

研究指導 中澤 真 教授

企業の商品改善のための化粧品サイトによるクチコミ分析

関 由理恵

1. はじめに

企業にとって市場の中で顧客の変化するニーズを掴み、不満をくみ取り、商品を改善につなげることが重要である。しかし、ユーザのニーズを掴むことは企業にとって容易ではない。

特にニーズを掴むことが難しい商品の1つとして、化粧品が挙げられる。その理由は、化粧品のニーズが細分化しているからである。化粧品は年齢・性別を問わず広く使用されており、商品に対する意見も幅広いものとなる。そのため、このような意見を確実に効率よく掴んでいくことが求められる。

ユーザの多種多様なニーズを把握するには大量のデータが必要となるが、これを簡単に収集できる情報源の一つがインターネット上のクチコミである。クチコミには商品を使用したユーザが感じた商品の良いところと悪いところが記されているため、悪い点に関する記述を集めることができれば企業は商品改善に活かすことができる。実際にテキストマイニングの技術を活かし、クチコミによる分析結果をマーケティングに活かす試みも行われている[1][2]。

本研究では否定的な表現に着目してテキストマイニングを行い、商品の改善点を発見する手法を提案する。

2. 化粧品サイトにおけるクチコミの現状と課題

2.1 化粧品クチコミサイトの現状と課題

クチコミには商品を使用したユーザが感じた商品の良い点と悪い点が記されているため、企業は商品改善の材料とすることができる。しかしどの部分がどのように悪いのか具体的に書かれている文でないと商品改善に活かすことはできない。例えば、「一度商品を買ったのですが結局返品しました」という文は、商品が気に入らなかったという情報は含まれているが、商品のどの部分が気に入らなかったのかという商品改善に役立つ情報はない。そのため商品の使用感、デザイン、金額、サイズ、色、香り等に関する問題点を指摘している文を、大量のクチコミデータの中から効率的に抽出することが必要になる。

一般的なクチコミサイトでは自由記述文の他に段階評価の情報も入力してクチコミを投稿する。商品改

善を目的とする場合、低評価のクチコミを集めることになるが、このようなクチコミの中にも部分的に商品を褒める言葉が含まれていたりする。この点からも、低評価のクチコミを単純にフィルタリングするのではなく、その内容を分析して必要な文だけを抽出することが求められる。それにあたり、まず必要な文にはどのような語が入っているのか明らかにする必要がある。

2.2 クチコミ文の分析

商品やサービス改善のためにクチコミを利用し、肯定的な表現、否定的な表現を抽出する研究はいくつか行われている。門田ら[1]は、ホテルのサービス改善のためクチコミ内の否定的な表現を収集し、さらに品詞に基づいて意見・原因が入る文を抽出し、極性を付与している。この研究は肯定的な表現と否定的な表現の両方に注目していること、極性を付与していることなどの特徴がある。しかし精度が低く、否定的な表現を十分に集められていない。

定政ら[2]の研究は車の購入理由を探ることを目的としている。そのため商品名や、理由を表す助詞である「ので」という単語が入るクチコミを収集している。そして、購入した商品とその理由を表す箇所にラベル付けし、商品の購入を表す表現とその理由を表す表現の組を抽出する方法を提案している。この研究は、助詞の「ので」に注目し、車の購入理由を抽出することを試みている。だが購入理由だけでなく、商品の不満な点を表現する際にも、「ので」のような特徴的な語が出現すると考えられる。そこで本研究では低評価のクチコミ内で批判する記述がある文、かつその批判する語の対象が具体的である文のみを抽出する方法を提案する。

3. 分析方法

本研究では低評価のクチコミの中から商品改善に必要な情報を文単位で抽出する方法を提案する。

3.1 クチコミの収集

分析には化粧品クチコミサイトの中でクチコミ数が多い「@cosme」¹のクチコミを利用する。クチコミ収集・分析には統計分析ソフト「R」²、形態素解析ソフト「MeCab」³を用いた[3][4]。主に口元に使用する化粧品のカテゴリを対象に、0から7までである段階評価のうちの評価1のクチコミのみを収集した。この理由は、

¹ <https://www.cosme.net/>

² <https://www.r-project.org/>

³ <http://taku910.github.io/mecab/>

高評価のクチコミに比べ、低評価のクチコミにはより多く悪い点が記述されており、低評価のクチコミを収集すれば有効にデータを扱えると考えたからである。このサイトでは評価0のクチコミも存在するが、ユーザが段階評価を付け忘れた場合にも0が設定されるため、今回は収集の対象としなかった。また商品を使用した期間とクチコミを投稿した期間にずれがある場合も考えられるため、期間は問わないこととした。⁴

収集するクチコミは学習用データと評価用データの二つに分けられる。学習用データは商品改善に役立つ情報を含む文か否かの判定ルールを生成するためのものであり、評価用データはこのルールの有効性を評価するためのクチコミ群である。

学習用データとしては、ディオールの「ディオール アディクトリップマキシマイザー」⁵(以下商品Aとする)のクチコミ全176件を使用する。また評価用データはオペラの「リップティント」⁶(以下商品Bとする)から全116件と、キャンメイクの「ステイオンバームルージュ」⁷(以下商品Cとする)から全124件を用いる。収集した件数にばらつきがある理由は、評価1のレビュー数に偏りがあったためである。

3.2 商品改善のための必要な文の判定

商品改善に役立つ文の抽出を考えるために、商品改善に必要な文(以下「必要文」)の定義をここでは考える。まず、商品特性を考慮し、必要文は以下の条件のうち、条件A, B, C, D, E, Fのいずれか1つを含むこととする。

- A) 商品の使用感について記されている
- B) 商品のデザインについて記されている
- C) 商品の金額について記されている
- D) 商品のサイズについて記されている
- E) 商品の色について記されている
- F) 商品の香りについて記されている

この条件に加えて、文中に具体的な批判内容が含まれているものを必要文と定義する。例えば、「キャップを外したら香る独特の香りは、私はちょっと苦手です」という文は条件Fにあてはまり具体的な批判内容が含まれているため必要文となる。

この定義に基づいて、学習用データに必要文と不要文のラベル付けを人手で行う。今回の学習用データとして使用する176件のクチコミに含まれる文の総数は871であり、ラベル付けの結果必要文は363、不要文は508となった。不要文はすべての文から必要文を除いたものとする。この学習用データの単語頻度をみることで、必要文中の単語にどのような出現傾向があるのかを明らかにする。

3.3 必要文判定のための特徴語の選出

次に学習用データを用いて、必要文と不要文を判別するのに適した単語の選出を試みる。考え方としては、必要文に頻出し、不要文では出現しない単語を見つけることができれば、これを必要文を判断する際の判断材料とできる可能性がある。本研究ではこれを特徴語とする。今回は選出する単語が単独で意味をなす語となるように、品詞が形容詞(自立)、動詞(自立)、副詞(一般、助詞類接続)、名詞(サ変接続、一般、形容動詞語幹)のいずれかであることを条件とした。

また、3.2の定義により分類した必要文中の頻出単語の傾向をみることによって、他の商品のクチコミでも批判的な記述がある文を抽出できるのではないかと考えた。

表 1は、学習用データの必要文、不要文のそれぞれに出現する単語の頻度を算出し、単語の出現頻度差の降順で単語を選出した結果である。

表 1 必要文による特徴語の選出

単語	品詞	品詞の分類	必要文内の単語の出現頻度	不要文内の単語の出現頻度	出現頻度の差
唇	名詞	一般	158	55	103
荒れる	動詞	自立	69	18	51
する	動詞	自立	264	214	50
なる	動詞	自立	92	42	50
皮	名詞	一般	50	7	43

表 1の出現頻度の差は、必要文中の単語の出現頻度から不要文中の出現頻度を引いたものの値である。そのため出現した特徴語は必要文に現れやすいため、これらの特徴語を含む文が必要文だと判断できる可能性が高い。そこでこの特徴語を用い、他商品のクチコミから正確に必要な文を抽出できるのか確かめる。

3.4 特徴語の出現割合

特徴語を用いて評価用データからの必要文の抽出を試みた。その際、抽出漏れがないかの指標となる再現率と、余分なデータが入っていないか、いわゆるノイズがないかの指標となる適合率を算出した。その結果を表 2に示す。

表 2 商品B, Cの再現率, 適合率

	商品B	商品C
再現率	68.3%	57.5%
適合率	77.7%	79.3%

⁴ 2019年1月時点に公開されていた。

⁵https://www.cosme.net/product/product_id/2913937/reviews/ord/point

⁶https://www.cosme.net/product/product_id/10119418/reviews

⁷https://www.cosme.net/product/product_id/10055980/reviews

商品B, Cの適合率はどちらも7割以上となっており、商品改善に役立つような文を抽出することができた。しかし抽出漏れやノイズが入ってしまったものもある。それぞれを含めた結果を表 3, 表 4, 表 5, 表 6に示す。

表 3 商品Bの抽出漏れクチコミの例

No.	クチコミ文
1	これのテカテカ感も苦手です
2	時間が経つとケバケバの色味で、安っぽい色です
3	ん?? 全て赤っぽいというか私は好みではありませんでした

表 4 商品Cの抽出漏れクチコミの例

No.	クチコミ文
1	潤うわけでもツヤが出るわけでもなく、すぐ落ちます
2	塗りたてはツヤもあり綺麗な仕上がりがでしたがツヤはすぐに消えてしまい、表面が乾燥してきて縦じわが目立つようになりました
3	薄づきなので何度も塗り直すと、逆に蛍光っぽい色になってしまいます

表 5 商品Bの必要文と判定されたクチコミの例

No.	クチコミ文	唇	荒れる	する	なる	皮	人手によるラベル
1	リップクリームを下地にしても乾燥がひどく、皮もめくれます	0	0	0	0	1	○
2	一度使用しただけで唇は荒れるし皮は剥けるし2度と使わないリップです	1	1	0	0	1	○
3	どこのリップを使っても荒れるのですが、こちらもやっぱり、すぐに乾燥してガサガサに荒れました	0	1	0	0	0	○
4	本当に皮剥けが凄い…	0	0	0	0	1	○
5	もともと外資系の口紅や使えないメーカーもあり、唇が敏感な方です	1	0	0	0	0	×
6	私の唇には合いませんでした	1	0	0	0	0	×

表 6 商品Cの必要文と判定されたクチコミの例

No.	クチコミ文	唇	荒れる	する	なる	皮	人手によるラベル
1	特に赤みが強い色を塗布すると、ピリリと刺激を感じました	0	0	1	0	0	○
2	発色はいいのですがすぐ乾燥して、皮がペリペリ剥がれてきて唇がボロボロになりました	1	0	0	0	1	○
3	発色もよく、お手頃価格で気に入っていたのですが、どうも長時間使用した後は唇の皮がめくれてしまいます	1	0	0	0	1	○
4	塗るとすぐ乾燥して皮が剥けるので、私には合わなかったようです	0	0	0	0	1	○
5	パールやラメも一切無いのでウルルン唇にはなれません	1	0	0	0	0	×
6	ジルスチュアートやエステイローダだと光の加減で唇が不思議に輝いてたのですが…	1	0	0	0	0	×

表 3, 表 4の結果は、抽出漏れがあったクチコミ文である。表 3のNo.1に注目すると、「テカテカ感が苦手」という記述がある。苦手という言葉から使用感について不満があることが解釈できるが、ユーザがこの言葉を使用する頻度が少なかったため、特徴語とはならず抽出漏れしてしまったと考えられる。また表 4のNo.1については「落ちる」という記述があるが、頻度が少なかったため特徴語として現れず、必要文とされなかったと考えられる。他のクチコミ文についても同様の考察ができる。

商品Bの必要文と判定した中で誤判定のものが表 5のNo.5,6と表 6のNo. 5,6のクチコミ文である。それぞれが誤判定されてしまった理由については、他商品使用時の感想が記されていたため、特徴語が含まれてノイズとなってしまったことが原因であると考えられる。

だが表 5と表 6には使用感、色などにおいて具体性を持った不満な内容が記された必要文を多く抽出できている。このことから、商品改善に役立つようなクチコミ文が抽出できるようになったと結論付ける。

4. まとめ

本研究では化粧品の効率的な商品改善のため、テキストマイニングを用い、必要文と不要文の単語の出現頻度差から特徴語を選出し、その特徴語から必要文を抽出する方法を提案した。その結果、商品改善の要素となりうる、批判する文を抽出できることを示した。しかし今回の分析では商品の件数は2件と少なかった。

また必要な文を抽出する際、単語の活用形を考慮した抽出を行っていないため、文中の語と特徴語が

一致していない場合も考えられる。

以上のことから、特徴語の選出方法について見直すこと、また適合率の精度を上げる抽出条件を考察することを今後の課題とする。

参考文献

- [1] 門田拓真, 竹内孔一, “宿泊施設のリニューアルを指向した宿泊施設評判分析システムの構築”, 研究報告自然言語処理(NL), 2014-NL-219(16), pp.1-6, 2014.
- [2] 定政邦彦, 細見格, 赤峯享, 中澤聡, 石澤善雄, “原因表現抽出のマーケティング支援への応用”, 研究報告自然言語処理(NL), 2013-NL-210(1), pp.1-4, 2013.
- [3] 石田基弘, Rによるテキストマイニング入門 第2版, 森北出版, 2017.
- [4] 小林雄一郎, Rによるやさしいテキストマイニング, オーム社, 2017.