

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

松波 友稀[†] 上田 真由美^{††} 中島 伸介[†]

[†] 京都産業大学コンピュータ理工学部 〒603-8047 京都府京都市 北区上賀茂本山

^{††} 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

E-mail: [†]{g1245108,nakajima}@cc.kyoto-su.ac.jp, ^{††}Mayumi_Ueda@red.umds.ac.jp

あらまし 近年、商品に対するレビュー情報が閲覧可能なサイトが数多く存在しており、多くのユーザが商品の購入を検討する際に商品レビューを利用している。中でもコスメアイテムは、ユーザが直接自分の肌に使用するものであり、ユーザが自分に合わない商品の購入を嫌う傾向は、他の商品に比べて非常に高い。したがって、コスメアイテムの購入においては、このようなレビューサイトの存在意義は非常に大きいといえる。本研究では、個々のユーザにとって真に有用なコスメリビューを提供するためのコスメリビュー推薦システムの構築を目指し、コスメアイテムに対する評価項目別の評価表現辞書を用いたレビュー自動スコアリング手法を提案している。本稿では、コスメアイテムに対するレビューデータを用いて、評価表現辞書の構築およびこれを用いた自動スコアリング手法の有効性の検証を行ったので報告する。

キーワード コスメ、レビュー・評判情報・レビュー分析、情報推薦

1. はじめに

近年、@cosme[1] や MAQUIA [2] といった購入者によるレビュー情報が閲覧可能なレビューサイトが数多く運営されており、レビューを閲覧して商品購入の参考にするということが盛んに行われている。特にコスメアイテムは、直接肌に使用するものであり、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、個人によって肌質や好みが異なることから、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、我々は好みおよび使用感に着目し、価値観を共有しうるユーザによるレビューを推薦するシステムの構築を目指している。

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値を付与されている。しかし、既存のサイトの多くでは、対象アイテムに対する総合評価が行われており、投稿者の好みおよび使用感に関する評価がわからず、価値観を共有しうるか否か判断することは困難である。

そこで本研究では、多くのユーザが着目している“うるおい効果”、“美白効果”、“毛穴・角質ケア効果”などの項目を評価項目として設定し、レビュー文から各評価項目の評価値を算出する手法を提案している(図1)。具体的には、既存のレビュー文と評価値を分析することにより構築した評価表現辞書を用いて、評価項目別の自動スコアリングを実現する。我々は先行研究[3]で、コスメアイテムに関するレビューでは、アイテムの種類により異なる特徴的な評価表現が存在することを確認している。そこで、まず、「化粧水」に関するレビューを対象とし、ランダムに抽出した80件のレビュー文から評価表現を抽出し、

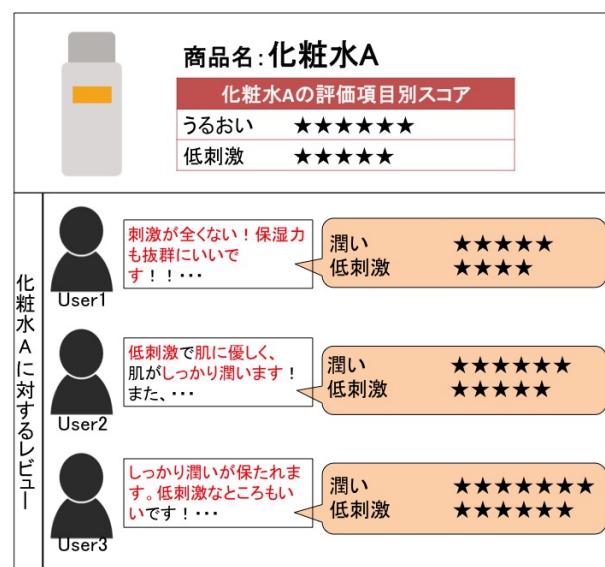


図1 各レビューデータに対する自動スコアリングの例

キーワードの共起に基づく評価表現辞書を構築した[4]。本稿では、先行研究における予備実験の結果を踏まえ、評価表現辞書への登録を改善させた上で、自動スコアリング方式の妥当性検証および、評価表現辞書の精度確認実験を行うことにより、提案手法の有効性を検証する。

以下、2章では関連研究との比較を述べ、3章では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について説明し、4章では自動スコアリング手法の妥当性の確認実験、また作成した評価表現辞書の精度確認実験について述べ、最後に5章でまとめと今後の課題を述べる。

2. 関連研究

レビュー情報を扱うサイトが多く存在する。Amazon.com [5] や価格.com [6] などの様々な商品を取り扱うオンライン通販サイトや、レストラン情報を提供する食べログ [7] などが有名である。また、近年では、コスメ・美容に特化したサイトである @cosme が有名である。@cosme の運営会社によると、2016 年 10 月時点で月間 2.8 億ページビュー、メンバー数が 390 万人、総レビュー数が 1300 万件と報告されており [8]、多くの女性がコスメ・美容に関した情報を @cosme 経由でやり取りしていることがわかる。更に近年では、アルゴリズムに着目した研究に加えてレビューデータの提示方法に関する研究についても研究者達は注目している [16]。

@cosme では、様々なブランドのコスメアイテムが掲載されており、コスメアイテム別のブランド比較が容易にできる。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、評価の良いレビュー・悪いレビューなど、目的に応じてレビューを検索することが可能である。また、会員登録の際に年代と肌質を記録するカルテ機能があり、自分の年齢や肌質と同じユーザによるレビューを検索することが可能である。

また、レビュー情報を扱うサイトが広く普及したことで、レビューを分析する研究が多く行われている。我々の先行研究でも、@cosme から化粧水とマスカラに関するレビューを分析し、各コスメアイテムレビュー分析に特化した化粧水とマスカラについての形態素解析用辞書を作成し、コスメレビューのポジティブ・ネガティブ判定を行っている [9]。この取組では、書き込みユーザの特徴も考慮したコスメレビューサイトの構築を目指し、課題として、広くブログや SNS からデータを収集し分析するための技術の開発に取り組むことを目的としている。本研究は最終的に評価項目別に評価値を推定し、自動で付与していくこと、また、将来的には一つの商品に対してレビューから得られた情報がより詳細にユーザに提供することを目的としている。

金平らは、実際の商品レビューに形態素解析・係り受け解析を行い、評価軸とそれに対応する評価表現辞書の自動構築手法、及び構築した評価軸情報を用いた分析手法を提案している [10]。評価表現辞書の評価の程度は、肯定表現と否定表現の比率を利用して表しているが、本研究では各評価表現に対して実際の人間による評価値を用いて辞書を作成している。

二本木らは、レビュー検索サイトに投稿されている多商品分野に渡る大量の自由記述文から、評価に関する表現を抽出、定量化し、効率的な分析・検索機能を提供する手法の提案をしている [11]。本研究は、コスメを対象としており、アイテムそのものの評価により好みに合うか否かを推定するだけでなく、そのアイテムの使用感がユーザ自身にとって適切か否かについても推定する。

姚らは、消費者がインターネット上に投稿されたレビューを信用する場合、どのようなレビューがより信用され、購買行動に影響するかを検証するため、以下の 3 つの仮説を立て、調査を行っている。(1) 消費者が購入したいと考えていた商品の

良いレビューはその商品の購買を促進し、購入を検討していなかった商品への悪いレビューはその商品の購買を阻害する、(2) 詳細データを書き込んだレビューは、客観的なレビューより信用出来る、(3) アドバイスを書き込んだレビューは、使い心地や個人的な感想・印象を書き込んだレビューより、信用出来る。211 名の女子大生を対象にアンケート調査を行ったところ、仮説検証は行えなかったが、分散分析からレビューが購買行動に影響を与えていることを確認している。

白田らは、@cosme のマスカラのレビューを手作業で収集し、概念グラフによるグラフ作成、モジュラリティによるコミュニティ抽出を用いて、レビュー分析を行っている [13]。この取組は、gephi を拡張した Consumer Behavior Analyzer を用いており、コミュニティの色分け表示及びグラフのレイアウト表示などの可視化を行っている。RIDF(残差 idf) を用いて、レビューから特徴語の抽出を行っている。

駒田らは、商品に関する意見や評判を含んだツイートから、どの属性に対してどのような評価なのかを自動的に収集・解析するため、属性語の自動抽出に取り組んでいる [14]。この取組では、ツイート中の商品評価文は、対象商品が明確でない点や崩れた文章が多い点などに対処するため、ツイート内の属性度と評価語の出現パターンを設定し、抽出した候補語の関連度を定義することで、属性語の自動抽出を行っている。

Ivan Titov らは、オンラインのレビューから対応するトピックを抽出し、格付けを取り出す統計モデルを提案している [15]。

John O'Donovan らは、オンラインオークションにおけるレビューの否定の情報を抽出し、この情報をグラフ式に示すシステムを構築している [17]。

上述の通り、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対する評価表現別のスコアリング手法に関する研究開発は行われていない。

3. 評価表現辞書を用いた評価項目別レビュー自動スコアリング方式

本節ではコスメアイテム評価表現辞書を用いた評価項目別レビュー自動スコアリング方式に関して説明する。まずは提案手法の概要 (3.1 節) を述べると共に、評価表現辞書の構築 (3.2 節) および自動スコアリング方式 (3.3 節) について説明する。

3.1 提案手法の概要

我々は、個々のユーザにとって真に有用なレビューを推薦するコスメレビュー推薦システムの実現を目指している。このようなシステムを実現するため、

- (1) 精度の高い推薦を実現するため、評価項目別のスコアリングの実現
- (2) 価値観を共有しうる類似ユーザの発見を実現することが必要である。

ここで本研究の位置づけを明確にするために、図 2 に将来的に構築を目指すレビュー推薦システムの概要を示す。(1)～(4) で記載 (青字) された処理がレビュー自動スコアリング処理の内容であり、(a)～(e) で記載 (赤字) された処理はレ

×フリーズ 潤いがかなり続く ≠ かなり潤いが続いた
○共起 潤いがかなり続く = かなり潤いが続いた
(終止形:続く) (終止形:続く)

図3 フレーズ表現による判定とキーワード共起による判定の違い

「化粧水」に関するレビューデータを80件取得し、特徴的な表現を手動で抽出し、女子大生2名で評価値を付与した。ただし、類似した意味を持つ表現が多数存在することから、@cosmeで「効果」として扱われている項目を参考に、39項目に分類した(表1の小カテゴリ)。また、「潤いがかなり続く」と「かなり潤いが続いた」といった表現を同一のものとして扱うため、キーワードの共起に基づく評価表現辞書とした(図3)。

3.3 評価表現辞書に基づく自動スコアリング方式

未知のレビューデータの自動スコアリングは、以下の手順で行う(図4参照)。また、図5に以下のレビュー自動スコアリング処理についての詳しい内容を示す。

- (1) レビューテキストを取得。
- (2) レビューテキストを外部から読み込む。
- (3) 形態素解析をした際に、句読点などを使って、文の区切りを判定。
- (4) 1つのレビューテキストからk個の文を抽出。
- (5) 辞書キーワードが含む文を判定。つまり、キーワードを見つけて、そのキーワードを含む文をピックアップ。
- (6) キーワードを含む文の中に、辞書中に含まれる共起条件を満たすものをピックアップ(1文の中に複数の共起条件を満たすものがあれば、優先順位を考えて、その共起条件を採用するか判定)
- (7) 採用された共起条件に該当する評価項目とその星の数を集計し、レビューテキストに対する評価項目毎のスコアを算出。

4. 評価

本節では、3.3節にて述べた自動スコアリング方式の妥当性の評価(4.1節)と3.2節で作成した評価表現辞書の精度確認実験(4.2節)について述べる。

4.1 自動スコアリング方式の妥当性の評価

キーワードの共起に基づく評価表現辞書を用いた自動スコアリング手法の妥当性を検証するため、実データを用いてスコアリング実験を行った。

a) 実験方法

ここで、レビュー実データに対するスコアリング実験の方法を示す。レビューデータは、@cosmeに投稿された「化粧水」に関するレビューからランダムに抽出した20件のレビューデータを用いた。レビュー1～レビュー10は「潤い」に対してポジティブな表現、レビュー11～レビュー20は「潤い」に関してネガティブな表現を含むレビューを用いた。本実験は、以下に示す3つの手法を比較することで行う。

- (1) 辞書を使用しない人手によるスコアリング手法(正解データ)

- (2) キーワードの共起に基づく評価表現辞書を用いた自動スコアリング手法(提案手法)
- (3) ペイジアンフィルタを用いたスコアリング手法(比較手法)

(1)では、辞書を用いずに被験者が実際にレビューテキストを読み、評価値0～7の範囲で評価項目「潤い」に関してスコアリングを行った。(2)は、事前に作成したキーワードの共起に基づく評価表現辞書を用いてスコアリングを行った。人手によるスコアリングと同様に0～7の範囲で評価項目「潤い」に関してスコアリングを行った。(3)は、新たに抽出したレビューテキストのうち、「潤い」に関してポジティブなレビューであるポジティブデータ56件、「潤い」に関してネガティブなレビューであるネガティブデータ28件を学習データとし、テストデータとして20件のレビューデータを判別した。なお、本実験では、1の大カテゴリを評価軸として用いた。

b) 結果

レビュー20件に対する人手によるスコアリング結果と、作成した評価表現辞書による自動スコアリング結果およびペイジアンフィルタを用いたスコアリング結果を図6に示す。

人手によるスコアリングの全ての項目の平均値(星の数)が3.36であり、提案手法(辞書に基づく自動スコアリング)の全ての項目の平均値が4.05であった。また、平均絶対誤差(MAE)は1.76であった。ペイジアンフィルタを用いたスコアリング手法の全ての項目の平均値は6.11であった。また、平均絶対誤差(MAE)は2.92であった。

人手によるスコアリングの方が、提案手法である辞書に基づくスコアリングに比べて、やや低い評価値となっているという傾向があるが、0～7という評価値の範囲に対して平均絶対誤差が1.76であり、提案手法の自動スコアリング手法では正解データに近いスコアリングを行えているといえる。しかし、レビュー11～レビュー20は「潤い」に対してネガティブな内容のレビューを用いたのに対し、図6からわかるよう正しく判定できていないレビューがいくつかみられる。これは辞書のネガティブ表現が不十分であったためであると考えられるが、評価表現辞書を改良することにより、さらなる判定性能の向上を見込むことができる。

ペイジアンフィルタを用いたスコアリング手法では平均絶対誤差が2.92であり、正解データと近いとはいえない結果となった。これは、ペイジアンフィルタを用いたスコアリング方式では、単語の出現回数などを考慮しているが評価表現及びその程度を表す単語との共起を考慮していないため、このような結果となったと考えられる。従って、本研究が提案する評価表現辞書を用いた自動スコアリング手法の方が有効であることを示す結果となった。

4.2 評価表現辞書の精度確認実験

次に、「化粧水」に関するレビューデータから、ランダムに抽出したレビュー10件を用いて2つのスコアリング手法によって得たそれぞれのスコアリング値の差の平均値である平均絶対誤差(MAE)を用いて、作成した評価表現辞書の精度を確認した。

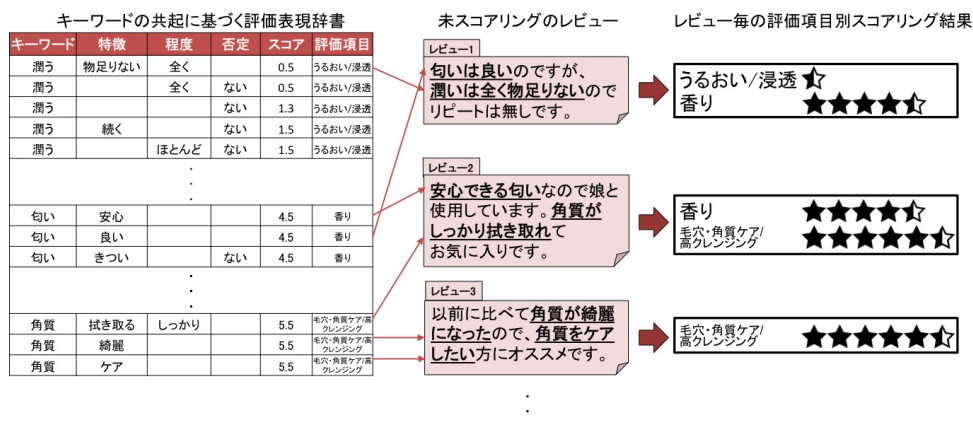


図 4 キーワードの共起に基づく評価表現辞書による自動スコアリング方式

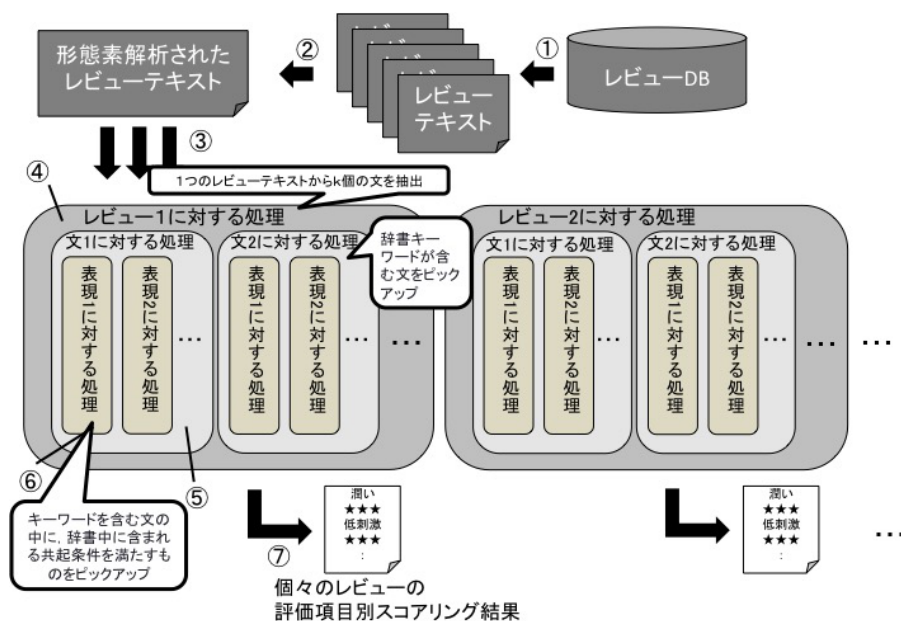


図 5 レビュー自動スコアリングの処理内容

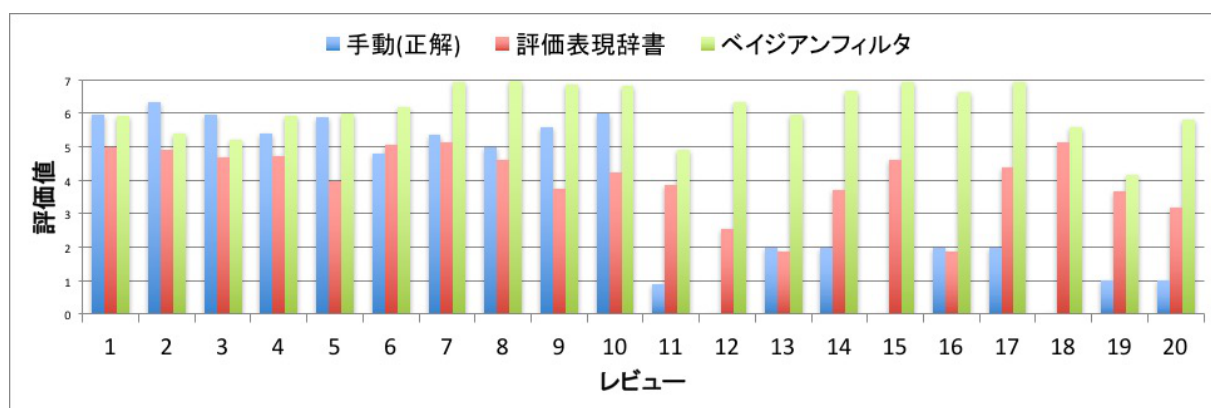


図 6 キーワードの共起に基づく評価表現辞書によるスコアリング結果

本実験は、以下に示す 2 つの手法を比較することで行う。

- (1) 辞書を使用しない人手によるスコアリング手法 (正解データ)
- (2) キーワードの共起に基づく評価表現辞書を用いた自動スコアリング手法 (提案手法)

(1) では、辞書を用いずに被験者が実際にレビューテキストを読み、事前に設定された 10 件のレビューに対して 0～7 の範囲でスコアを付けた。世代による差も考慮し、20 代から 50 代までの女性 (大学生および社会人) 30 名に被験者として評価項目 10 件に関してスコアリングを行ってもらった。提案手法である

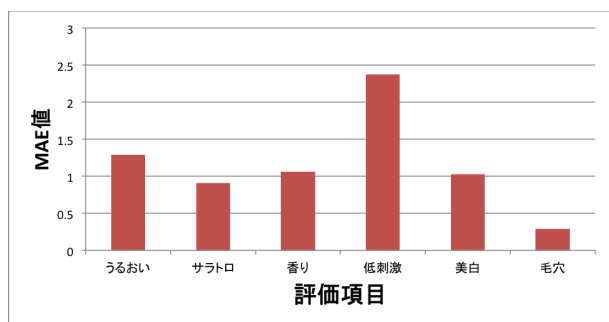


図7 レビュー10件に対する人手と辞書によるそれぞれのスコアリング値の差のMAE結果

(2) では、人手によるスコアリングと同様に事前に設定した10件のレビューに対して0~7の範囲で評価項目10件に関してスコアリングを行った。

評価項目毎の平均絶対誤差の結果を図7に示す。評価値が0~7なので、平均絶対誤差の幅が7であるうち、ほとんどの項目は1.30以下に抑えることができた。ただし、評価項目「低刺激」のみ平均絶対誤差が2.00を超える結果となってしまった。これは、「私は肌が弱く刺激を感じる事が多い体質なのですがこの商品は大丈夫でした!」といった、投稿者の肌質問題を語った表現も評価してしまっていることが原因と考えられる。このようにレビューする対象商品ではなく、自身の肌質問題また他の商品を含めて評価しているケースがある。今後はスコアリングする際にノイズとなる評価表現の傾向を確認し、各パターンに対して対策を立てることによって、自動スコアリングの精度向上を目指す。

5. おわりに

本稿では、個々の利用者に価値観の合うレビューを推薦するために、既存のレビューサイトのレビューデータに対して、評価項目別の評価値を付与する自動スコアリング手法を提案している。本手法では、実際のレビューから抽出した評価表現を用いて、キーワードの共起に基づく評価表現辞書を構築している。キーワードの共起に基づく辞書を利用することにより、「潤いがかなり続く」や「かなり潤いが続いた」という表現を同一のものと判断することが可能である。

また、代表的なコスメ・美容に関するサイトである@cosmeから収集したレビューデータに対して、提案手法を用いてスコアリングを行い、自動スコアリングおよびページアンフィラによるスコアリングと比較を行った。提案手法である評価表現辞書を用いたスコアリングでは、自動スコアリングに近い値を得ることができた。評価表現辞書の精度確認においては、自動スコアリングと提案手法によるスコアリングの値の平均絶対誤差を用いて確認したところ、ほとんどの項目について1.30以下に抑えることができた。

今後、評価表現辞書のさらなる拡充・調整を行い、スコアリング精度を向上させ、レビュー推薦システムに組み込むことを予定している。

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のご提供をいただきました。また、本研究の一部は、JSPS 科研費 16K00425 および 26330351 による。ここに記して謝意を表す。

文 献

- [1] @cosme. <http://www.cosme.net/>
- [2] MAQUIA. <http://hpplus.jp/maquia/cosme/search/index/>
- [3] 松波 友稀, 上田 真由美, 中島 伸介, コスメアイテムの使用感および嗜好度判定を目的としたレビュー分析手法の提案, 第7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 2015) P3-1, 2015 年 3 月.
- [4] 松波 友稀, 上田 真由美, 中島 伸介, 階上 猛, 岩崎 素直, John O'Donovan, Byungkyu Kang, コスメアイテム評価表現辞書を用いた評価項目別レビュー自動スコアリング方式, 第8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 2016) B1-1, 2016.
- [5] Amazon.com, <http://www.amazon.com/>
- [6] 価格.com, <http://kakaku.com/>
- [7] 食べログ, <http://tabelog.com/>
- [8] @cosme サイトデータ資料<2016 年 10 月>, 株式会社アイスタイル, <http://www.istyle.co.jp/business/uploads/sitedata.pdf>, (2017.1.12 アクセス)
- [9] 濱岡 祐美, 上田 真由美, 中島 伸介, コスメアイテムレビューサイト構築のための種類別の評価観点の抽出手法の提案, 電子情報通信学会 第22回 Web インテリジェンスとインタラクション研究会, WI2-2012-15, pp.45-46, 2012.
- [10] 金兵 裕太, 沼尾 雅之, ネットショッピングサイトの商品レビューを利用したジャンル毎の評価軸の自動構築, DEIM Forum 2016, C2-3, 2016.
- [11] 二本木 智洋, 住田 一男, 文の構造化によるレビュー評価の分析・検索, 情報処理学会インタラクション 2002 論文集, 2002 巻 7 号, pp.175-176, 2002.
- [12] 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
- [13] 白田 由香利, 橋本 隆子, 久保山 哲二, インターネット上のレビューサイトにおける化粧品の評判分析, 学習院大学計算機センター年報, 33, pp.2-7, 2013.
- [14] 駒田 康孝, 山名 早人, 商品評価ツイートからの属性語自動抽出手法の提案, DEIM Forum 2014, B5-6, 2014.
- [15] Ivan Titov, Ryan McDonald, A Joint Model of Text and Aspect Ratings for Sentiment Summarization, 46th Meeting of Association for Computational Linguistics (ACL-08). Columbus, OH, USA, 2008.
- [16] 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://recex.ist.tugraz.at/intRS2015/>) at ACM Recommender Systems 2015. Vienna, Austria. 2015.
- [17] 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.