

## クチコミ情報からの関連商品マイニング

島松, 千春

九州大学大学院システム情報科学研究府 | 九州大学システム情報科学研究院

御手洗, 秀一

九州大学情報基盤研究開発センター | 九州大学システム情報科学研究院

伊東, 栄典

九州大学情報基盤研究開発センター | 九州大学システム情報科学研究院

廣川, 佐千男

九州大学情報基盤研究開発センター | 九州大学システム情報科学研究院

<http://hdl.handle.net/2324/17046>

---

出版情報 : 電子情報通信学会第19回データ工学ワークショップ (DEWS 2008), 2008-03. 電子情報通信学会

バージョン :

権利関係 :

## クチコミ情報からの関連商品マイニング

島松 千春<sup>†</sup> 御手洗 秀一<sup>‡</sup> 伊東 栄典<sup>‡</sup> 廣川 佐千男<sup>‡</sup>

<sup>†</sup>九州大学システム情報科学研究院情報理学専攻 〒812-8581 福岡市東区箱崎 6-10-1

<sup>‡</sup>九州大学情報基盤研究開発センター 〒812-8581 福岡市東区箱崎 6-10-1

E-mail: <sup>†</sup> c-shima@i.kyushu-u.ac.jp, <sup>‡</sup> {mitarai, itou, hirokawa}@cc.kyushu-u.ac.jp

**あらまし** 近年, blog, wiki, SNS 等の普及により一般利用者がコンテンツ作成者となる Web2.0 と総称される動きを作りだしている. その中で, 様々な商品に対して個人が自分の嗜好・評価を投稿するクチコミサイトも増加している. クチコミには大衆の感性に基づくデータが蓄積されているため, そこからマーケティングや商品推薦のための情報発見システムが検討されている. 我々もクチコミサイトに着目し, 商品間の関係発見についての研究を行っている. 本研究では, 概念束を用いて商品間の関係を発見する手法を提案する. 概念束は 1979 年に Rudolf Wille らが形式概念解析として提案したもので, 対象(ここでは商品)と対象が持つ属性との関係を行列で表現し, そこから対象間または属性間の関係を分析する. また, 商品間の関係分析結果に基づいて商品提案を行う手法を検討した.

**キーワード** クチコミ, 推薦, 検索, 概念束, 関連商品分析

## Discovery of Related Items from Word of Mouth

Chiharu SHIMAMATSU<sup>†</sup> Shuich MITARAI<sup>‡</sup> Eisuke ITO<sup>‡</sup> and Sachio HIROKAWA<sup>‡</sup>

<sup>†</sup> Dept. of Informatics, Kyushu University 6-10-1 Hakozaki, Higashi-ku, Fukuoka, 812-8581 Japan

<sup>‡</sup> Research Institute for IT, Kyushu University 6-10-1 Hakozaki, Higashi-ku, Fukuoka, 812-8581 Japan

E-mail: <sup>†</sup> c-shima@i.kyushu-u.ac.jp, <sup>‡</sup> {mitarai, itou, hirokawa}@cc.kyushu-u.ac.jp

**Abstract** Recently, the Web2.0 movement changes a common user to a contents creator using CGM (contents generating media) such as blog, wiki, and SNS. Word of mouth services are becoming popular, and individuals submit their preference and evaluation to various items on those sites. Reputations by means folk are being accumulated and some IR systems are emerging to analyze them for marketing and recommendation. The authors pay attention to a word of mouth site, and consider the relation discovery problem between items. In this paper, we propose a technique to discover relations between items using the formal concept analysis and the concept lattice. The formal concept analysis is proposed by Rudolf Wille et.al. in 1979. The relation between objects (items, in this paper) and the attributes of the objects is represented by the concept matrix. The hierarchy of items and attributes are represented by a concept lattice. We studied a technique of item relation analysis using concept lattice technique. Moreover, we propose an item recommendation technique based on word of mouth data and concept lattice.

**Keywords** word of mouth, recommendation, information retrieval, concept lattice, relation discovery

### 1. はじめに

Web 上に莫大な情報があふれ, 情報過多という状況に陥っている. 混在するそれらの情報から, 目的に適合するものを探し出すために, 検索エンジンや情報推薦システムの開発が活発に行われてきた. 検索とは, データ集合から目的のデータを探し出すことである. 推薦とは, 利用者のプロフィールなどに基づいて利用者に適した情報を提示することである. しかしながら, 検索したいものが一つに定まっていない場合や, 利用者のプロフィールが複数の側面を持っている場合には検索や推薦が有効とは言い難い.

近年, blog, wiki, SNS 等の普及により一般利用者

がコンテンツ作成者となるメディア (CGM: Consumer Generated Media) が増大しており, それらは Web2.0 と総称され動きを作りだしている. その中で, 様々な商品に対して個人が自分の嗜好・評価を投稿するクチコミサイトも増加している. クチコミには大衆の感性に基づくデータが蓄積されているため, そこからマーケティングや商品推薦のための情報発見システムが検討されている. 我々もクチコミデータから商品間の関係を発見する研究を行っている.

人は多面性を持っており, 同じ人物であっても状況や服装, 気分, 季節等によって求める商品は異なる. 利用者が過去の嗜好と異なる商品を求める場合, 過去

の履歴に基づく推薦や協調フィルタリングでは多面性を考慮した商品推薦は困難である。

検索は目的が明確に分かっている場合に有効な手段であるため、「なんとなくこんなイメージのものが探したい」という目的には適していない。自分の目的に適した商品が推薦されない場合、利用者は検索で商品を探そうとする。一般に利用者は求める商品について漠然としたイメージしか持たないため、イメージから連想される単語を検索語として入力する。しかし、商品の説明テキストに検索語を含む商品が提示されるだけであり、それらの商品の一つ一つを確認するのは手間のかかる作業である。

本研究では、多くの利用者が商品に対して意見を投稿したクチコミデータから、概念束を用いて商品間の関係を発見する手法を提案する。概念束は 1979 年に Rudolf Wille ら[1]が形式概念解析として提案したもので、対象(ここでは商品)と対象が持つ属性との関係を行列で表現し、そこから対象間または属性間の関係を分析する。クチコミが持つ情報から商品間の関係分析ができれば、その集合を目的やイメージから連想される単語で検索するシステムを構築できる。

本論文の構成を述べる。第 2 節では関連研究について述べる。第 3 節では、本研究で提案する商品間の関係分析手法について述べる。第 4 節では、本研究が対象としたデータと提案手法の適用結果について述べる。また結果についての議論と考察を行う。最後に第 5 節でまとめと今後の課題を述べる。

## 2. 関連研究

クチコミに関する研究は 1950 年代から取り組まれている。Arndt[2]は非公式な会話を通して商品やブランドに対する意見が広まる現象の広告効果を強調した。Katz and Lazarsfeld[3]は、クチコミが消費者の行動に対し大きな力を持つことを発見した。消費者が行動決定を行う際、クチコミが重要な役割を果たすことは、消費者行動研究において広く認められている[4]。CGM の増大からクチコミは個人の Web サイトや blog 上で出現することも多くなった。blog などから文章を解析し[7]、クチコミのマイニングを行う研究[8]が行われている。

一般利用者からの評価値を用い商品の推薦を行う協調フィルタリングについての研究開発が進んでいる。推薦システムとは大量の対象の中から、個々の利用者にとって適当と思われる対象を候補として提示するシステムである。GroupLens[9]や Balabanovic ら[10]の web ページ推薦システム Fab, Shardanand ら[11]の音楽推薦システム Ringo などがあげられる。これらは対象に対する評価値のみを用いて推薦を行っている。しかし、

クチコミの文章そのものを分析し商品提案を行う研究は少ない。

## 3. クチコミに基づく商品間の関係分析手法

ここでは、クチコミ文章の調査結果と、本論文で提案する商品間の関係分析手法について述べる。

### 3.1. クチコミ文章の調査

本研究ではクチコミ文章の情報に基づいて商品の関係分析を行う。そこで、クチコミ文章を調査した。複数のクチコミサイトで調査した結果、クチコミ文章には、その商品を使った感想や特徴、使い方などが書かれていることが分かった。当然ともいえる情報以外に、サイト提供者側で用意された商品カテゴリ分類にこだわらない情報や、製造社からの商品説明に明示されていない特徴や利用目的、使い方、商品のイメージがクチコミ文章には記述されていた。

クチコミ文章を解析することで、製造社や商品カテゴリにとらわれない、利用者の観点からの商品間の関係分析ができると考えられる。我々は、クチコミ文章を解析し、その文章に基づいて商品間の関係を解析する方法を検討した。

### 3.2. 単語の出現頻度および TF-IDF 値

クチコミ文章を解析するため、商品の特徴を表す単語、利用者がその商品を使った目的を表す単語を調査する。文章の単語解析のために、形態素解析ツールとしては、Chasen<sup>1</sup> や MeCab<sup>2</sup>がある。これらのツールを利用することで、単語での解析が可能になる。

ある商品が、ある目的に使われる場合、その目的について商品を紹介するクチコミ文章には、特定の単語群が頻出すると考える。たとえば、レストランを料理の種類・価格帯・立地で紹介するのではなく、結婚式の披露宴会場として紹介する場合などは、目的に応じた文章が記述される。

そこで、ある商品に対するクチコミ文書群から出現単語を抽出して解析すれば、その単語内には商品の利用目的やイメージを表す単語が発見できると考えられる。文書群と単語の関係を算出する手法に TF-IDF 値の計算がある[13]。商品ごとにクチコミ文書群を解析し、各単語の TF-IDF 値を算出する。TF-IDF 値の計算方法を以下に示す。

$$w_{i,j} = tf_{i,j} \log \left( \frac{N}{df_i} \right)$$

$tf_{i,j}$  = 単語  $i$  が文書  $j$  で出現した頻度

<sup>1</sup> <http://chasen.naist.jp/hiki/ChaSen/>

<sup>2</sup> <http://mecab.sourceforge.net/>

$$df_i = \text{単語 } i \text{ を含む文書の総数}$$

$$N = \text{全文書数}$$

一部のクチコミ文書例で解析したところ、形態素解析で出力される一意な単語の数は、後の関係解析処理が困難になるほど膨大であった。解析の際、品詞が名詞・動詞・形容詞・副詞の単語だけに制限することや、自立語の単語のみに制限することなど、解析対象とする単語数を減らす工夫が必要であろう。

### 3.3. 概念束を用いた商品間の関係分析

我々はクチコミ文章に基づく商品間の関係分析手法として、概念束を用いる手法を提案する。概念束は1979年に Rudolf Wille ら[1]が形式概念解析として提案したものである。形式概念解析は、束論に基づいた理論であり、データに隠れた理論的な構造について束を通して解析する手法である。また、形式概念の間の関係を数学的構造として表現したものを概念束という。

概念束を使えば属性による分類が可能である。そのため、概念束はデータマイニング、情報検索、知識管理などのように、データに潜むパターンや関係を発見するために用いられてきた。

概念束を作るには、対象と属性の関係を表した行列（概念行列）が必要である。概念行列の要素値は 0,1 の 2 値である。対象  $i$  が属性  $j$  を持つとき、行列の  $i,j$  要素は 1 の値を持つ。

本研究では、商品の分類と利用者への推薦を目的とする。そのため、概念行列における対象は、商品または利用者となる。関係を分析するには、対象である商品あるいは利用者の特徴付ける属性として何を選択するかが問題となる。

我々はクチコミ文章に着目して、商品进行分析することを検討している。そこで、以下に述べる二つの手法を提案する。

#### 3.3.1. 方式 1：特徴語による概念束

方式 1 は、クチコミ文章群から抜き出した特徴語を、概念行列の属性として用いる。クチコミ文章群を解析し、商品  $i$  のクチコミ文章群において単語  $j$  の TF-IDF 値が高ければ、単語  $j$  は商品  $i$  を特徴付ける単語と考えられるため、商品と単語の関係を概念行列にし、行列の  $i,j$  要素を 1 とする。

方式 1 をより詳細に述べる。最初に、解析対象とする商品を、クチコミが  $n$  件（たとえば 10 件）以上投稿されている商品だけに制限する。クチコミ投稿数の少ない商品は一般的な意見が反映されていないと考えられる。また、クチコミ件数が少数の場合、特異な意見が大きな影響を与えてしまうためである。

次に、ある商品についての特徴語を算出する。ある

商品に対するクチコミ全件を結合したものを 1 文書とする。この結合を 10 件以上のクチコミが投稿されている商品全てについて行う。クチコミ文章の結合を全商品について行うと、商品数と同じだけの結合クチコミ文書が作成される。これを集めて全文書集合とし、文書に出現する単語について、商品ごとに各単語の TF-IDF 値を算出する。TF-IDF 値が閾値以上の単語を、その商品の特徴語とする。

最後に、商品と単語との関係を概念行列で表現する。商品  $i$  について、単語  $j$  の TF-IDF 値が閾値以上なのは  $i,j$  要素が 1、閾値未満ならば 0 とする行列を作成する。この行列に基づいて概念束を作成する。出来た概念束は、クチコミ文章に基づいた商品間関係を表現する。

前述の方式 1 は、商品を利用者と置き換えても同様に計算可能である。投稿件数が 10 件以上の利用者に制限して、クチコミ文章を利用者毎に結合し、その利用者のクチコミ文書を作成する。それらから利用者ごとに各単語の TF-IDF 値を算出する。TF-IDF 値が閾値以上の場合、その利用者の特徴語として概念行列を作成する。

#### 3.3.2. 方式 2：目的やイメージを表す単語による概念束

方式 2 では、クチコミ文章内に記述されている商品の利用目的、利用場面、商品に対するイメージを表す単語を用いる。これらの単語を概念行列の属性として用い、商品と単語間の関係から概念束を作成する。ある単語が閾値以上の割合で出現する場合、その商品はその単語を属性に持つとする。

目的やイメージを表す単語の抽出方法については、決まった方法は無い。商品の分野ごとに違った単語になると考えられる。例えば商品が自動車である場合、レジャー、家族サービス、通勤、などが目的を表す単語として出現するであろう。一方、化粧品やアクセサリの場合、癒し系、結婚式向け、就職活動用、といった単語が目的を表す単語になるだろう。

上記のような目的やイメージを表す単語をクチコミの文章から抜き出す。商品  $i$  のクチコミ文章内で単語  $j$  の出現頻度が高い場合、行列の  $i,j$  要素を 1 とする商品と単語の間の概念行列を作成する。この行列に基づいて概念束を作成する。この概念束は、商品の利用目的として考えている事や、商品に対する利用イメージなどに基づいた商品間関係を表現することができる。

## 4. おわりに

本研究では、様々な商品に対して個人が自分の嗜好・評価を投稿するクチコミサイトデータから、商品

間の関係を抽出する手法について研究した。クチコミには大衆の感性に基づくデータが蓄積されているため、そこからマーケティングや商品推薦のための情報発見システムが検討されている。新たな関係解析手法として、概念束を用いて商品間の関係を発見する手法を提案した。クチコミの文章に対して形態素解析を行い、TF-IDF 値を計算し、概念束を用いて分析を行う手法を提案した。

今後は前述の関係解析方法を具体的なデータに適用する。また、出力された概念束を用いた検索・推薦システムを構築する。このシステムの概要は以下の通りである。まず、利用者が検索語を選択する。次に概念束上で、検索語を含む最も下位のノードを表示する。このとき表示される情報は、ノードに存在する商品、上下に位置する概念との差分である。差分を表示することで、検索結果の絞り込み・拡張の支援が可能になる。目的の商品集合が得られるまでノードを移動する。実際の検索にこのシステムを組み込めば、これまでとは違った観点での検索が可能になる。

## 文 献

- [1] Bernhard Ganter, Rudolf Wille, C Franzke, "Formal Concept Analysis: Mathematical Foundation," ISBN-13: 978-3540627715, Springer, 2005.
- [2] Arndt, Johan, "Word of Mouth Advertising: A Review of the Literature," The Advertising Research Foundation, INC., 1967.
- [3] Katz, E. and P. F. Lazarsfeld, "Personal Influence: The Part Played by People in the Flow of Mass Communications," The Free Press, 1955.
- [4] Brown, Jacqueline-Johnson and Peter H. Reingen, "Social Ties and Word-of-Mouth Referral Behavior," Journal of Consumer Research, Vol.14, pp.350-62, Dec.1987.
- [5] 稲葉 真純, 長野 伸一, 長 健太, 溝口 祐美子, 川村 隆浩, "CGM 分析技術の現状と課題," 人工知能学会研究会資料, SIG-SWO-A603-06, Mar.2007.
- [6] 安藤和代, "相互作用性に着目したクチコミ研究の系譜と展望: 関係性パラダイムでのクチコミ活用に向けて," 商学研究科紀要, pp.43-57, Nov.2004.
- [7] 鍛冶伸裕, 喜連川優, "文構造を考慮した評価文書分類のための確率モデル," 電子情報通信学会第 17 回データ工学ワークショップ第 4 回日本データベース学会年次大会, 5A-i4, Mar.2006.
- [8] 斉藤典明, 菅野元之, 佐野直美, "インターネット上の口コミ解析手法の提案," 情報処理学会研究会報告, GN45-12, pp.65-70, Nov.2002.
- [9] Poul Resnick, neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, John Riedl, "GroupLens: An Open Architecture for Collaborative Filtering of News," In Proc. of ACM 1994 Conference on Computer Supported Cooperative Work(CSCW'94), pp.175-186, 1994.
- [10] Marko Balabanovic, Yoav Shoham, "Fab: Content-Based, Collaborative Recommendation," "comm.of the ACM, Vol.40, No.3, pp.66-72, 1997.
- [11] Upendra Shardanand, Pattie Maes, Social Information Filtering: Algorithms for Automating "Word of Mouth", ISBN:0-201-84705-1, ACM Press/Addison-Wesley Publishing Co, pp.210-217, 1994.
- [12] V. Choi, "Faster Algorithms for Constructing a Galois/Concept Lattice," Presented at SIAM Conference on Discrete Mathematics, 2006.
- [13] 北研二, 津田和彦, 獅々堀正幹, "情報検索アルゴリズム," 共立出版, 2002. (ISBN4-320-12036-1).