

スコア重畳型の WordCloud を用いたレビュー提示手法

大野湧^{†1} 塚田浩二^{†1}

概要：近年、オンラインショッピングは市場を拡大し続けている。オンラインショッピングでの商品の比較には、レビュースコアやレビューコメントが重要な役割を持つ。本研究では、複数の商品の性質を比較する手法として、WordCloudに着目した。まず、各商品のレビューコメント群を形態素解析して頻出語を抽出する。次に、各語が出現するコメントに対応するスコアを累計し、各語毎のスコアを算出する。こうして得られた頻出語に対して、各語毎のスコアで採色することで、スコア情報を重畳表現できる WordCloud を作成する。本論文では、基礎的なアイデア、実装と表現例について示す。

1. 背景

近年、オンラインショッピングは市場を拡大し続けている。生田ら[1]や伊木ら[2]によると、Amazonや楽天市場などの電子商取引サイト（以下ECサイト）におけるレビューは商品決定において重要な役割を担っている。しかし、ECサイトにおけるレビューの表示方式は大きな変化が見られない。

本研究では、ECサイトの中でも代表的な楽天市場のレビュー表示に焦点を当てる。現在の楽天市場におけるレビュー表示の例を図1に示す。基本的には、各レビュー（スコア／コメント）がリスト形式で表示され、選択することでレビューの詳細を表示することができる。こうした表示方法はシンプルで扱いやすいが、レビュー全体を見通すことが難しく、複数の商品比較をする際の操作量も多くなってしまふ。そこで本研究では、レビューコメント群から抽出したWordCloudに対して、レビュースコアを元にした色情報を重ね合わせることで、各商品の傾向を俯瞰的に表現可能なレビュー提示手法を提案する。



図 1 楽天市場のレビュー一覧画面

2. 関連研究

レビューを対象とした信頼性判断支援システム[2]は、レビューに対するスコアを可視化し、スコアの根拠となる情報を提示するシステムである。スコアは類似性、協調性、集中性、情報性に基づいて求められる。ユーザが信頼性に対して意識を向けることができ、有用な判断支援を行うことを目指している。

購買行動促進のための重要語抽出とタグクラウド生成[3]は、レビューデータを整理するために、KNP*1や日本語WordNet*2を用いて機械的に商品紹介やそのレビューからタグにふさわしく、表記ゆれのない語を抽出する。その後、抽出した語をタグクラウドを用いて視覚化するシステムである。

RakuTempo[4]は、楽天市場等のオンラインモール上の店舗の特徴と店舗間の関連性の情報視覚化により、ユーザの嗜好に合った店舗の発見や回遊を支援するシステムである。従来の店舗検索システムと比べ、各店舗の特徴を把握しやすくなり、素早く店舗の回遊が可能となったと報告している。

秋山ら[5]の研究では、多くのユーザから参考にされているレビューの印象に着目した分析を行っている。レビュー文中にポジティブ、ネガティブ両方の感情が含まれている場合や、ニュートラルの箇所が多いレビューは評価が高くなりやすい傾向にあることが確認されている。

本研究では、レビューコメント群から抽出したWordCloudに対して、レビュースコアを元にした色情報を重ね合わせることで、各商品の傾向を俯瞰的に表現可能なレビュー提示手法を提案する。

3. 提案

ここでは、従来のレビュー表示の課題分析と、本研究の

^{†1} 公立はこだて未来大学

*1 <http://nlp.ist.i.kyoto-u.ac.jp/?KNP>

*2 <http://compling.hss.ntu.edu.sg/wnja>

提案について述べる。

3.1 課題分析

現在の楽天市場におけるレビュー表示は、図1のようなリスト型となっており、多くのECサイトに共通している。これらのレビューの特徴や課題について、我々は予備的なインタビュー等を通して以下のように分析した。

- 個別のレビューに対しては多様な方法でアクセスできるが、レビュー全体を見通した判断は難しい。なお、レビュースコアの平均を見れば商品の全体的な評価は確認できるが、レビューコメントで記述された詳細情報は抜け落ちてしまう。
- 多数のレビューコメントを見比べる場合、操作量が增大する。

3.2 提案システム

上記の問題点を解消するために、本論文ではスコア情報を重畳表現できるWordCloudを用いたレビュー表示手法を提案する。

具体的には、まず、各商品のレビューコメント群を形態素解析して頻出語を抽出する。次に、各語が出現するコメントに対応するレビュースコアを集計し、各語毎のスコア分布を算出する。こうして得られた頻出語に対して、各語毎のスコアで採色することで、スコア情報を重畳表現できるWordCloudを作成する(図2)。

単なるWordCloudとは異なり、各語がPositive/Negativeな意味で使われているかも併せて一覧できる点が特徴である。

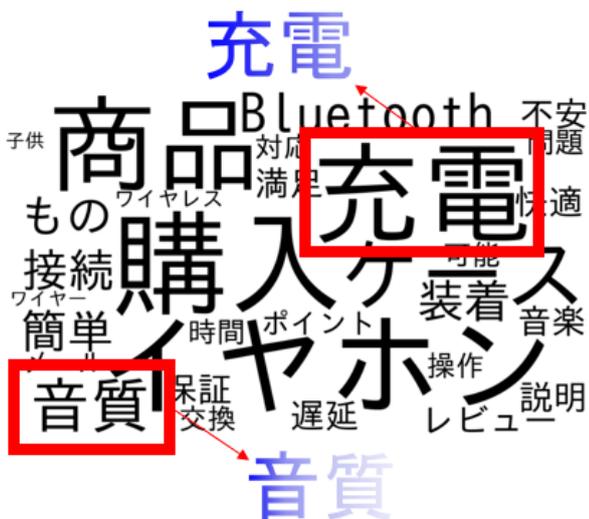


図2 スコア情報を重畳表示可能なWordCloudのコンセプト

4. 実装

本システムの実装は、Pythonを用いてGoogle Colab上で行った。まず、楽天市場からの商品情報を楽天商品検索APIを用いて取得する。具体的には、検索キーワードを指定すると、検索結果の上位(現在は20件)を取得/表示する。

次に、各商品に対して、レビュー(スコア/コメント)をスクレイピングを用いて収集する。現在は上位20件のレビューデータを対象としているが、任意に拡張できるように実装している。

ここで、レビューコメント群に対してMcCabを用いた形態素解析を行い、頻出語を抽出する。その後、商品の特徴を表しやすい品詞として名詞のみを抽出した。なお、一部の不要な語はNGワードに指定することで排除した。また、WordCloudの視認性を考慮して、頻出語は上位30個を利用するものとした。各頻出語が出現するコメントに対応するスコアを累計し、コメント数で割ることで、各語毎の平均スコアを算出する。

こうして得られた頻出語に対して、各語毎のスコアで採色することで、スコア情報を重畳表現できるWordCloudを作成する(図3)。ここで、単語の大きさは単語の出現頻度を表し、色は濃淡が濃い程高いスコアを表している。最終的には、各単語をスコアの分布でグラデーション状に塗り分ける予定だが、実装上の問題で今回はスコアの平均値を用いて単色で塗ることにした。



図3 WordCloudの作成例(色の濃淡が濃いほど平均スコアが高い)

なお、スコアと色の対応付けとしては、文献[6][7]等を参考に、青を主体として彩度や明度を調整した色パターンで評価を表した。今回利用した色パターンを図4に示す。



図 4 スコアと色の対応付け

最後に、商品の一覧と合わせてWordCloudを表示した例を図5に示す。図5の左側の商品は、「ストラップ」や「見た目」というワードが大きく表示されており、外観に関するコメントが多いことや、「子供」や「プレゼント」というワードの色が濃いことから、子供用のプレゼントとして高評価なことが分かる。右側二つのWordCloudを見比べると、全体的に右側の方が色が濃いため、評価が高いことが分かる。例えば、「音質」等の共通語に着目しても、中央は色が薄く、右側は色が濃くなっている。また、右側の商品はワイヤレスであり充電しやすさなども気になるが、「充電」や「ケース」の色が濃いことから、ケース型の充電方式が概ね高評価であることが分かる。



図 5 商品一覧画面上での提案システム実行例

5. 議論

ここでは、スコア重畳型のWordCloudの生成に関連して、「品詞や表示数の調整」「塗り方の調整」「文書間での頻出語調整」「Webサービス化」について議論する。

まず、現在は名詞のみを抽出対象としているが、システムとしては形容詞／動詞のみも扱えるよう実装している。スコア分布の影響が出やすい品詞を今後調査し、調整していく。表示する語の数については、現在は頻出語の上位30件のみを利用している。出現頻度の分布は考慮していないため、例えば図5の中央の例（出現頻度の高い語が多い）場合だと、右の例（出現頻度の低い語が多い）場合と比べて、かなり見にくい印象になる。こうした観点も考慮したうえで、今後は単語のサイズや表示数に反映していきたい。

塗り方については、最終的には図2のように各語の出現するスコアの分布が一目でわかるように、語の中でグラデーションを作成する予定である。ただ、この場合彩度を変化させる単色グラデーションでは差が目立ちにくくなる可能性があるため、色相の変化なども併せて利用を検討していく。

文書間での頻出語調整については、現在の実装では行っていない。よって、図5の例でも「購入」「商品」といった一般的な名詞が存在している。これは、同じ単語同士でも色の違いでスコアが比較できると考えた意図的な仕様だが、あまりに一般的な単語が増えすぎると特徴把握が困難になる可能性もあるので、今後はTF/IDF等を用いた文書間での頻出語調整についても一定の範囲で活用する予定である。

最後に、システムをWebサービスとして公開し、誰でも活用できる状態としたい。図5の左上にあるような検索欄に任意の検索キーワードを入力するだけで、多数の商品間の特徴を俯瞰的に比較できるようなサービスとしての実装を検討している。

6. まとめと今後の展望

本稿では、レビューコメント群から作成した WordCloud に対して、スコア情報を重畳表示可能なシステムを提案・試作した。今後は、議論で述べたような、「品詞や表示数の調整」「塗り方の調整」「文書間での頻出語調整」「Webサービス化」といった課題を整理し、より商品の特徴を把握しやすい視覚化システムの構築を図る。

参考文献

- 生田目崇, 鈴木元也. EC サイトにおけるサイト閲覧行動と購買行動の分析, 経営情報学会, 全国研究発表大会要旨集, 2013s(0), 274-277, 2013.
- 伊木淳, 亀井清華, 藤田聡. レビューを対象とした信頼性判断支援システムの提案, 情報処理学会論文誌, Vol.55, No.11, 2461-2475, 2014.
- 松崎朋美, 波多野賢治. EC サイトにおける購買行動促進のための重要語抽出とタグクラウド作成, 情報処理学会研究報告, 2014-IFAT-114(10), 1-6, 2014-03-22.
- 大河原一輝, 平野廣美, 益子宗, 星野准一. ショッピングモール型 EC サイトのための店舗情報視覚化システム, 情報処理学会論文誌, Vol.56, No.3, 847-855, 2015.
- 秋山和寛, 熊本忠彦, 灘本明代. レビューの印象に基づいた評価の高いレビューの分析, 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(DEIM2019), 2019.
- 槇究, 渡部裕子, 飯島祥二. 単色の印象操作—背景色と個人差に着目して—, 日本色彩学会誌, 31(1), 2-13, 2007.
- 大森正子, 橋本令子, 加藤雪枝. 色彩刺激に対する心理評価と生理反応評価, 日本色彩学会誌, 26(2), 50-63, 2002.