

化粧品レビュー文における商品間関係の抽出

Extracting Product Comparisons from Cosmetic Reviews

安部 小百合 小林 一郎
Sayuri anbe Ichiro Kobayashi

お茶の水女子大学大学院人間文化創成科学研究科理学専攻

Advanced Sciences, Faculty of Sciences, Graduate School of Humanities and Sciences, Ochanomizu University

In recent years, CGM(Consumer Generated Media) have huge information of users' opinion. Especially in product reviews, users often compare different products, from viewpoints of their favorite, the reasons why they chose, etc. In this paper, we have investigated the relation among shopping items, especially the same shopping items of other brands, in consumers' reviews. We show a case study of extracting information about comparisons among cosmetics, and visualize the relation among cosmetic brands.

1. はじめに

近年、インターネットにおけるCGM(Consumer Generated Media)においてユーザの意見が多数発信されるようになった。その一つにレビュー文がある。レビュー文には多くの評判情報が含まれている。中にはその商品のみの評判情報だけでなく、他の商品との比較や同時使用に関する情報が書かれている場合も多くあり、商品を選択し購入する大きな動機づけになっている。しかし、これらの文書量は膨大であり、全てに目を通して判断することは時間や労力の面から不可能であるため、情報収集を効率的に自動化をする解決法が求められている。この問題へのアプローチとして、レビュー文の分類や情報抽出、要約の手法等が活発に研究されている。

本研究では、レビュー文に現れる化粧品ブランド間の関係に着目する。例えば、ブランドAを対象にしたレビュー文の中に「ブランドBを前は使っていたけれど」という文面が出現した場合、投稿者はブランドAとブランドBの双方を使用した経験があると言え、「ブランドBと重ねるとよかった」とのレビューからはそれらの組み合わせの知識が得られる。さらには、どちらの場合もブランドAとブランドBに関連性があると捉えることもできる。商品検索の際に、現在自分が使用している、また自分が関心を持っているブランドと関連性が深いブランドもそのユーザにとって選択肢となる可能性が高いため、これらの情報は非常に有益であると言える。このことから、本研究の先行研究となる[4]にてユーザ任意のブランドにとって関連性のあるブランドをレビュー文から抽出し、さらに、化粧品ブランドに関する比較関係の抽出を行った。本稿では特にそれを提示するために関係性を可視化すること目的とする。

2. 比較関係の抽出

先行研究[4]において、レビュー文における比較関係の抽出に取り組んでいる。手法として、まずブランド名辞書を構築し、他のブランド名が出現するレビュー文を抽出した。その後ブランドの関係をSVMやナイーブベイズ分類期の機械学習手法を用いて比較と併用とに分類した。また、「NEより」「前はNE」(NEはブランド名)といったルールベースの抽出も行った

連絡先: 安部小百合, お茶の水女子大学大学院人間文化創成科学研究科理学専攻情報科学コース, 〒112-8610 東京都文京区大塚 2-1-1, g1020502@is.ocha.ac.jp

本稿ではブランド同士の比較関係をユーザに提示する手法として比較関係グラフを構築する。

2.1 使用データ

今回使用したデータは、化粧品レビューサイトアットコスメ[5]のレビュー文の中でカテゴリ名マスカラの11020件を用いた。含まれているデータには、アイテム名、ブランド名、商品名、レビュー文章、おすすめ度、投稿ユーザの属性等がある。この中で、アイテム名、ブランド名、レビュー文章を使用した。各レビューは商品ごとに書かれており、商品にはそれぞれブランドとメーカーがある。化粧品の商品名は非常に長い場合が多く、商品名でなくブランド名で呼ばれることが多い。よって、本稿ではブランド名を使用する。

他のブランド名が出現するレビュー文の例を表1に示す。

表1: 他のブランド名が出現するレビュー文の例

アイテム名	ブランド名	レビュー文
マスカラ	マリ・クレール	... 個人的にはマジョマジョより好き。多分リビします
	デジャヴュ	... 私はいつも、 <u>KATE</u> マスカラ下地→ <u>ヘレナ</u> なんです...

2.2 ブランド名辞書の構築

ブランド名を抽出するためにブランド名を登録した辞書を構築する。多くのブランド名には表記の揺れが存在し、Feldmanら[1][2]は車や靴のブランド名の表記揺れについて数字の表記の違い(5とVなど)や単語の省略をそれぞれのブランドにおいて一致させる正規表現でパターンを作成して対応している。

化粧品ブランドにおいても略語や表記揺れが多くみられる。例えば、デジャヴュというブランド名をデジャブ、デジャヴといった英単語をカタカナに直した際の日本語特有の表記揺れもある。また、デ`シ`ャウ`ユのように濁点や半濁点を別の文字で表記しているケースや、マリ・クレール→マリクレールのように中点を省略するケースも見られる。今回は略語を人手で追加し、アイテム名マスカラを対象にした16件のブランド名を判定する辞書を構築した。

3. グラフによる比較関係の可視化

第2章において説明したレビュー文と辞書を使用したところ、他のブランド名が出現するレビュー文は1687件存在した。これを各ブランドをノードとし、レビュー文のブランド名→出現したブランド名とエッジを張り、出現数を重みとした有向グラフとして構築した。表1に示すレビュー文の場合のグラフを用いた可視化結果を図1に示す。

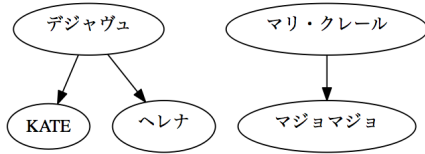


図1: グラフを用いた表1のレビュー文の可視化結果

3.1 商品間関係の可視化とエッジの削減

構築したグラフを可視化したものを図2に示す。出現したエッジをすべて表示した場合、図2のように非常に煩雑で見にくいグラフになってしまうためエッジの重みが小さい順に全体との比率を考慮してエッジを削減する。例として全体の60%を削減したグラフを図3に示す。

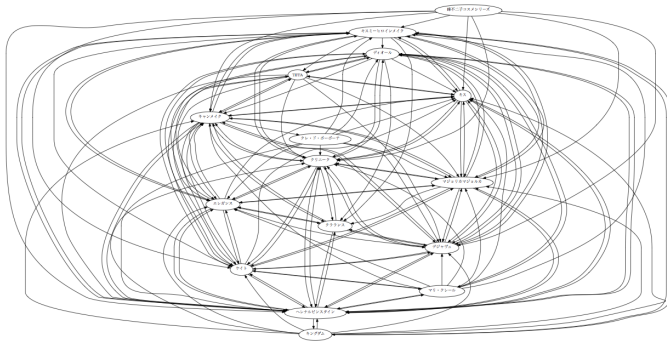


図2: ブランド間の関係をすべて図示したグラフ

3.2 考察

レビュー中に他のブランド名が出現した数を重みとし、ブランド間の関係の強さを示したグラフの構築を行った。全ての関係を表示するとグラフが煩雑となり可読性が低下することより、重みが小さいものを削減した。グラフを見ると多数のブランドからエッジが張られているブランドもあるがひとつも張られていないもの、またエッジを出している数もブランドにより偏っている。ブランド別のレビュー数は少ない物で91件、多い物で2160件であり、このことも偏っている原因ではないかと考えられる。重みの調整は今後の課題としたい。

4. おわりに

本稿では、先行研究[4]において化粧品レビューから抽出されたブランド同士の関係をユーザにわかりやすく提示するためにグラフを用いて可視化を行った。関係性を特定する手法として他のブランド名が出現する回数を用い、それらの可視化手法としてブランド名をグラフのノードとして関係性を示した。また、可読性および重要な関係のみの提示を目的として、ブランド間の関連性の少ない削減を行った。

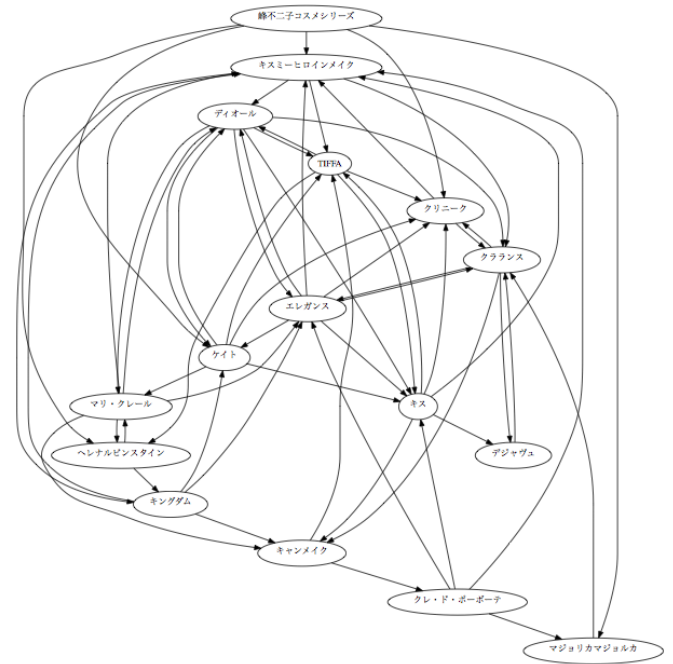


図3: 関連の少ないエッジの60%を削減したグラフ

今後の課題として、より多くの情報を得るためにブランド名の表記揺れに対応するブランド名辞書の拡充を行う。また、テキストの係り受け解析等を使用して関係性があるか否かだけでなく、「香り」「色み」「価格」「組み合わせ」のような評価項目やその良さ悪しのようにどのような関係性があるかを抽出し、さらにユーザが商品選択をする際に有益な情報を抽出したいと考えている。

謝辞

本研究では、株式会社アイスタイル様よりデータを提供していただきました。ここに感謝の意を表します。

参考文献

- [1] Feldman R, Fresko M, Goldenberg J, Netzer O, Ungar L, Extracting product comparisons from discussion boards. Proc. Seventh IEEE Internat. Conf. Data Mining 2007 (IEEE, Piscataway, NJ), 469-474, 2007.
- [2] Feldman R, Fresko M, Goldenberg J, Netzer O, Ungar L, Analyzing Product Comparisons on Discussion Boards, Language, Culture, Computation. Computing - Theory and Technology, pp 399-408, 2014.
- [3] 佐藤敏紀, 奥村学, blogからの比較関係抽出. 情報処理学会自然言語処理研究会, pp.7-14, 2007 レビューテキストを対象とした評価条件の抽出手法, 言語処理学会第19回年次大会, A4-2, 2013.
- [4] S. Anbe, I. Kobayashi, "An Approach to Category Classification of Cosmetics Reviews based on Brand Names", Proceedings of SCIS-ISIS, Kitakyusyu, pp. 971-975, 2014.
- [5] <http://www.cosme.net/>