する各評価項目の類似度の平均を算出し、その値の  $+\alpha$  とする。

閾値を用いた提案手法の有効性と、その閾値の大きさを検証するために  $+\alpha$  の部分を 0.3, 0.2, 0.1 で、4.3.2 節同様、人手による評価項目と評価キーワードの組合せと提案手法による評価項目と評価キーワードの組合せとどの程度一致するかの一貫率で比較する。比較した結果が表 8 の通りである。

平均+0.3 の時は閾値が高すぎるため、一致率が低く、平均+0.1 の時は一致率は高いが、閾値が低すぎるため、適切でない評価項目に判定された評価項目が多くあった。例えば、「エイジングケア」という評価項目に「コスパ」という評価キーワードが登録されるなど関連性の低い評価項目に判定されるケースが多くあった。そのため、中間である平均+0.2 が閾値として妥当であると判断した。本検証により、提案手法に閾値を用いることの有効性とその閾値の値は平均+0.2 が妥当であることを確認できた。

## 5. まとめと今後の課題

本稿では、コスメレビューを学習した Word2Vec を用いた評価キーワードが属する評価項目の自動判定手法を提案し、提案手法に用いる Word2Vec の学習に用いるコーパスにコスメレビューを適用することが有効であることを証明した上で、提案手法の有効性を検証した。

一般的な wikipedia を学習した Word2Vec と比較した時の、コスメレビューを学習した Word2Vec の有効性を評価項目「潤い」との類似度上位単語と比較した。結果、コスメレビューを学習した Word2Vec ではコスメ特有単語が類似度上位単語として多く抽出できたため、提案手法の Word2Vec の学習に用いるコーパスにコスメレビューを適用することの有効性を確認できた。提案手法の評価では、各評価項目に対する評価キーワードの振り分け精度が 67.2 % となり、提案手法のある程度の有効性を確認できた。評価表現辞書の構築手法を効率化、自動化することは、全 17 種類の辞書を作る上でコスト縮小になるのはもちろん、辞書作成者の主観による判定、識別が減るため、評価表現辞書を用いた自動スコアリングの精度向上にもつながると考えている。

今後は、本稿で述べた提案手法を用いて、構築していないコスメアイテム分類の評価表現辞書を構築するため、4.4節で述べた閾値を用い、評価キーワードが属する評価項目の判定が行えるかを確認し、提案手法を確立していく。

また、未構築のコスメアイテム分類の評価表現辞書を完成させ、全コスメアイテム分類に対する自動スコアリングシステムの開発を目指す。さらに、評価表現辞書構築手法におけるスコア付けの効率的な方法を検討する。

謝辞 株式会社アイスタイル様にコスメアイテムに関するレビューデータをご提供いただいた。また、本研究の一部は、JSPS 科研費 19K12243 および 17H01822 による。ここに記して謝意を表す。

## 文 献

- [1] 経済産業省,『我が国におけるデータ駆動型社会に係る基盤整備 平成 30 年』
- [2] @cosme, <http://www.cosme.net>
- [3] Yuuki Matsunami, Mayumi Ueda, Shinsuke Nakajima, Takeru Hashikami, John O'Donovan, and Byungkyu Kang, "Mining Attribute-Specific Ratings from Reviews of Cosmetic Products", *Transactions on Engineering Technologies (International MultiConference of Engineers and Computer Scientists 2016)*, pp. 101-114, Springer, 2017.
- [4] Asami Okuda, Yuki Matsunami, Mayumi Ueda and Shinsuke Nakajima, "Finding Similar Users Based on Their Preferences against Cosmetic Item Clusters", *The 19th International Conference on Information Integration and Web-based Applications & Services(iiWAS2017)*, pp. 154-158, December, 2017.
- [5] ランキング アイテム カテゴリー一覧 ( アイテム ) , <http://ranking.cosme.net/category/items>
- [6] Amazon.com, <https://www.amazon.com>
- [7] 価格.com, <http://kakaku.com>
- [8] @cosme サイトデータ資料< 2017 年 6 月>, 株式会社アイスタイル, <http://www.istyle.co.jp/business/uploads/sitedata.pdf>, (2018 年 12 月 17 日 アクセス)
- [9] Byungkyu Kang, Nava Tintarev and John O'Donovan, "Inspection Mechanisms for Community-based Content Discovery in Microblogs" *IntRS'15 Joint Workshop on Interfaces and Human Decision Making for Recommender Systems (http://reccx.ist.tugraz.at/intRS2015/)* at *ACM Recommender Systems 2015. Vienna, Austria. September 2015.*
- [10] 松波友稀, 上田真由美, 中島伸介, コスメアイテムに対する項目別スコアを用いた効果タグ推薦手法の提案. 第 10 回 Web とデータベースに関するフォーラム (WebDB Forum 2017), DE2017-22(2017-09), pp.139-144, 2017.
- [11] 奥田麻美, 松波友稀, 上田真由美, 中島伸介, コスメアイテム推薦のための類似ユーザ判定方式. 第 10 回 Web とデータベースに関するフォーラム (WebDB Forum 2017), DE2017-23(2017-09), pp.144-150, 2017.
- [12] 中嶋良樹. "化粧品サイトを対象にした口コミと売上に関する分析研究.", 経営情報学会 全国研究発表大会要旨集 一般社団法人 経営情報学会 2011 年秋季全国研究発表大会. 一般社団法人 経営情報学会, 2011.
- [13] 山田朋輝, 高山毅. "口コミを利用した感情可視化マップシステム.", 第 79 回全国大会講演論文集 2017.1 (2017): 785-786.
- [14] Kanayama, Hiroshi, and Tetsuya Nasukawa. "Textual demand analysis: detection of users' wants and needs from opinions." *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1. Association for Computational Linguistics*, 2008.
- [15] 石野亜耶, 村上浩司, 関根聡. "商品レビューからの購買意図の抽出とそれを用いた商品検索システムの構築", 言語処理学会 第 20 回年次大会 (NLP2014), 2014 年 3 月.
- [16] John O'Donovan, Vesile Evrim, Paddy Nixon and Barry Smyth, "Extracting and Visualizing Trust Relationships from Online Auction Feedback Comments.," *International Joint Conference on Artificial Intelligence (IJCAI'07)*, Hyderabad, India, January 2007.
- [17] Ivan Titov and Ryan McDonald, "A Joint Model of Text and Aspect Ratings for Sentiment Summarization," *46th Meeting of Association for Computational Linguistics(ACL-08)*, Columbus, USA, pp.308-316, 2008.
- [18] 姚佳, 井戸田博樹, 原田章. "インターネットのレビューが購買行動に及ぼす影響 -女子学生の化粧品購買のアンケート調査から-", 経営情報学会 2014 年春季全国研究発表大会, 2014 年 6 月.
- [19] 入江政行, 佐々木一, 坂田一郎. "Yelp レビュー情報における潜在的トピックを用いた類似・競合店舗の特定." 研究報告知能システム (ICS) 2018.14, pp.1-6, 2018.
- [20] 吉田朋史, 北山大輔, 中島伸介, 角谷和俊. "ユーザレビューの分散表現を用いた主観的特徴の意味演算による観光スポット検索システム". 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 (DEIM Forum 2017) , No. P6-5, 2017.
- [21] John O'Donovan, Shinsuke Nakajima, Tobias Höllerer, Mayumi Ueda, Yuuki Matsunami, Byungkyu Kang, "A Cross-Cultural Analysis of Explanations for Product Reviews" , *Proceedings of the Joint Workshop on Interfaces and Human Decision Making for Recommender Systems (IntRS 2016) co-located with ACM Conference on Recommender Systems (RecSys 2016)*, pp55-58, September 2016.
- [22] Duc-Hong Pham, Anh-Cuong Le, Thi-Kim-Chung Le, "A least square based model for rating aspects and identifying important aspects on review text data," *In Proceedings of the 2nd National Foundation for Science and Technology Development Conference on Information and Computer Science*, pp.265-270, 2015.