

# 化粧品レビューテキストからの機械学習による評価推定

塚越 三蘭† 馬 青† 村田 真樹‡

† 龍谷大学大学院理工学研究科

‡ 鳥取大学大学院工学研究科

## 1. はじめに

消費者が商品を購入する際、商品本体に記載されている情報を参考に購入する。しかし、昨今ではインターネットにおけるレビュー投稿サイトが多く存在するため、レビューを参考に購入を検討する動きが数多く見られる。特に化粧品においては、肌に合うかどうかというような実際に使用してからでないと気づけないことがあるため、使用者からのレビューは重要である。そのようなレビューを扱っているサイトとしては、「コスメ・美容の総合サイト@cosme」[1]が第一にあげられる。このサイトには膨大な数のレビューが投稿されており、@cosmeがそれらのレビューを用いて独自に算出を行ったランキングが存在する。ランキングによって購入する商品を決める利用者も多い。しかしながら、世間で人気がある商品が必ずしもその利用者本人に合うわけではないため、ランキングが参考にならない場合がある。よって、膨大なレビューから自分に合う商品を探すしかないが、利用者個人によるレビューの分析は容易ではないため、本当に利用者に向く商品を見つけることは困難であるといえる。

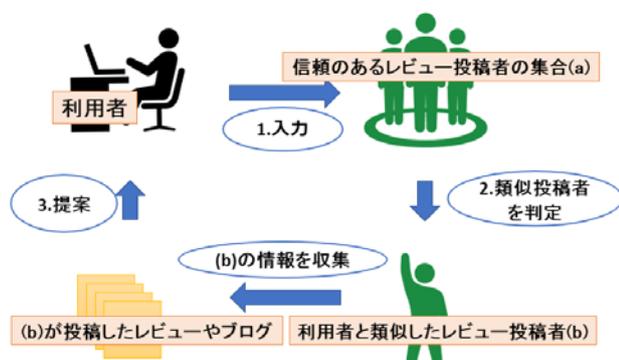


図1 将来的に開発したいシステムの概要

そこで、本研究では図1に示すような、利用者に向いた化粧品を提案するシステムの開発を目的とした。本稿ではその最初のステップとして、図1の「2. 類似投稿者の判定」を実現するためのベースとなる手法を提案する。信頼のあるレビュー投稿者の集合(a)に属する投稿者たちと利用者との、商品に対する使用感の共感度を数値化できれば実現できる。そこで本研究では、機械学習を用いて、化粧品のレビュー情報を扱うサイトに投稿されたレビューテキストから、化粧水に対する「美白」「毛穴」「ニキビ」「ハリ・ツヤ・弾力」「保湿」の5つの評価項目に、それぞれ「上」「中」「下」

「無」の評価値を推定する手法を提案する。その評価値が得られれば、共感度を算出できる。学習データには、レビューテキストに人手でラベル付けを行い作成したコーパスを使用する。レビューテキストをベクトル化する際に、係り受け解析で文節の繋ぎなおしと文の分割を取り入れた手法も提案する。機械学習がレビューの評価推定に有効であるかを確認するため、人手で作成したベースラインとの比較も行う。

## 2. 関連研究

商品に対するレビューを発信する機会が増加しており、そのようなレビュー情報を扱うサイトが数多く存在する。しかし、情報量の多さにより本当に必要とする情報を得るための作業が、利用者への負担となっている。そのため、負担の軽減を目的とした、レビューテキストを分析する研究が盛んに行われている。

レビューテキストの分析手法として、ニューラルネットワークを用いてタスクに適した評価項目(例えば、ホテルのレビューなら「サービス」、「立地」等)の特徴を抽出する手法の提案がなされている[2][3][4]。本研究では、化粧品に特化したサイトを対象としているため、これらの先行研究とはタスクが異なる。学習データにレビューテキストに手動でラベル付けを行ったコーパスを利用した点と、ベクトル化の方法も異なる。

一方、@cosmeのような化粧品に特化したレビューサイトが存在しており、そのようなサイトを対象とした研究も行われている。松波らは、コスメアイテムに対するレビューを用いて独自の辞書を構築することで、ラベル付けがされていないレビューに関して、各観点(例えば、アイシャドウのレビューなら「パール・ラメ」「ツヤ」「発色がよい」「立体感」「塗りやすい」)に評価値(0~7)を推定し、自動で付与する手法を提案している[5][6][7]。これらの研究では評価値推定に手動で構築した辞書を使っており、評価値推定に機械学習を用いている本研究と異なる。

## 3. 課題設定

本研究の最終目標は、利用者に向いた化粧品を提案するシステム(図1参照)の開発である。その開発過程で用いる、レビューテキストの機械学習を用いた評価推定手法の提案が本稿の課題である。

化粧品には様々なカテゴリが存在するため、研究の対象をその中でも最も使用感を重視するであろう「化粧水」に限定した。評価項目は「美白」「毛穴」「ニキビ」「ハリ・ツヤ・弾力」「保湿」の5つに分け、それぞれに対して「上」「中」「下」「無」の4段階で評価

推定を行う。機械学習は、5項目に対してそれぞれ別々に構成し評価推定を行うため、本研究では5項目の内から代表として「美白」に限定した。

レビューテキストを機械学習に用いるために重要な単語を抽出しベクトル化する手法を提案する。また、機械学習の有効性を確認するため、ベースライン手法を作成し比較する。これらの手法の詳細は次節で述べる。また、機械学習にはコーパスが必要で、これは、@cosmeから「化粧水」に対するレビューテキストを自動収集し、それに手動でラベル付けを行うことにより作成した。コーパス作成の詳細は5節で述べる。機械学習にはSupport Vector Machine(SVM)を用いる。機械学習による実験の詳細は6節で述べる。

## 4. 提案手法

機械学習にはSVMを用いる。そのためには、重要な単語を抽出し、ベクトル化する手法を提案する必要がある。4.1節でその手法について述べる。また、機械学習の有効性を確認するための、ベースライン手法については4.2節で述べる。

### 4.1 単語抽出とベクトル化

レビューテキストから重要な単語を抽出して、Bag-of-Words(BoW)で表現したベクトルを機械学習に用いる。

本節では形態素解析を用いた場合と、更に構文解析を用いた場合に分けて、重要な単語の抽出とベクトル化の手法を検討する。

#### 4.1.1 形態素解析を用いる場合

形態素解析を用いることでテキストを品詞毎に抽出することができる。そこで、重要な単語を品詞と頻度で限定する手法を提案する。指定したn種類以外の品詞を不要とし、以下のように「重要単語の集合」の決定とベクトル化を行う。これをpos\_mix手法と呼び、nを2(名詞, 形容詞), 3(名詞, 形容詞, 副詞), 4(名詞, 形容詞, 副詞, 動詞)と変化させるためpos\_mix(n=2, 3, 4)と区別することにする。

1. 収集したレビューテキスト<sup>1</sup>を形態素解析し、n種類の品詞の単語を原型で抽出する。その際にカタカナをひらがなに統一する。ストップワード<sup>2</sup>を取り除く。
2. 1で抽出した単語の出現頻度トップN以内の単語と、助動詞「ない」を含んだN+1個の単語を「重要単語の集合」とし、ベクトルの要素とする。
3. 個々の要素は対象のレビューテキストにその単語が出現すれば1、出現しなければ0の2値を取るone-hotで表現する。

しかし、pos\_mix手法の「重要単語の集合」は、n個の

品詞に優劣をつけずに頻度を用いるため、品詞に偏りが生じる。そこで、各品詞の数をN/n個ずつに揃えるように作成することも考えられる。つまり、それぞれ品詞の出現頻度トップN/nの単語を用いたn(N/n)+1個の単語を「重要単語の集合」として用いる。これを手法posと呼び、品詞数の変化は同様にpos(n=2, 3, 4)と区別することにする。

#### 4.1.2 構文解析を用いる場合

前節で述べた2つの手法の推定における問題点は、「美白」の評価推定に不要な単語をベクトルに用いてしまう点である。例えば、表1に示す文を見ると、n=4の場合、表1に示す6個の重要な単語が得られる。この場合の不要な単語とは「毛穴、綺麗、なる」のことである。なぜなら、これらは「毛穴」について述べられている単語であり、「美白」とは無関係だからである。そこで本節では、構文解析を用いて係り受けを考慮することにより、これらのような単語を不要とする手法を提案する。表1の文を例に以下に手順を示す。

表1 レビューテキストに含まれる文の例

文	抽出した重要単語(n=4)
毛穴が綺麗になったり、色白になったりします。	毛穴、綺麗、なる、色白、なる、する

1. 文を形態素解析し、係り受け解析を行う。以下の図2のような結果が得られる。文節の係り先は、係り受け解析のツールCabochaによる解析結果をそのまま用いる。
2. 1の結果から、文節の繋ぎなおしと文の分割を行うことで、以下の2つの係り受け関係を持つ文(以下、係り文)を得る。
  - ① 毛穴が綺麗になったり、なったりします。
  - ② 色白になったりします。
3. 上記で得た係り文が、「美白」に対して言及された文であるかの判定を、人手で作成した美白言及単語リスト(図3参照)内の単語が含まれているか否かで判定し、含まれていれば抽出する。文②のみが抽出できる。



図2 文を係り受け解析した結果

美白 シミ しみ ソバカス そばかす クスミ くすみ  
色白 白い 肌<sup>3</sup> トーン 透明感

図3 美白言及単語リスト

<sup>1</sup> テストデータは含まない。

<sup>2</sup> <http://svn.sourceforge.jp/svnroot/slothlib/CSharp/Version1/SlothLib/NLP/Filter/StopWord/word/Japanese.txt>

<sup>3</sup> 「美白」とは無関係に思えるが、「肌が明るくなった。」のように「美白」について言及している可能性があるため加えた。

上記の手順で抽出した文を、以下美白文と呼ぶ。前節のベクトル化は、ベクトルの次元構成に用いる「重要単語の集合」を求めるステップと、それらに対する要素に値をセットするステップの2段階に分けることができる。値セットのステップにおいて、本節の方法では、この美白文から抽出した単語「色白, なる, する」のみに対し、ベクトルのそれらに対応する要素に値1をセットする。つまり、不要な単語「毛穴, 綺麗, なる」に対応する要素の値は0のままで、不要な単語の影響をベクトルから除去することになる。

一方、次元構成のステップは以下の二通りの方法で求める。1つはレビューテキストの全文を対象に求める方法である。これはつまり、前節の手法pos(n=4)に用いた「重要単語の集合」の求め方と同じである。これを本節の値セットの方法と合わせてdep1と呼ぶ。もう1つはレビューテキストの全文を対象とせず、美白文のみを用いて「重要単語の集合」を求める方法である。これを本節の値セットの方法と合わせてdep2と呼ぶ。

## 4.2 ベースライン

ベースラインとして、「美白」を評価する単語を用いて、レビューテキストの評価値を算出する手法が考えられる。そのような単語には前節で述べた手法dep2の4(N/4)+1個の単語で作成した「重要単語の集合」を用いればよいと考えた。ベースライン手法の詳細を以下に示す。

1. dep2の「重要単語の集合」である、4(N/4)+1個の単語に対して、図4のように手動で値を付けて登録する。値は、効果を意味する単語に2、無関係の単語に1、程度を表す単語に0.5、否定の単語に-1、効果がないことを意味する単語に-2を付ける。

白い 2
明るい 2
薄い 2
悪い -2
やわらかい 1
(以下省略)

図4 値を付けて登録した単語の例

2. レビューテキストを文単位に分割する。
3. 1つの文に対して美白文を抽出する。
4. 3で抽出した美白文に対して、1で登録した単語が出現する度に、その単語の値の積を取る。値が出現する単語が、否定なら反転、程度なら半分、無関係のときは変化なし、にするために積を取る。登録単語が1つも存在しない場合は値を1とする。すべての美白文の値を得た後、各美白文の平均の値を1つの文の値とする。
5. すべての文の値を得たあと、各文の値の平均をレ

ビューテキストの最終の値とする。美白文が1つも存在しない場合、その値を0とする。

6. 最終の値が1または0なら「無」、1より大きいなら「上」、0と1の間なら「中」、0より小さいなら「下」と推定する。

## 5. コーパスの作成

本研究では、@cosmeに投稿されている化粧水カテゴリのレビューを収集してコーパスを作成した。その中から、収集したレビューに手動で評価値のラベル付けを行い、1000件のデータのタグ付きコーパスを作成した。ラベル付けの基準を表2に示す。

表2 「美白」項目のラベル付け基準

値	概要	レビューの例
上	① 美白になるとわかる文章	美白になります。
中	② ①に程度が加わる	まあまあ美白になります。
下	③ ①に否定が加わる ④ 美白にならないとわかる文章	美白になりません。 美白効果無しです。
無	⑤ 美白について述べられていない	潤います。 毛穴が閉じます。

ラベルの付け方は、著者の一人がレビューの文章を読み、評価値を付けるという方法で行った。ラベル付けには2つの問題点がある。

1. 客観性が保証できない
2. ベースライン手法との独立性が保証できない

問題点1の理由は、ラベルを著者一人の主観のみで付けたからである。しかし、複数の人間の手でラベルを付けると基準が一定でない問題もある。そこで、著者が付けたラベルが100件に達した時点で他者(3名)に確認と指摘をしてもらうことにした。

問題点2の理由は、基準に忠実にラベルを付けた場合、4.2節のベースライン手法に有利なバイアスがかかってしまうからである。解決策として、設けた基準に反したとしても、文章の前後関係から読み取った情報を優先してラベル付けを行うことにした。

## 6. 実験

### 6.1 実験条件

4.1節で述べた4種類のベクトルを用いる。SVMは機械学習ライブラリscikit-learn[8]のSVM(RBF)を用いた。ベクトルの次元数はNの数を120とした121次元であった<sup>4</sup>。実験には5節で述べたタグ付きコーパスを使用した。コーパスの各ラベルの割合は「上」:249件、「中」:142件、「下」:184件、「無」:425件であった。それらと同じラベルの割合で、学習データとテストデータを9:1の割合で分割した。学習データを用いて、グリッドサーチと5-fold交差検証を行うことにより、各機械学習

<sup>4</sup> Nの数を80(その結果、ベクトルが81次元)、120(同121次元)、160(同161次元)と3パターン用意し、予備実験により決定した。

における最適なハイパーパラメータを決定した。ハイパーパラメータの組み合わせ数は、1024通りで設定した。詳細は表3に示す。

表3 グリッドサーチに用いるハイパーパラメータ

機械学習	パラメータ	値
SVM (RBF)	C	$10^{-4}$ - $10^4$ 間に対数(基底10)スケールで32分割
	$\gamma$	$10^{-4}$ - $10^4$ 間に対数(基底10)スケールで32分割

## 6.2 実験結果と考察

SVMに手法 pos\_mix と pos を用いて品詞数 n が変化した場合の推定精度を表4に示す。表4の精度を比較すると、n=4が pos\_mix, pos 共に最もよい精度で pos においてはn=3がn=2を下回っていることから、「重要単語の集合」に動詞を加える手法が有効であるとわかる。しかしながら、pos においてはn=3がn=2を下回っていることから、「重要単語の集合」に副詞は不要である可能性がある。また、n=2のとき pos が上回っている以外に明らかな差がないため、どちらが有効かは断言できないが、品詞数の割合を調節することで精度が変化することを確認できた。

表4 品詞の数 n が変化した場合の推定精度

	n=2	n=3	n=4
pos_mix	0.54	0.56	0.66
pos	0.61	0.55	0.64

本実験よりすべての品詞を用いた場合の精度が一番よいことが確認できたので、各手法の比較実験はn=4に限定して行った。実験結果を表5に示す。

表5 各手法の推定精度

ベースライン	0.27
pos_mix	0.66
pos	0.64
dep1	0.62
dep2	0.70

表5の最も精度が高かったのは、dep2である。SVMを用いたすべての場合がベースラインより精度が高かった。dep2が最もよいことから、係り受け解析を取り入れた手法が有効であったとわかる。しかし、dep1の精度が最も低いが、これは「重要単語の集合」内の単語が美白文に出現しない場合が多く、個々の値をセットする際に0が大半を占めてしまったことが原因であると考えられる。

上記実験結果は、ハイパーパラメータの組み合わせの内、検証誤差のもっとも小さいパラメータセットを用いたものである。図5はn=4における、パラメータセットを検証誤差の小さい順に並べ、その上位N個を用いたときの各手法の平均精度を表している。もっとも性

能のよいdep2では右肩下がり傾向がみられることから、ハイパーパラメータをグリッドサーチにより決定することは有効であると考えられる。

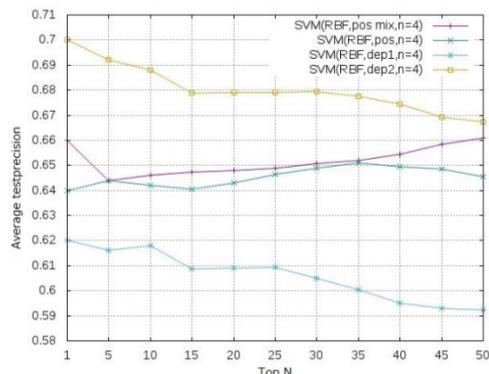


図5 検証誤差が小さい上位N個 (N=1~50) のハイパーパラメータセットを用いた場合の平均推定精度

## 7. おわりに

本稿では、化粧品のレビューテキストから、機械学習を用いて評価推定を行う手法を提案した。全ての場合において機械学習手法がベースライン手法より精度が高いことから、化粧品のレビューテキストの評価値の推定において、機械学習が有効であることが確認できた。また、レビューテキストのベクトル化を行う際に係り受け解析で文節の繋ぎなおしと分割を行う手法を取り入れることが有効であると示した。ベクトルの素性に動詞をいれることが有効であるとわかった。

今回は「美白」の項目に限定してレビューの評価推定を行った。今後は、残りの項目を同様の手法で評価推定し、最終目標の、利用者に向う化粧品を提案するシステムの開発を進めていく予定である。

## 参考文献

- [1]@cosme <https://www.cosme.net/>
- [2]ZHANG Bo, 白井清昭, レビューテキストの書き手の評価視点に対する評価点の推定, 言語処理学会第23回年次大会, pp. 803-806, 2017.
- [3]藤谷宣典, 三輪誠, 佐々木裕, 隠れ状態を用いたホテルレビューのレーティング予測, 言語処理学会第21回年次大会, pp. 764-767, 2015.
- [4]外山洋太, 三輪誠, 佐々木裕, 文書・文間及びカテゴリ間の関係を考慮したレーティング予測, 言語処理学会第22回年次大会発表論文集, pp. 158-161, 2016.
- [5]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(DEIM2015) P3-1, 2015.
- [6]松波遊稀, 上田真由美, 中島伸介, 階上猛, 岩崎素直: 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, データ工学と情報マネジメントに関するフォーラム(DEIM2016), B1-1, 2016.
- [7]松波遊稀, 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, データ工学と情報マネジメントに関するフォーラム(DEIM2017), B5-3, 2017.
- [8]scikit-learn, <http://scikit-learn.org/stable/>