

# インターネット上の口コミサイトにおける化粧品の評判分析

白田由香利<sup>1</sup>, 橋本 隆子<sup>2</sup>, 久保山 哲二<sup>3</sup>

<sup>1</sup> 学習院大学 経済学部 yukari.shirota@gakushuin.ac.jp

<sup>2</sup> 千葉商科大学 商経学部 takako@cuc.ac.jp

<sup>3</sup> 学習院大学 計算機センター ori-cc2012@tk.cc.gakushuin.ac.jp

## 1. 始めに

近年、ブログやソーシャルネットワークなどの web サービスの普及に伴い、商品のレビューを読むことで商品購入を判断する消費者が増加している。化粧品分野においては、@コスメ [1] というサイトが注目され、そこに書かれる口コミが商品選択に影響を与えている。消費者だけではなく、化粧品を製造・販売する企業にとっても、消費者の評判情報のソースとして、その口コミ内容は価値が高い。インターネット上に書かれた口コミから商品に関する評判情報を抽出する技法は、口コミ分析技術と呼ばれ、これに関する研究・開発が盛んに行われている。我々も、価格ドットコム、twitter などの口コミを分析してきた [2-5]。

本稿では、@コスメの mascara の口コミについて、口コミ分析を行ったので報告する。マスカラとは、まつ毛を長く見せるための化粧品であり、黒く、長いほど美しく見える、と評価されている。マスカラの商品としての課題は、時間が経過すると落ちやすい、落ちた時目の周辺が黒くなり美しくない（パンダ目、と呼ばれる）、マスカラが固まってダマになり美しくない（ヒジキ状態、などと呼ばれる）、化粧落としの際に、落ちにくい、などがあげられる。

本分析は、学習院大学経済学部の白田ゼミ演習(3年)のゼミ研究として2011年度に行われた。@コスメからの口コミ情報のクロールは、学生の手作業で行われた。まず、特定の商品について、口コミ情報を収集する。そして口コミ分析ツールとして、我々が gephi[6] に付加機能をアドインした Consumer Behavior Analyzer を用いて分析した。

次章では、我々が利用した口コミ分析手法について説明する。第3章では、マスカラについて分析した結果を述べる。第4章はまとめである。

## 2. 口コミ分析手法

我々が行った口コミ分析手法を説明する。

### (A) データ収集

特定のマスカラ商品に対して、期間を固定して、その期間内に書かれた口コミを収集する。これ

は学生の手作業によって行われた。ひとつの商品に対して約 200 の口コミを集めた。以下に口コミの例を示す。全文ではなく抜粋である。

- ▷ マスカラだけはジプシー状態からどうしても抜け出せず必要以上に何本も持っている状態ですが（▽▽；唯一リピートした2本の中の1本がコレです。カールの力が一番よかったです！あと、にじみにくいカナ！
- ▷ ここの口コミを見て購入しました★とにかく使いやすいです！！コームタイプで塗りやすく、落ちにくいです（^O^） / マスカラ初めての人でも使いやすいと思いますよ私がそうだったので（σ・▽・σ）σおすすめですよ♪

価格ドットコムなどの書き込みに比較して、@コスメの口コミは、絵文字が多いことが特徴と言える。また、相対的に、非難叱責などの激しい表現はなく、明るく感じの文章で、読み手が楽しく感じられるものが多い。これが、@コスメの人気の素であるかもしれない。

## (B) 言語処理

ユーザの口コミ 1 件を 1 ドキュメントとみなして、形態素解析により、名詞、形容詞、動詞を抽出した。その結果に対して、重要なものを抽出する。各単語の重みづけ手法としては、RIDF（残差 idf, residual inverse document frequency）を用いた。一般的には、TF-IDF が広く使われているが、人間の直感に合わないという問題がある。RIDF では、ポアソン分布を利用して単語の IDF 値を推定し、実際の IDF 値との差を測ることで、重みを計算する [7]。ここで抽出された単語を、特徴語と呼ぶ。

## (C) グラフ生成

次に特徴語間の共起関係に注目して、有向グラフを作成する。キーとなる単語から、それに関連する単語をノードとして結合することで、有向グラフを作成する。有効グラフとしては廣川ら [8] が提案した概念グラフを用いる。概念グラフは、ドキュメント集合に現れる単語の共起関係と頻度情報に基づいて、語の上下関係を動的に算出する手法である。たとえば「マスカラ」がキーとなる単語だった場合、「マスカラ」をルートとして、それに共起する下位の単語を下位のノードとし、エッジで接続する。すべての上位語にエッジを描くのではなく、隣接上位にのみエッジを描く。口コミ件数の合計が、そのエッジの重みとなる。

以下、概念グラフについて簡単に説明する。

### 概念グラフ

検索対象全体のドキュメント集合を  $U$  とする。  $U$  の部分集合を  $X$  とし、  $|X|$  は  $X$  に含まれる文書の個数を表す。単語  $u, v$  について、  $df(u, X)$  は  $u$  を含む  $X$  中の文書数、  $df(u * v, X)$  は  $u$  と  $v$  の両

方を含む X 中の文書数を示す。このとき、単語の関係を以下のように定義する。

$$r(v,u)=df(u*v,X)/df(v,X)$$

ここで、 $r(v,u)>0.5$  かつ  $df(u,X)>df(v,X)$  ならば単語 u は単語 v の上位にあると考える。具体的な可視化においては、単語 v についてすべての上位語に枝を描くのではなく、その隣接上位だけとすることで枝の数を抑えている。

#### (D) コミュニティの抽出

概念グラフは口コミ文書の集合に対して、ひとつ作成される。我々は、口コミ文書の中で、どのような話題が書かれているかを知りたい。そのため、この概念グラフを話題毎に分割したい。これをネットワークからのコミュニティ抽出と呼ぶ。そこでは、コミュニティ内にエッジが多く、コミュニティ間にエッジが少なくなるような分割が望まれる。コミュニティ抽出の際、指標としてモジュラリティ Q が用いられる [9]。

$$Q = \sum_i (e_{i,i} - a_i^2) = \sum_i e_{i,i} - \sum_{i,j} (e_{i,j}^2)$$

$$a_i = \sum_j (e_{i,j})$$

ただし、 $e_{i,j}$  はコミュニティ i 内のノードと、コミュニティ j 内のノードを結ぶエッジ数の全体に対する割合である。 $\{e_{i,j}\}$  はネットワークの隣接行列で、コミュニティの分割数 k として、 $e_{i,j}$  を要素にもつ  $k \times k$  の対称行列となる。

図 1 は、コミュニティ抽出のようすを示す。ツール gephi では、ノードの大きさをノードの次数に応じて大きくなるように表示し、モジュラリティに注目したネットワーク(グラフ)分割をもとに、コミュニティごとに色を変えて表示できる。

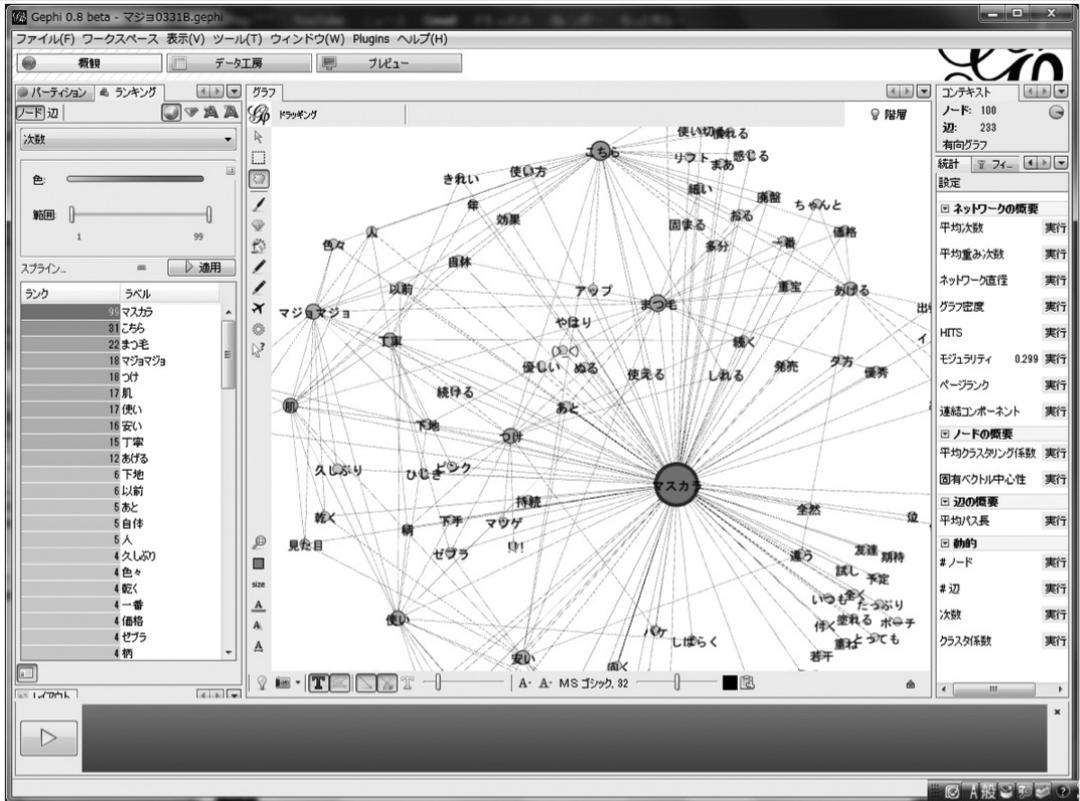


図 1 : gephi でコミュニティを色分け表示したようす。

### 3. マスカラの分析結果

前節の分析手法により、あるマスカラ X についての口コミ分析を行ったので、その結果を報告する。重要な単語（特徴語）として、100 語を抽出し、概念グラフを作成し、コミュニティを抽出した結果、6 個のコミュニティが抽出できた（図 2 参照）。

コミュニティ 1 は、一般的な表現が多かった。使ってみることへの期待、試してみる、という気持ち等が書かれている。内容として、商品に対する重要なコメントは殆ど得られなかった。

コミュニティ 2 は、下地が必要、乾く、など、肌への影響を論じている。

コミュニティ 3 は、重要語「ひじき」が表すように、長時間経過したのちに、マスカラが固まり、ひじき状態にならないで、きれいな状態が続くことを論じている。

コミュニティ 4 は、コミュニティ 3 と内容が類似している。追加されている内容として、夕方になっても落ちない（持続性の良さ）、泣いても落ちない、それなのに価格が安いのでおすすめ、などが読み取れる。利点とともに、価格の安さを賞賛する表現は、他の利点においても、頻出している。

コミュニティ 5 は、化粧品を落とす際の特徴を論じている。このマスカラが落としやすい、と

いう利点を述べている。

コミュニティ6は、マスカラの最も重視される点、まつ毛のリフトアップ機能が論じられている。マスカラでは、いかに長く太いまつ毛がクルンと上向きに上がるかが、重要となる。それをリフトアップと呼ぶ。そして、よくリフトアップできた場合、ゴージャス、という形容詞により賞賛する。

コミュニティ1				
発売	重ね	とつても	いつも	(><)
期待	友達	試し	塗れる	選ぶ
あんまり	前	付く	微妙	ぬる
多分	予定	とつても	入れる	たつぷり
ポーチ	マスカラ	全く	優しい	使い
下手	柄	使い方	しばらく	使える
戻る				
コミュニティ2				
下地	久しぶり	色々	マジヨマジヨ	乾く
以前	コツ	普段	自体	肌
人	見た目			
コミュニティ3				
ピンク	マツゲ	丁寧	!!!	続く
年	つけ	全然	ひじき	きれい
しれる	あと	続ける		
コミュニティ4				
出会う	絶対	まつ毛	細い	優秀
イメージ	一番	価格	嬉しい	付け
持続	おすすめ	ちゃんと	泣	夕方
向く	違う	束	重宝	あげる
コミュニティ5				
めっちゃ	最初	ゼブラ	必要	落とし
失敗	落とせる	バケ	安い	オススメ
コミュニティ6				
固まる	廃盤	おる	ロコミ	慣れる
使い切る	こちら	やはり	アップ	効果
感じる	リフト			

図2：抽出された6個のコミュニティ

#### 4. まとめ

本研究では、化粧品のマスカラの口コミ分析を行った。口コミ分析手法としては、概念グラフによるグラフ作成、モジュラリティによるコミュニティ抽出を用いた。また、gephiによるコミュニティの色分け表示およびグラフのレイアウト表示、などの可視化を行った。我々がgephiを元に作成したConsumer Behavior Analyzerでは、時間の要素を入れた口コミ分析が行える。我々は、Consumer Behavior Analyzerを用いて、東日本大震災に関するSNSの分析において、話題がどのように変遷したか、この時系列分析機能を用いて分析した[2]。化粧品の口コミ分析においても、新商品による既存商品への影響の時系列変化、消費者のニーズの時系列変化などが分析可能である。今後は、そうした事柄も分析していきたい。

## 謝辞

2011年度、経済学部演習3年のクラスで、マスカラ等の口コミ分析を行ってくれた白田ゼミ3年生の諸君の協力を感謝します。本研究の一部は、学習院大学計算機センター特別研究プロジェクト「概念グラフを用いた化粧品のニーズ分析」により支援を受けました。ここに記して謝意を表します。

## 参考文献

- [1] "@cosme," <http://www.cosme.net/>.
- [2] T. Hashimoto, T. Kuboyama, and Y. Shirota, "Topic Detection about the Great Japan East Earthquake based on Emerging Modularity," *Proc. of 22nd European Japanese Conference on Information Modelling and Knowledge Bases, Prague, Czech Republic, June 4-9, 2012* (to appear) .
- [3] T. Hashimoto, T. Kuboyama, and Y. Shirota, "Rumor Analysis Framework in Social Media," *Proc. of IEEE TENCON2011, pp. 331-333, Indonesia, 2011.*, 2011.
- [4] T. Hashimoto, T. Kuboyama, and Y. Shirota, "Consumer Behavior Analysis from Buzz Marketing Sites over Time Series Concept Graphs," *A. Konig et al. (Eds.), KES 2011 (15th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems), Part II, LNAI 6882, Springer, Heidelberg, pp. 73-83, 2011.*, 2011.
- [5] 橋本隆子, 久保山哲二, 白田由香利, "ソーシャルメディアを対象としたマーケティング解析 - 時事問題をきっかけとした想定外の消費行動抽出 -," *学習院大学経済論集, Vol.47, No.4, 2012年1月*, pp.263-280, 2011.
- [6] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: An open source software for exploring and manipulating networks," *International AAAI Conference on Weblogs and Social Media, 2009.*
- [7] 北研二, 津田和彦, 獅々堀正幹著, *情報検索アルゴリズム*: 共立出版, 2002.
- [8] Y. Shimoji, T. Wada, and S. Hirokawa, "Dynamic Thesaurus Construction from English-Japanese Dictionary," *The Second International Conf. on Complex, Intelligent and Software Intensive Systems*, pp. 918-923, 2008.
- [9] 林幸雄 (編), 大久保潤他, *ネットワーク科学の道具箱: つながりに隠れた現象をひもとく*: 近代科学社, 2007.