

ユーザーコメントと含有成分情報を活用した 有害化粧品成分の特定

斎藤明子[†] 岩淵莉央[‡] 中島陽子* 本間宏利* 増山 繁**

[†] 釧路高専 情報工学科

[‡] 釧路高専 電子情報システム専攻

* 釧路高専 創造工学科

** 豊橋技術科学大学 情報・知能工学系

{yoko, honma}@info.kushiro-ct.ac.jp

1 はじめに

近年、レビューサイトの普及に伴い、製品に対するユーザーコメントを製品購入前に参考にすることが一般的になった。代表的な製品レビューサイトに、家電のレビューサイトである価格 com¹、飲食店のレビューサイトである食ベログ²、化粧品のレビューサイトである@cosme³ やコスメニスト⁴ などがある。化粧品レビューサイトの@cosmeでは、ページビュー数が月間3.1億PV、サイトを閲覧しているユニークユーザー数は月間15,000,000人であり、製品の選定や情報収集の手段として利用しているユーザーが多く存在する。@cosmeサイトのユーザーの多くは、サイト独自のリコメンド機能を利用したり、「肌質」や「年齢」の類似したユーザーのレビューを参考にして、自分の肌と相性が良いと思われる製品の購入が可能である。青島 [1] と岩淵 [2] らはそれぞれ、ユーザー評価と化粧品含有成分に基づいたリコメンドシステムを開発している。

基礎化粧品においては、ユーザー個人の肌と相性のよくない製品を誤って使用することで、肌の荒れ、痒み、痛みといった皮膚異常を引き起こす恐れがある。しかしながら、ユーザーの肌質に相性が良くないと思われる製品に対しては、購入や使用に対する注意喚起を促す機能は非常に少ない。そのような製品の識別には膨大なユーザーコメントを参照する必要が生じる。また、発売直後の新製品や使用者数の少ない製品は、ユーザーレビュー数も少ないため情報に乏しく、ユーザーと相性のよくない製品の識別は困難である。

本研究では、肌質や年齢が類似したユーザー群ごと

に皮膚異常を引き起こす共通の成分、有害成分が存在すると仮定し、@cosmeサイトのユーザーレビュー情報と美肌マニア⁵サイトの化粧品含有成分情報を利用して、各ユーザー群に対する有害成分を特定した。そして、有害と判定された製品の含有成分を元に、各ユーザー群に有害と思われる製品に対して購入や使用の注意喚起を促すシステムを開発した。

2 利用サイト

2.1 @cosmeについて

@cosmeは株式会社アイスタイルが運営する、国内最大級の化粧品レビューサイトであり、クチコミと呼ばれるユーザーレビューの投稿・閲覧を可能としている。2017年9月時点で、クチコミの投稿権限を持つ@cosmeメンバーが4,500,000人、総クチコミ件数約14,000,000件、総製品数290,000商品であり、クチコミ数は年々増加している。@cosmeでは登録時に肌質・年齢などの情報を登録するため、ユーザー自身に類似した肌質や年齢のクチコミを参考にできる。ユーザーは類似ユーザーによるクチコミを製品購入の判断材料にできるほか、製品の使用方法などの参考にしている。

@cosmeのクチコミのイメージを図1に示す。おすすめ度、効果タグ、ユーザーコメントなどの評価項目が存在する。おすすめ度はユーザーが製品に対して‘0’から‘7’までの8段階で評価したものである。効果タグはユーザーが感じた製品の効果を付与できるもので、基礎化粧品であれば、“うるおい”や“美白”などのタグが存在する。効果タグ以外にも、2,000文字以

¹価格 com: <http://kakaku.com>

²食ベログ: <https://tabelog.com>

³@cosme: <http://www.cosme.net>

⁴コスメニスト: <https://cosmenist.com>

⁵美肌マニア: <http://www.bihada-mania.jp>

内のユーザーコメントを記載することができる。“うるおい”や“美白”といった効果に関しては効果タグを利用してクチコミを検索可能であるが、“肌の荒れ”や“痛み”といった皮膚異常に関するタグは存在しないため、製品の有害性を知るにはユーザーコメントを参照しなければならない。



図 1: @cosme におけるクチコミの例

2.2 美肌マニアについて

化粧品には成分の公開義務があるが、@cosme には化粧品の成分情報が記載されていない。そのため、美肌マニアという化粧品成分を公開しているサイトを利用して各化粧品の含有成分情報を獲得する。美肌マニアは化粧品の成分情報と成分の用途や毒性を公表しているサイトで、2018年1月7日現在、31,336製品の情報が公開されている。

美肌マニアにおける化粧品成分の記載例を図2に示す。成分は含有量順に記載され、成分の簡易説明が掲載されている。

スキンコンディショナー(ハトムギ化粧水)の全成分

●成分名をクリックすると、その成分を含む他の化粧品を探すことができます。

成分名	成分分析
水	溶剤
BG	保湿剤、溶剤、粘度低下剤
グリチルリチン酸2K	香料、消炎作用
ハトムギエキス	油剤、保護剤、美白性
(ステレン/アクリルアミド)コポリマー	合成ポリマー、不透明化剤
エタノール	溶剤、可溶化剤、収れん剤
クエン酸	キレート剤、pH調整剤
クエン酸Na	緩衝剤、キレート剤
EDTA-2Na	キレート剤、殺菌防腐剤
メチルパラベン	防腐剤

図 2: 美肌マニアにおける成分掲載の例

3 提案手法

本研究では、化粧品レビューサイトの化粧品に対するユーザーコメントと含有成分情報の分析により、ユーザー群ごとに有害である可能性の高い成分を抽出し、それらの成分を含有した製品を特定、警告を行うシステムを構築する。

3.1 本研究の設定

本研究では、基礎化粧品のうちもっとも使用率の高い“化粧水”を対象製品とする [3]。@cosme と美肌マニアの両サイトに存在する化粧水 2,014 種類を本研究の対象製品とした。両サイト間でのブランド名や成分名の表記揺れを手動で修正した。

@cosme のユーザー登録時に、ユーザーが設定した“肌質”と“年齢”を用いてユーザー群を設定する。ユーザーは肌質として、6種類の肌質(普通肌、脂性肌、混合肌、乾燥肌、アトピー肌、敏感肌)のうち1つを選択して登録する。本研究では、6種類の肌質のユーザーを5グループの年代(10代、20代、30代、40代、50~60代)に分割し、計30個のユーザー群を設定した。また、本研究では肌が敏感であり、皮膚異常を起こしやすい“アトピー肌”と“敏感肌”の2つの肌質カテゴリに実験対象を限定した。

3.2 有害判定コメントの抽出

ユーザーコメントのうち、製品の使用により肌の荒れや痛みなどの皮膚異常を引き起こしたと考えられるコメントを有害判定コメントと定義する。ここでは有害判定コメントを抽出する手法について述べる。

有害成分を特定するため、ある製品がユーザーに有害であったかどうかはユーザーコメントから判断する必要がある。本研究では“荒れ”、“腫れ”、“しみる”などの皮膚異常が起こったと考えられるコメント表現を有害表現と定義する。有害判定コメントの獲得において、有害表現を含むコメントを有害判定コメントとして単純に抽出すると、所望の皮膚異常が引き起こされた旨のコメントとは別に、皮膚異常が治癒した内容のコメントも一緒に抽出してしまう。

皮膚異常が発生した場合、一般的にユーザーは製品に低い“おすすめ度”評価をつけると考えられる。しかし、おすすめ度が低い場合でも、“使用感”や“価格”といった別の要因の理由で低評価をつける場合もある。そこで、有害表現とおすすめ度を併用するこ

とで、所望の正しい有害判定コメントのみの抽出を実現した。おすすめ度の閾値は4以下に設定した。これはアトピー肌の全クチコミのおすすめ度の平均値をもとに設定した。有害表現はおすすめ度4以下のユーザーコメントから人手で収集した55表現を利用する。収集した表現の例を図3に示す。収集した有害表現の中には、「ムズムズ」と「むずむず」のような表現の揺れも含まれる。

荒れ 痛い 赤み
ムズムズ むずむず
etc...

図 3: 人手で収集した有害表現の例

3.3 有害製品判定

ある製品に対し、特定のユーザー群から多くの有害判定コメントがある場合、製品はそのユーザー群に対し有害であると考えられる。このような製品を有害製品と定義する。

ある製品の総コメント数に対し、有害判定コメント数がある一定以上の割合を占める製品を有害製品として抽出する。有害判定コメントが占める割合の閾値を10%刻みで変化させ、有害製品推定の精度の差を比較した。

3.4 有害成分の特定

有害成分を特定するとき、有害製品に共通して含まれる特徴的な成分の特定が必要である。本研究では、TF-IDF の概念を応用した IF-IDF 値 [1, 2] を算出して有害成分を特定する。

IF (Ingredient Frequency)

IF は有害製品群 X に含まれる成分 i の割合である。美肌マニア上で、化粧品成分は含有量順に記載されている。記載順を重みを付与し、IF 値を算出する。

$$IF(i, X) = \sum_{i=1}^m \frac{n_{i,p} - \alpha_i}{n_{i,p}} \quad (1)$$

$n_{i,p}$: 有害製品群 X に存在する製品 p の含有成分数

m : 有害製品群 X に存在する製品数

α_i : 成分 i の製品 p における掲載順位

IPF (Inverse Product Frequency)

IPF 値は全製品に対し、ある成分 i の含まれている割合の逆数である。有害製品に限らず、有害でない製品にも多く含まれている成分は IPF 値が低く算出される。

$$IPF(i) = \log \frac{N}{pf(i)} \quad (2)$$

N : 化粧品の製品数

$pf(i)$: ある成分 i を含む製品数

IF-IPF

IF-IPF 値の導出式を以下に示す。

$$IF \cdot IPF(i, X) = IF(i, X) \times IPF(i) \quad (3)$$

3.5 含有成分による有害製品の推定

IF-IPF 値を利用した有害製品の推定手法について述べる。IF-IPF 値上位の成分は、有害製品群に含まれる成分の中でも特に特徴的な成分と考えられる。IF-IPF 値上位の成分を多数含む製品ほどユーザーに有害である可能性が高い。本研究では上位3成分の全てを含有する製品を有害製品として抽出した。

4 システム評価実験

Python 言語にて、提案手法のシステムを実装した。アトピー肌のユーザーの中で、レビュー数が5,943件と最多であった“アトピー肌20代”のユーザー群を実験対象とした。ある製品の総レビュー数に対する有害判定コメントの割合を10%ごとに区切ってそれぞれ有害成分を求めた。また、IF-IPF 上位3成分の全てを含有する製品を抽出して有害製品とした。推定した有害製品のユーザーコメントに占める有害判定レビューの割合を有害判定率とする。この有害判定率を、本システムの正当性評価に用いる。

表1は各閾値に対しての導出された有害成分を示す。IF-IPF 値上位3つの成分名とそのIF-IPF 値をまとめている。“精製水”は、閾値10~30%の場合に有害成分として判定されているが、閾値を調整し、抽出条件を厳しくすることで、精製水のような一般的な成分が有害成分として判定されなくなることがわかる。

表2はシステムの評価結果である。有害判定率上位2商品とその有害判定率を記載した。10%と50%~90%は有害製品に該当する製品が存在しなかった。

閾値 30%の有害判定率が最高で 100%ともっとも高かった。しかし、この製品のクチコミは 1 件しか存在しなかった。したがって、閾値として 40%を使用するのが適切だと結論づけた。

表 1: 有害成分の特定結果

閾値	成分名	IF-IPF 値
10%	L-アスコルビン酸 2-グルコシド	16.6931
	精製水	13.0571
	ラウリルグリコールエーテル ⁶	12.9745
20%	L-アスコルビン酸 2-グルコシド	11.9910
	精製水	11.3081
	グリチルリチン酸ジカリウム	10.6573
30%	L-アスコルビン酸 2-グルコシド	8.3937
	DPG	7.8086
	精製水	6.9845
40%	キシリトール	6.4714
	DPG	5.1271
	BG	4.1592
50%	DPG	3.4814
	ラフィノース	3.4653
	キシリトール	3.1807
60%	DPG	3.4814
	ラフィノース	3.4653
	キシリトール	3.1807
70%	セイヨウニワトコ水	2.8893
	ナイアシンミド	2.1913
	コンニャク根エキス	2.1568
80%	コンニャク根エキス	2.1568
	セトリモニウムプロミド	1.9412
	イロハモミジ葉エキス	1.7421
90%	コンニャク根エキス	2.1568
	セトリモニウムプロミド	1.9412
	イロハモミジ葉エキス	1.7421

表 2: 有害成分の特定結果

閾値	製品名	有害判定率
20%	美容液からつくった化粧水 しっとり	46%
	薬用ホワイトニング化粧水 H	44%
30%	薬用ホワイトローション	100%
	雪肌粋 薬用美白化粧水	50%
	ホワイトキャプチャーローション	50%
40%	ハイドレーティング ローション	66%
	クリーム肌水	12%

5 おわりに

本研究では、ユーザー群固有の化粧品の有害成分の特定、ユーザーに対し有害と思われる製品の推定を行うシステムを実装した。その後、推定した製品のレビューから有害判定率を求め、実装したシステムの評価を行った。

実験の結果より、一般的な成分を有害成分と誤判定しないためには、有害判定レビューの閾値を 40%に設定するのが適切だとわかった。今回は IF-IPF 値上位 3 成分全てを含む製品を有害製品と判断していた。これにより推定する製品数を絞りこめるという利点があったが、閾値 50%以上では該当する製品が存在しなかった。有害製品として判断する条件を 3 成分のうち 2 成分を含む製品に変更することで、40%以上の閾値でも有害製品を推定できる可能性があり、システムの精度向上が期待できる。

今回は手動で抽出した有害表現を使用した。低おすすめ度のユーザーコメントの特徴語を有害表現とすることで、有害製品のさらなる推定精度向上を図る予定である。

参考文献

- [1] 青島春花, “化粧品口コミサイトにおけるユーザーの評価と化粧品成分間の分析に基づくリコメンドシステムの提案,” 豊橋技術科学大学修士論文, 2016.
- [2] R. Iwabuti et al., “Proposed of the Recommender System Based on User Evaluation and Cosmetic Ingredients,” International Conference on Advanced Informatics, Bali, 2017.
- [3] ポーラ文化研究所, “スキンケア化粧品詳細調査,” 2009.