

推薦論文

ブログに記述された不満表現からの潜在ニーズの発見

坂井 俊之^{†1} 藤村 考^{†1}

本論文は、ユーザが記述したテキストからユーザの不満を読み取り、その不満を解決する商品を見出すことを目的としている。そのために、ユーザの実体験が記述されたブログのテキストを大量に収集し、その中から不満表現の抽出と、商品と不満表現との関係性抽出を行う。これらの実体験に基づいた知識を利用することで、ユーザの様々な不満を解決する商品の提示が可能となる。そして、これらの不満表現抽出と関係性抽出について検討した結果、それぞれの抽出において、ネガティブな状態をポジティブな状態に変化させる動作として「改善動作」を新たに定義し、その共起情報を利用することで、抽出精度を向上させることができるという知見が得られた。また、この知見に基づき、ユーザの不満に対して不満を解決する商品を提示するシステムのプロトタイプを作成し、いくつかの不満に対して、複数の商品から不満を解決する商品を選択し、提示できることを確認した。

Discovering Latent Solutions from Expressions of Dissatisfaction in Blogs

TOSHIYUKI SAKAI^{†1} and KO FUJIMURA^{†1}

This paper aims to find the techniques or goods that solve user's problems or dissatisfaction extracted in texts created by the user. We collected a large number of texts describing user experiences from blogs to extract expressions of dissatisfaction. These texts also contain information about the techniques or goods that solve the dissatisfaction and about their effectiveness. We found that the co-occurrence frequency of words that indicate problem prevention or solution, such as "protect" and "cure", is an effective measure for realizing these extractions with high accuracy. We implemented a prototype system on the proposed method and tested it. The results show that the system can identify some useful goods for solving user's problems extracted in the texts.

1. はじめに

近年、ブログや SNS, Twitter 等のソーシャルメディアの普及と情報検索技術の発展により、インターネット上に提供される情報がユーザの購買行動に大きな影響を与えるようになってきた。このような中で、インターネット上に蓄積されるユーザの情報を活用することによって、ユーザに適した商品情報を発見し、広告や推薦を行うような試みが行われている。その中で代表的なものとしては以下があげられる。

- (1) ユーザの検索語に関連する情報を提供する。
- (2) ユーザがインターネット上で公開しているテキストに関連する情報を提供する。
- (3) ユーザの過去の購買行動に関連する情報を提供する。
- (4) 他のユーザの (1)~(3) の情報を利用して情報を提供する。

この中で、(1) は Adwords¹⁾, (2) は Adsense²⁾, (3) は Amazon³⁾ による商品推薦 (広告提示) が有名である。また、(4) は (1)~(3) と併用されることで提供する情報の精度が向上する。

しかし、上記のいずれも、スポーツ、音楽、ファッションといったような「ユーザの興味」や「文書の主題」をターゲットとした情報提供であり、「潜在的にユーザがかかえる不満」を解決する商品情報の提供を目的としたものではない。本論文では、この「潜在的にユーザがかかえる不満」を解決する商品情報の提供を目的とする。「潜在的にユーザがかかえる不満」とは、花粉症が辛い、アプリケーションの動作が遅い、部屋が散らかっている、といったような、「漠然となんとかしたいという思いをかかえているが、解決のための具体的な行動に結び付いていない不満」を指す。このような不満に対し、具体的な商品を示すことで、ユーザが簡易に不満を解決できたならば、非常に有益であると考えられる。応用例として、ブログに記述された不満に関し、ブログの著者に対して商品広告を提示する例を図 1 に示す。近年、ブログや Twitter をはじめとするソーシャルメディアのように、自己の備忘録あるいはライフログという位置づけのソーシャルメディアは増加してきている。これらの最大の読者は第三者ではなく、著者であるケースも少なくない。そこで、これらのメディアにおける広告配信のターゲットとしては、第三者だけではなく、著者も視野に入れること

^{†1} 日本電信電話株式会社 NTT サイバーソリューション研究所
NTT Cyber Solutions Laboratories, NTT Corporation

本論文の内容は 2010 年 5 月のグループウェアとネットワークサービス研究会にて報告され、同研究会主催により情報処理学会論文誌ジャーナルへの掲載が推薦された論文である。

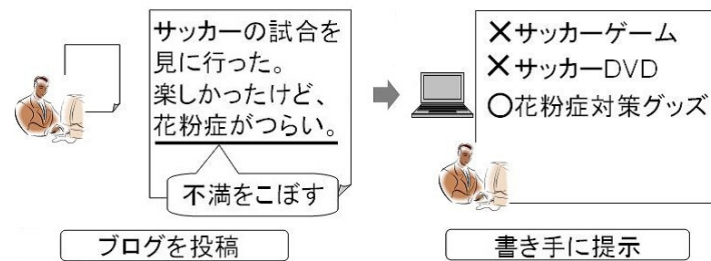


図 1 不満に対する商品広告配信

Fig. 1 Advertisement to the author of a blog.

が重要である。

そこで本論文が提案する商品情報が提供されるまでの概略を示す。商品情報の提供までの流れとしては、大きく以下の3段階で行う。

- (1) ユーザの不満表現を検出する(不満表現検出)。
- (2) 検出したユーザの不満表現と商品の関係性の強さを抽出(関係性抽出)。
- (3) (2)の結果から、ユーザの不満を解決する商品を選択(商品発見)。

このうち、(1)の不満検出には、ユーザがインターネット上で公開している情報として、本論文ではブログ記事を利用する。たとえば、以下のような例を考える。

昨日、サッカーの試合を見に行った。やはりサッカーの試合はTVで見るより、実際に見に行ったほうが良い。自分の応援していたチームが2-0で勝ったこともあり、とても楽しかったが、花粉症がつかった。

このような、「花粉症がつかった」という不満表現をユーザのブログ記事から検出する。AdSenseのようなトピックを検出する方式では、サッカー等のトピックに関する単語が検出されてしまい、花粉症という単語が検出されないため、新たに不満表現を収集した辞書を用意し、その辞書とのマッチングによって不満表現を検出する。また、(2)の関係性抽出では、インターネット上に記述された人々の実体験を利用する。つまり、ユーザが記述したブログ等のWebテキストの中には、「花粉症が辛い」といったような不満表現に対し、「～によって解決した」等の、すでに何らかの解決策にたどり着いた人が発信した商品情報も存在するため、これらの情報を処理することにより、不満表現と商品の関係性の強さを抽出す

る。最後に、(3)の商品発見では、(2)の関係性抽出を様々な商品に対して適用した結果から、ユーザがかかえる不満を解決するのに適した商品のランク付けを行う。

本論文では、ブログから不満表現の検出、関係性の抽出を行い、既知の商品リストに記述された商品のランク付けに関して検討を行った。その結果、不満表現検出や関係性抽出においては、ネガティブな状態をポジティブな状態に変化させる動作である「改善動作」との共起情報を用いることが有効であるという知見を得ることができた。また、この改善動作を用い、2008年5月11日から2008年8月26日までの健康商品に関するブログ記事(35,267件)を分析した結果、不満を解決する商品が発見可能であることを確認した。ここで、商品のランク付けに関しては、関係性抽出で抽出した関係性の強さを降順に並べ、ランク付けを行った。

以下に、本論文の構成を示す。まず、2章では、関連研究について述べる。次に、3章では、「不満表現検出」、「関係性抽出」、「商品発見」に関する本論文のアプローチについて述べる。4章では、まず、「不満表現検出」と「関係性抽出」を特定の商品に対して適用した場合の抽出精度の評価結果について述べ、そして、健康商品推薦のプロトタイプシステムへ適用した結果について述べる。最後に、5章でまとめを述べる。

2. 関連研究

本論文の提案内容はサービスという観点から情報推薦に分類される。情報推薦の技術としては、これまで協調フィルタリングやコンテンツフィルタリング、そして、これらのハイブリッド方式が提案されてきた⁴⁾⁻⁶⁾。それぞれについて、以下に述べる。

協調フィルタリングは、アイテムの内容(商品の説明文等)は考慮せず、アイテムに対するユーザの評価点のみを考慮する方式である。ここで、評価点とは、商品の購入履歴やページの閲覧履歴、アンケート結果等を指す。協調フィルタリングの方式としては大きく以下の3種類に分かれる。

- ユーザベース方式：推薦対象者とアイテムへの評価点のつけ方が似ている他のユーザを発見し、そのユーザが高い評価点をつけるアイテムを推薦する。
- アイテムベース方式：推薦対象者がアイテムAに対し高い評価をつけていると仮定する。他のユーザからの評価の傾向がアイテムAとアイテムBで似ている場合、アイテムBを推薦する。
- モデルベース方式：クラスタリングやベイジアンネット、EMアルゴリズム等を用いてユーザやアイテム間の関係をモデル化しておく。ユーザと似たモデルのアイテムを推薦

する。

コンテンツフィルタリングは、アイテムの内容とユーザプロフィールを比較し、ユーザプロフィールと近い内容のアイテムを推薦する方式である。なお、ユーザプロフィールとは、ユーザの嗜好や興味を抽出し、モデル化したものを指す⁹⁾。コンテンツフィルタリングの方式としては大きく以下の3種類に分かれる。

- ルールベース方式：ユーザプロフィールに対し、人手で作成したルールに基づいてアイテムを推薦する方式。
- メモリベース方式：アイテムの内容とユーザプロフィールをそれぞれベクトルで表現し、ベクトルの類似度が高いアイテムを推薦する方式。
- モデルベース方式：アイテムの内容に対し、ユーザが正負の評価をつけて機械学習を行うことでユーザプロフィールを作成する。作成したユーザプロフィールに従ってアイテムを推薦する。

ハイブリッド方式は、アイテムに対するユーザの評価点と、アイテムの内容をあるモデルの下で統合することで推薦を行う方式である。モデルの学習には最大エントロピー法⁷⁾やEMアルゴリズム⁸⁾等が用いられる。

本論文で行う推薦は、ユーザが具体的な行動を起こしていない場合を想定しているため、ユーザの不満を解決するような商品に関しては過去の購買履歴が存在しない(あるいは、少ない)と考えられ、協調フィルタリングは不向きである。そこで本論文では、メモリベース方式のコンテンツフィルタリングを用いている。

コンテンツフィルタリングを適用するためには、ユーザプロフィールとアイテムの内容の抽出が必要である。一般的には、ユーザプロフィールとしては嗜好を抽出するが、本研究では不満という新たな種類のプロフィールを抽出している。これにより、ユーザに対して推薦できる商品の幅が広がることが期待できる。

ユーザの不満のように、人や物に対する「不評」を抽出する技術としては、Web上の文書から人や物に対する評判を抽出し、「好評」か「不評」かを判定する技術がある¹⁰⁾⁻¹²⁾。これらの方式では、まず「悪い」等の不評であることがはっきりしている評判を用意しておく。そして判定対象の表現と、その周囲に存在する「悪い」等の不評表現との間の関係性を用いることで「好評」、「不評」を判定する。文献12)では、本論文と同じようにネガティブな名詞の判定もしているが、本来の目的が節の全構成要素のpn判定であるため、係り受けの判定や節どうしの因果関係の判定を行う必要がある。しかし、本論文の目的としては、節の構成要素の一部である動作名詞の対象がネガティブであるかどうかを判定できればよい

ため、精度を下げる可能性のある要素(係り受けの判定や因果関係の判定)を排除し、問題を単純化できる。具体的には、節全体がポジティブであることに限定し、また、節の構成要素の一部である動作名詞も、「解消」や「予防」のように、ネガティブな名詞を対象に持ち、かつ、節全体の極性をポジティブに反転させるようなものに限定している。

また、本論文と同じように「不評」と共起しやすい述語を用いて「不評」を抽出する技術も存在する¹³⁾。文献13)では、目的が名詞のpn判定であるため、汎用的なモデルとして様々な述語を用いている。これに対し、本論文では不満表現の抽出を目的としている。そのため、ポジティブな極性の名詞を極力排除できるように、ネガティブな名詞をポジティブな状態に変化させるような述語、つまり、ポジティブな名詞を対象としない、「解消」、「予防」等の動作名詞を用いている。

以上をふまえ、本研究では商品によって実際に不満が解決されたときに、ユーザがブログ記事等にその体験を記述した際の文の構造を手がかりに不満表現を抽出する方式を提案する。

また、アイテムの内容の抽出では、{アイテム, 不満}というペアの中から{アイテム, アイテムが解決する不満}という関係性を抽出する必要がある。不満に関する関係性抽出技術としては、モノとトラブルの間関係性を抽出する技術がある¹⁴⁾。これは、トラブルを不満と置き換えると、{アイテム, アイテムが引き起こす不満}という関係性抽出であり、アイテムの内容の抽出としてはそのまま利用することができない。

そこで本論文では、ユーザが記述した体験から、不満が解決されたことを表す表現との共起情報を用いてアイテムを抽出する方式を提案する。このように、個人の体験を利用する研究としては文献15)が存在するが、本研究と文献15)では以下のように個人の体験の利用目的が異なる。

- 本研究では個人の不満を解決する商品の推薦メカニズムを発見することを目的としている。
- 文献15)は個人のエピソードに関する意味解析を行い、個人の体験に関して有用な情報を検索可能とすることを目的としている。

そのため、文献15)では具体的な推薦メカニズムが提示されておらず、この手法で推薦が可能かどうかは明らかになっていない。それに対して、本研究では具体的な推薦メカニズムを示したうえで、不満を解決する商品の推薦が可能であることを示している。このような研究は筆者の知る限り過去には存在していない。

3. アプローチ

3.1 処理の流れ

不満を解決する商品を発見する際の手順は以下ようになる(図2)。

- 手順1 ユーザが記述したブログ記事集合から不満表現を抽出し、データベースに保存する(不満表現辞書の作成)。
- 手順2 商品名のリストと不満表現辞書を用いて、商品に関するブログ記事集合から、{商品名, 不満表現}の共起ペアを抽出する。その中から「商品が不満を解決する」という関係性を満たす共起ペアを抽出し、抽出した件数とともにデータベースに保存する(関係性の抽出)。
- 手順3 商品情報を提示する対象者のブログ記事から、データベースに保存された不満表現を用いて不満表現検出を行う(不満表現検出)。
- 手順4 検出した不満表現に対応する商品名をデータベースから抽出し、手順2で{商品名, 不満表現}のペアとともに保存した抽出数を用いて、提示する商品のランキングを行う

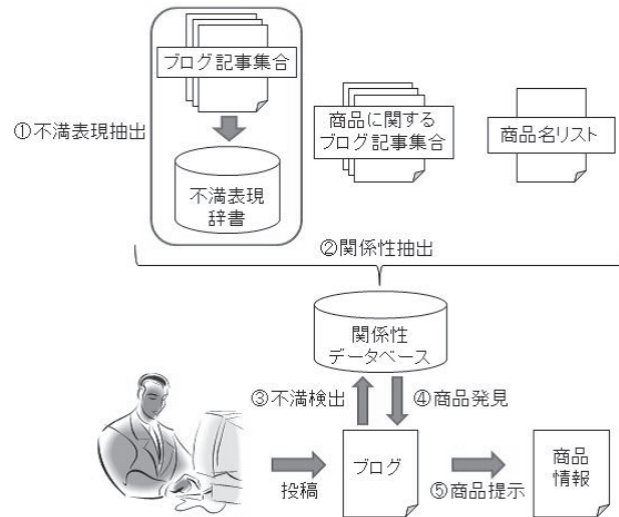


図2 システム概要

Fig. 2 System configuration of the proposed recommendation system.

(商品発見)。

手順5 ランキングの結果、順位の高い商品を提示する。

本論文で提案するシステムでは、ユーザへ商品情報を提示するまでの時間を短縮するため、「関係性の抽出」までのステップをあらかじめ「不満表現検出」より先に行い、不満表現とそれを解決する商品との関係をデータベースに保存しておく。一方、「不満表現検出」は、ブログ記事等が投稿されたタイミング等で随時実行し、利用者に提示する。また、利用者に提示する候補となる商品名は広告主等により既知のリストとして与えられていることを前提とする。それぞれの処理の詳細について以下に述べる。

3.2 不満表現辞書の作成

まず、本論文で抽出対象とする不満表現について述べる。ブログ記事集合に記述されている不満表現としては以下のようなものが存在するが、本論文では、(1)の名詞形の不満表現を抽出する対象とした。

- (1) 名詞(複合名詞)のみでネガティブな意味を持つ不満表現(ex. 花粉症, 停電, 汚れ, 焼きむら, 生活習慣病)
- (2) 形容詞のみでネガティブな意味を持つ不満表現(ex. 暑い, 寒い, 臭い, 汚い)
- (3) 名詞+形容詞のペアでネガティブな意味を持つ不満表現(ex. 起動が遅い, 容量が少ない, 火力が弱い)

ここで、名詞形の不満表現が出現する文の例として以下を考える。

例1. 商品Aは花粉症[不満表現]を解消[動作]してくれます[モダリティ]。

例2. 商品Bは花粉症[不満表現]を予防[動作]してくれます[モダリティ]。

本論文では例で示したような文から不満表現を抽出するために、ある特徴を持つ「動作」を利用する。不満表現は、この特徴を持つ「動作」の対象を抽出することで得られると考えられる。

以下では、「動作」の特徴について述べる。まず、上記の例で発生している出来事はポジティブな意味を持っており、かつ、「不満表現」と「動作」で表されている。「不満表現」がネガティブな意味を持つにもかかわらず、出来事がポジティブな意味となっているということは、「動作」が出来事をポジティブな意味に変化させていると考えられる。これは、ネガティブな名詞とともに出現し、その状態をポジティブな状態に変化させる「動作」が存在することを意味する。本論文では、このように、ネガティブな状態をポジティブな状態に変化させる特徴を持つ「動作」を改善動作と呼ぶこととする。不満表現は、まず改善動作を抽出し、その対象を抽出することで得られると思われる。

改善動作自体は、ポジティブな出来事が発生している文に出現し、かつ不満表現とともに出現しやすい動作を抽出することで得られると思われる。ポジティブな出来事が発生している文の判定には恩恵構文¹⁶⁾を利用する。恩恵構文は例1, 2のように「～てくれる」、「～てあげる」、「～てもらう」のようなモダリティを持ち、動作の受け手の利益を表す。したがって、文全体はポジティブな意味になると考えられる。改善動作の抽出としては、この恩恵構文のモダリティと不満表現との共起確率が高い動作名詞を抽出する。ただし、不満表現は少数のサンプルを事前に与えるものし、この少数のサンプルから改善動作を利用して最終的に多数の不満表現を得る形とする。たとえば、ある不満表現サンプルから改善動作 A を抽出した場合、改善動作 A が対象にとりうる不満表現を X 個と仮定すると、X-1 個が新たに獲得した不満表現となる。

共起確率を算出するために、不満表現サンプルは出現頻度が高いもの、つまり、多くの人がかかえているような不満表現が好ましい。web 上には不満のアンケート等の結果が載っているため、本論文ではその中から上位の不満表現のうち、名詞で表されているものを用いた。不満表現サンプルの抽出に利用したアンケートについては 4.1.1 項で述べる。

以上をふまえ、不満表現の抽出手順を以下に示す。

- (1) 不満表現抽出用のコーパスとして利用するブログ記事集合 B を事前に用意する (ブログ記事数を $|B|$ とする)。
- (2) 不満表現サンプル集合 D を事前に用意する。
- (3) ブログ記事集合 B を形態素解析し、動作名詞集合 V を抽出する。
- (4) ブログ記事集合 B から (2) で準備した不満表現サンプル集合 D と (3) で抽出した各動作名詞 $v (\in V)$ との共起確率 $P_d(v)$ を計算する。

$$P_d(v) = \frac{\sum_{d \in D} DF(d, v)}{|B|} \quad (1)$$

ただし、 $DF(d, v)$ は d, v が共起するブログ記事数を表す。

- (5) ブログ記事集合 B から恩恵構文のモダリティ M と (3) で抽出した各動作名詞 $v (\in V)$ との共起確率 $P_m(v)$ を計算する。

$$P_m(v) = \frac{\sum_{m \in M} DF(m, v)}{|B|} \quad (2)$$

ただし、 $DF(m, v)$ は m, v が共起するブログ記事数を表す。

- (6) (4), (5) で算出した値を用いて、式 (3) のように、(3) で抽出した各動作名詞 v が改善動作である可能性の高低を表す指標 $P_{kai}(v)$ を算出する。

$$P_{kai}(v) = P_d(v) \cdot P_m(v) \quad (3)$$

- (7) (6) の値が高い順に複数個の動作名詞を改善動作として抽出する (ただし、明らかに改善動作と異なるものは人手で排除する)。
- (8) ブログ記事から改善動作の直前に出現する名詞を出現頻度とともに抽出する。ただし必要があれば、ここで抽出した名詞のうち、特定出現頻度以下の名詞は不満表現である可能性が低いと排除する。
- (9) (8) で抽出した名詞から、ネガティブな意味を持つ名詞を不満表現として人手で選定する。

上記で示した不満表現や改善動作の抽出の手順には人手での選定が入っているが、不満表現辞書や改善動作自体は、1 度作成すれば時間の経過等によって変化しにくいという静的な側面を持っているため、自動化までは必須ではないと考えられる。そのため、本論文のように、不満表現や改善動作である可能性が高いものを提示することで、人手での抽出を効率化するだけでも十分有用である。

3.3 関係性の抽出

作成した不満表現辞書と、既知の商品リストとを用いて、ブログ記事コーパスから文章内の共起情報として、{商品名, 不満表現} のペアを抽出する。「関係性の抽出」では、この抽出したペアの中から、「商品が不満を解決する」という条件を満たすペアを抽出する。抽出した {商品名, 不満表現} のペアの関係性としては以下が考えられる。

- (1) 商品と不満表現に関係がなく、別の話題として記述される (ex. この前、風邪 [不満表現] を引いた。やっと治ったので商品 A で遊んだ)。
- (2) 商品の購入 (使用) 理由として不満表現が記述される (ex. 運動不足 [不満表現] 解消のために商品 A を買った)。
- (3) 商品購入後 (使用後) に解決した不満として不満表現が記述される (ex. 商品 A で運動不足 [不満表現] を解消した)。
- (4) 商品購入後 (使用後) に新たに発生したものととして不満表現が記述される (ex. 商品 A で筋肉痛 [不満表現] になった)。

(1)~(4) のうち、「商品が不満を解決する」という関係性を満たすペアは (2), (3) である。そこで、(1), (4) の関係性を排除し、(2) と (3) の関係性のみを抽出する方式について検討した。

方式としては、共起情報によって関係性の強さを判定する手法をとる。一般的に関連語の抽出において 2 つの語の共起頻度が高いほど、2 つの関係性は強いとされ、それが本論文に

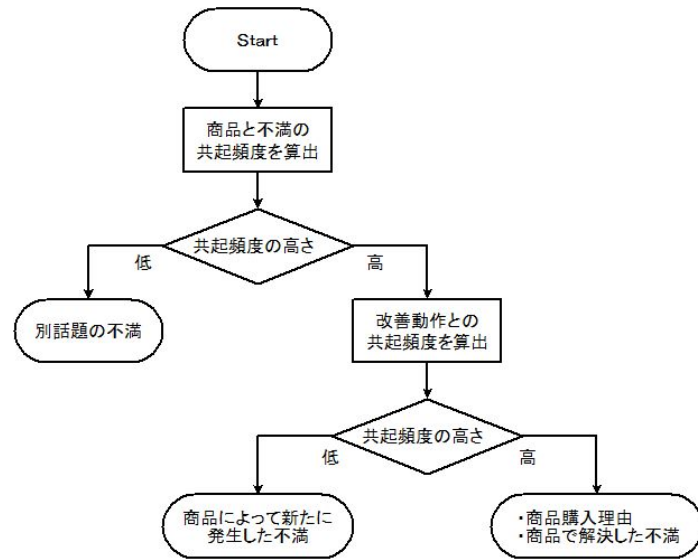


図3 関係性の抽出

Fig. 3 Extraction of the relations between dissatisfactions and items.

おける商品と不満表現の関係性にもあてはまると考えられる。3.2節で例示したような恩恵構文を抽出することでも上記の関係性は抽出可能であるが、「商品 X で不満 Y を予防した」のように恩恵構文として記述されていない場合は関係性を抽出できず、網羅性の低下が予想されることから、共起頻度による方式を選択した。

以下で方式の詳細を述べる。まず、(1)の関係性の排除（関係性が低いペアの排除）には、単純に商品名と不満表現の共起頻度を用いる。しかし、(4)の場合、すでに商品名と不満表現の間に強い関係性が存在してしまっているため、単純な商品名と不満表現の共起頻度では関係性を排除することはできない。ただし、(2)、(3)とは異なり、(4)は「不満をなくす、防ぐ」といったようなネガティブな状態をポジティブな状態に変化させる文脈を持たないため、(4)を排除するために、改善動作との共起頻度を用いることが可能である（図3）。

以上の理由から、関係性の抽出としては、{商品名, 不満表現, 改善動作}の3つの共起頻度をとることが有効であると考えた。

3.4 不満表現の検出

3.2節で抽出した不満表現辞書を用いて、商品推薦対象のユーザが投稿したブログ記事から不満表現を検出する。具体的には、ブログ記事の形態素解析を行い、各形態素と辞書に格納されている不満表現とのマッチングにより検出する。

3.5 商品の発見

次に、3.3節で抽出した関係性の強さを用いて、前節で検出した不満を解決する商品を見出す。

D_u をユーザ u の投稿記事集合から検出した不満表現の集合、 V を 3.3 節で抽出した改善動作集合、 i をユーザに提示する候補の商品、 d を不満表現としたとき、商品 i と不満表現 d との関係性の強さを表す関数 $r(i, d)$ は以下のように算出する。

$$r(i, d) = \sum_{v \in V} DF(i, d, v) \quad (4)$$

ここで $DF(i, d, v)$ はブログ記事コーパスの全体集合において、抽出された改善動作集合 V の要素と共起する商品 i と不満表現 d の共起頻度である。

そして、ユーザ u に提示する商品は、次式により、すべての i について $score_u(i)$ を算出して、この値が最も高い商品 $i (\in I)$ の情報をユーザ u に提供する。

$$score_u(i) = \sum_{d \in D_u} rank_{Dic}^{-1}(r(i, d)) \quad (5)$$

ただし、 $rank_{Dic}^{-1}(r(i, d))$ は、不満表現辞書 Dic 内のすべての不満表現 $d' (\in Dic)$ に対する商品 i との関係性 $r(i, d')$ を算出し、この値が大きい順にソートした際の不満表現 d の順位の逆数である。これにより、 $score_u(i)$ は、ユーザ u が言及した不満表現と商品 i との関係性の強さを評価する。

ここで、単純な共起頻度に基づき、商品 i を選択するのではなく、不満表現 d ごとの順位の逆数の和でスコアリングしたのは、 D_u に含まれる複数の不満表現を同じ重みで考慮するためである。たとえば、「花粉症」という不満表現と「鼻水」という不満表現が出現した場合に、これらの単語の出現頻度に偏りがある場合でも、いったん、順位の逆数でスコアリングすることで、それらに関連して抽出される商品の重みを同等に考慮することを可能にするためである。

4. 実験

前章で述べた方式を用い、2種類の実験を行った。1つ目は、単一の商品に対して不満表

現と商品の関係性を抽出する実験であり，抽出した関係性の精度と再現率を評価した．2 つ目は，複数の商品に対して不満表現との関係性を抽出し，あるブログの記事から抽出した不満表現に対して，妥当な商品が発見できることを確認する．

4.1 単一の商品に対する実験

4.1.1 実験方法

前章で述べた方式を用い，単一の商品に対して不満表現との関係性を抽出した．まず，改善動作を抽出するため，無作為に抽出したブログ記事 1,281,765 件を形態素解析し，動作名詞を抽出した．形態素解析には JTAG¹⁷⁾ を用いた．次に，式 (3) に従って {動作名詞，恩恵構文のモダリティ}，{動作名詞，不満表現サンプル} の共起確率から，抽出した動作名詞が改善動作である確率を算出した．恩恵構文のモダリティとしては，「～してくれる」を用い，不満表現サンプルとしては，goo ランキング¹⁸⁾ の「治したい！女性の身体の悩みランキング」，「30 歳を過ぎて気になり始めた体の変化ランキング」，「2007 年年間病名検索ワードランキング」から人手で抽出した 90 個のサンプルを用いた．また，共起範囲としては，モダリティが 1 形態素以内，不満表現サンプルが 2 形態素以内である．算出した改善動作である確率の高い順に，3 章で述べた改善動作の定義にあてはまる動作名詞を 10 個選定した．不満表現は，改善動作との距離が 10 形態素以内にある名詞を同じブログ記事集合から抽出し，その後，健康に関する不満表現を人手で 980 個選定した．選定した不満表現と改善動作を用いて，関係性の抽出を行った．商品は「WiiFit」を用い，ブログ記事は「WiiFit」，「Wii フィット」をクエリとして収集したブログ記事 46,349 件を用いた．また，共起範囲に関しては，不満表現と改善動作の共起範囲は 3 形態素以内で固定とし，不満表現と商品名の共起範囲は 50 byte から 1 記事以内まで変化させ，それぞれについて，精度，再現率を評価した．不満表現と商品名の共起範囲を変化させたのは，ブログに複数の話題が含まれていた場合に，商品と別話題の不満表現を排除するためである．評価に用いる正解データは以下のいずれかの項目を含む記事が 1 つでも存在した場合，その不満表現は正解であるとして作成した．

- (1) 商品によって不満がなくなった，あるいは防ぐことができたことが明記されている．
- (2) 商品によって不満をなくす，あるいは防ぐことを目的としていることが明記されている．

4.1.2 結果と考察

まず，今回抽出した改善動作と不満表現をそれぞれ表 1，表 2 に示す．表 2 には，「トラブル」，「症状」，「違和感」等の，単独では詳細を特定できない表現も含まれているが，これ

表 1 抽出した改善動作

Table 1 Extracted “kaizen-dousa” words that indicate problem prevention or solution.

改善動作	P_{kai}
解消	2.543870e-04
改善	0.7044562e-04
治療	0.6087893e-04
克服	0.2815650e-04
復活	0.1478488e-04
予防	0.1092559e-04
防止	0.02989590e-04
緩和	0.02446028e-04
回復	0.01576329e-04
除去	0.01386083e-04

表 2 抽出した不満表現の例

Table 2 Examples of the extracted dissatisfactions.

不満表現例			
ストレス	腰痛	ガン	トラブル
運動不足	虫歯	紫外線	ワキガ
違和感	夏バテ	疲れ	メタボ
アレルギー	老化	症状	寝不足
生活習慣病	便秘	風邪	角質

表 3 共起頻度の上位 10 件 (改善動作なし)

Table 3 Relations extracted by frequency of co-occurrence (without “kaizen-dousa”).

不満表現	共起頻度
筋肉痛	2,426
メタボ	546
風邪	392
ストレス	373
脂肪	367
運動不足	304
疲れ	298
内臓脂肪	210
メタボリック	208
リバウンド	185

表 4 共起頻度の上位 10 件 (改善動作あり)

Table 4 Relations extracted by frequency of co-occurrence (with “kaizen-dousa”).

不満表現	共起頻度
運動不足	208
ストレス	73
メタボ	40
便秘	24
太り	21
にきび跡	14
メタボリック	14
肥満	8
抜け毛	8
疲労	8

は「～のトラブル」，「～の症状」，「～の違和感」といった表現を抽出することを考慮してのことである．ただし，本論文における評価からはこれらの表現は排除した．実際に不満表現と商品の関係性の強さを抽出した結果を表 3，表 4，図 4 に示す．表 3，表 4 はそれぞれ {不満表現，商品名} の共起頻度上位 10 件と，{不満表現，商品名，改善動作} の共起頻度上位 10 件であり，図 4 は抽出した関係性を共起頻度で順位付けした際の，特定順位までの関係性の抽出精度を表す．ただし，不満表現と商品名の共起範囲は 1 記事内である．これらの結果から，改善動作を導入することによる以下のような効果がみられた．

- (1) 商品購入後に新たに発生した不満を表す不満表現の排除
- (2) 商品との間に複数の関係性が存在する不満表現に対しての正しい評価

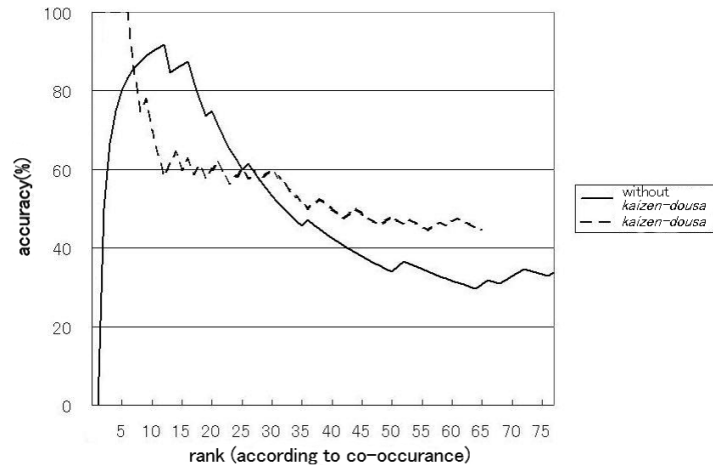


図 4 関係性抽出結果の精度 (共起範囲 1 記事内)

Fig. 4 Accuracy of extracted relation (the range of co-occurrence is within a blog). A solid line represents accuracy without “kaizen-dousa”, and a dotted line represents accuracy with “kaizen-dousa”.

(1) は 3.3 節で述べた効果である。表 3, 表 4 を比較すると, 商品購入後に新たに発生した不満表現である「筋肉痛」が改善動作を用いることにより排除され, これにより, 上位 6 件までは改善動作を共起条件に加えたほうが精度は高くなっている (図 4)。また, 同様の効果が 30 位以降においてもみられた。また, (2) について述べると, 本実験において, 商品と不満表現に関して以下のような複数の関係性に関する記述がみられた。

- 「不満」を解消しようとして「商品」を購入した。
- 「不満」によって商品が「使用」できなかった (商品と不満表現の間に関係性が存在しない)。

たとえば, 表 3 の「風邪」については, 確かに「体力をつけて風邪を防止するために WiiFit を購入した」という記述はあったが, 大多数は「風邪を引いたので, その日は WiiFit を使用しなかった」というものだった。このような例に対しては, 関係性は低いとして評価すべきである。これに関して, 改善動作を用いなかった場合には「風邪」は 3 位, 改善動作を用いた場合は 31 位となっており, 正しく関係性を評価できたことが分かる。しかし, 一方で 7 位から 29 位においては大きく精度を落としている。これは, 共起条件として改善動作を追加したことにより共起頻度が低下し, 商品と関係のない別話題の不満表現が出現してし

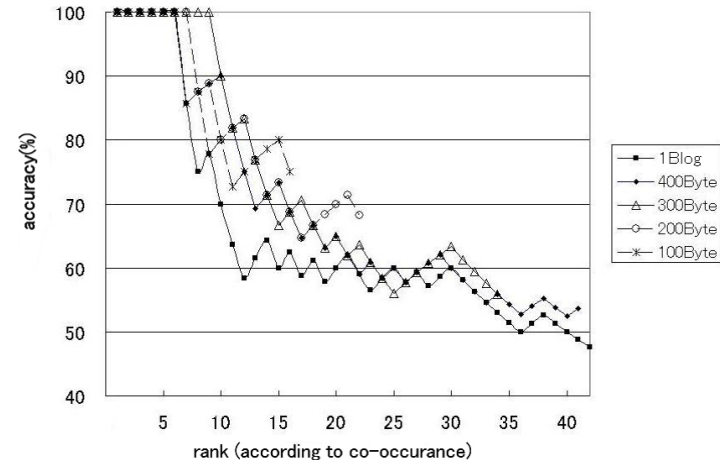


図 5 関係性抽出結果の精度 (改善動作あり, 共起範囲変化)

Fig. 5 Accuracy of extracted relations (with “kaizen-dousa”, the range of co-occurrence is changed).

まっている可能性がある。

確認のため, 不満表現と商品名の共起範囲を変化させた結果を見てみると (図 5), 共起範囲を狭めることにより別話題の不満が排除され, 精度が向上していることが分かる。ただし, その一方で図 6 を見てみると, 共起範囲を狭めることで再現率が低下してしまっている。

以上のことから, 精度と再現率をともに向上させるためには, 共起範囲を狭めずに別話題の不満表現を排除する, あるいは, 改善動作を増やすことにより, 共起頻度を増加させることが必要であると考えられる。前者としては, 改善動作を共起条件に加えない状態で足りるを行ってから, 改善動作を適用する等の方法が考えられる。後者としては, 単純に抽出する改善動作の数を増やす方法や, 「脂肪 (不満表現)」と「燃焼 (動作)」のように, ある不満表現に特有の改善動作が存在することから, これらの動作を特定の不満表現に対してのみ改善動作として定義する等の方法が考えられる。

また, 精度と再現率の向上とは別に, 推薦する商品を選択する際にも共起頻度が低い不満表現については考慮する必要があると思われる。共起頻度が低い不満表現が下位に出現する場合は, 他の適した商品が推薦されるため問題とはならないが, 上位に出現する場合はその商品が推薦されるため問題となる。そのため, 推薦する商品を選択する際に, 上位の不満表

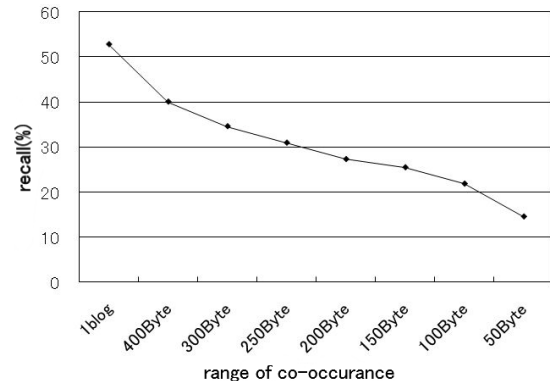


図 6 再現率の変化
Fig. 6 Recall of the extracted relations.

現の共起頻度が低い商品データに関しては無効とする必要があるが、どの程度の共起頻度とするべきかはこの実験だけでは判断できないため、今後の課題とする。

4.2 複数の商品に対する実験

4.2.1 実験方法

この実験の目的は、複数の商品が存在した場合に、ブログ記事から検出した不満表現に対して、不満を解決する商品が発見できることを確認することである。そのために、プロトタイプシステムを実装し、複数の商品に対して本論文で提案した方法を適用した。このシステムでは、{不満表現, 商品名, 改善表現}の共起頻度をあらかじめデータベースとして保持しており、不満表現が記述された文章が入力された場合に、その不満表現を自動で検出し、不満を解決する商品を出力する。出力する商品の選定としては、共起頻度が高い順に数個の商品を選定する。共起頻度を算出するために用いたブログ記事としては、「健康グッズ」や「健康商品」等のクエリで収集した 35,267 件のブログ記事を用いた。商品名に関しては、ブログ記事から「～を買いました」等の購買表現につながる名詞等を抜き出し、その中から健康に関して効果のある 289 個の商品を手で選定した。また、不満表現と商品名の共起範囲は 1 記事内とした。

4.2.2 結果と考察

実際に不満表現を含む文章入力した結果(図 7)と、検出した不満表現に対しての商品の出力を示す(表 5)。図 7 では、ユーザが入力した文章が上部に表示されており、その下に、

TITLE: 晴れのち曇り、ときどき夏バテ

これから、暑くなると食欲が上がるどころか下がってきますね。
 やだな～、夏バテの季節・・・。
 今日は夕々のピーカン天気です。
 夜からはまた、天気が下っていく・・・という、予報の関東地方。
 暑い中、図書館に借りていた本を返しに行ってきた、借りてきました。
 本が重し、帰りに買い物するので、カートを引きに行きます。
 ウチではなぜか、カートを昔から「コロコロ」と呼んでます。
 さすがに、平日の午前中は歩いていました。
 なので、今日は広い範囲で読みたい本などを検索。
 健康おかし365日一決定版(SSCムックレータスクラブX2005/03)不明商品詳細を見る普段は料理のコーナーは素通りするのですが、このことと「食」について書いていたので、ちょっとのぞいてみました。
 これから、夏バテの季節。
 図書館に歩いて来ただけで、軽く夏バテ気味なので、何か、効率的に栄養を摂取できる食べ合わせはないか・・・と
 思っていたのです。

あなたの不満:
夏バテ

- 商品:
1. つな井 point 1
 2. オクラ point 1
 3. 白菜キムチ point 1
 4. 生ウナギ point 1
 5. アイスクリューム point 1
 6. きゅうり point 1
 7. 原酒 point 1
 8. アイスノ point 1
 9. サツマイモ point 1
 10. 鰻 point 1

[戻る](#)

図 7 文章入力結果
Fig. 7 Output of the system.

表 5 各不満に対する上位 5 件
Table 5 Output of the system (top 5).

rank	運動不足	アトピー	疲れ目
1	ロデオボーイ	ネイチャーメイド	ブラックベリー
2	ヨガマット	オリーブオイル	黒豆
3	ジョーバ	石けん	スイカ
4	ピラティス	レメディ	ブルーベリー
5	サングラス	オーガニックコットン	いも
rank	メタボ	抜け毛	頭痛
1	ライ麦	シャンプ	バファリン
2	スポーツ自転車	マッサージオイル	頭痛薬
3	草履	ケラスターゼ	風邪薬
4	やせ薬	育毛剤	ぶり
5	黒烏龍茶	アロエ	いも

文章から検出した不満表現と、不満を解決する商品名が表示されている。表 5 は、このようにして出力された不満表現と、その不満を解決する商品の上位 5 件を提示したものである。表 5 を見ると、「運動不足(不満表現)」に対して「ロデオボーイ(商品)」や「ジョーバ(商品)」等、「疲れ目(不満表現)」に対して「ブラックベリー(商品)」や「ブルーベリー(商品)」等、「頭痛(不満表現)」に対して「パファリン」等が提示されている。これらの商品について web で検索すると、以下のような説明があった。

ロデオボーイ：ウォーキングと同等程度の運動量で、室内で手軽に有酸素運動ができます。たづな付なので、体力に自信がない方でも安心。手軽に毎日続けられます。運動不足や姿勢の悪さを気にされている方に。

ジョーバ：身体への負担が少なく、運動が苦手な方や体力のない方でも、ダイエットや筋力アップ、運動不足の解消等の効果を得られるのが、ジョーバフィットネスなのです。

ブラックベリー：目によいアントシアニンが豊富眼精疲労による肉体的・精神的疲労に効果
ブルーベリー：アントシアニンとは、植物に含まれる紫色の色素で、ポリフェノールの一種。主にワインの原料であるブドウや、ブルーベリー、紫芋、あずき等に含まれています。このアントシアニンには眼の疲れを癒したり、健康を維持する働きが大きいとされています。

パファリン：日常的な頭痛、生理痛、急な発熱まで、パファリンは、様々な症状に合わせてお使いいただけます。

このことから複数の商品から不満を解決する商品が選択され、提示されることが見て取れた。

5. ま と め

本論文ではユーザが潜在的にかかえる不満に対し、解決方法としての商品を提示するため、「不満表現の抽出」と「関係性の抽出」について検討を行った。不満表現の抽出では、不満表現、動作、恩恵構文のモダリティの 3 要素で構成されるタイプの文に着目し、ネガティブな状態をポジティブな状態に変化させる改善動作を利用することで、ヒューリスティックに抽出する方式を示した。さらに、関係性の抽出においては、商品と不満の間に成立する関係性を示し、改善動作を利用することで、目的とする関係性が抽出可能であることを示した。また、改善動作を利用した方式を用いてプロトタイプシステムを実装し、健康分野において抽出した不満表現に対して、複数の商品から不満を解決する商品を発見できることを明らかにした。

本論文の主な貢献は、不満表現からの商品の推薦という書き手をターゲットとする新しい広告、推薦の枠組みを提案し、その可能性を明らかにしたことである。しかし、本論文で提案する枠組みにおける各ステップで提案した個々の要素技術についてはまだいくつかの課題が残されている。特に以下の課題があげられる。

- 不満表現抽出の自動化
- 関係性抽出における再現率の改善
- 形容詞タイプ、名詞 + 形容詞タイプの不満表現への拡張
- 商品の発見にステップにおける商品のランキング方法の改良

参 考 文 献

- 1) Adwords, available from <http://www.google.co.jp/adwords/start/start.html>.
- 2) Adsense, available from <https://www.google.com/adsense>.
- 3) Amazon, available from <http://www.amazon.co.jp/>.
- 4) 土方嘉徳：嗜好抽出と情報推薦技術，情報処理，Vol.48, No.9, pp.957-965 (2007).
- 5) Jin, X., Zhou, Y. and Mobasher, B.: A Maximum Entropy Web Recommendation System: Combining Collaborative and Content Features, *Proc. ACM SIGKDD Conf.*, pp.612-617 (2005).
- 6) Popescul, A., Ungar, L.H., Pennock, D.M. and Lawrence, S.: Probabilistic Models for Unified Collaborative and Content-Based Recommendation in Sparse-Data Environments, *UAI-2001* (2001).
- 7) Berger, A.L., Della Pietra, S.D. and Della Pietra, V.J.D.: A maximum entropy approach to natural language processing, *Computational Linguistics*, Vol.22, No.1, pp.39-71 (1996).
- 8) Dempster, A., Laird, N. and Rubin, D.: Maximum likelihood from incomplete data via the EM algorithm, *J. Roy. Statist. Soc. B*, Vol.39, pp.1-38 (1977).
- 9) 土方嘉徳：情報推薦・情報フィルタリングのためのユーザプロファイリング技術，人工知能学会論文誌，Vol.19, No.3a (2004).
- 10) Hatzivassiloglou, V. and McKeown, K.R.: Predicting the semantic orientation of adjectives, *ACL*, pp.174-181 (1997).
- 11) 那須川哲哉，金山 博：文脈一貫性を利用した極性付評価表現の語彙獲得，自然言語処理研究会報告，pp.109-116 (2004).
- 12) 乾 孝司，乾健太郎，松本裕治：出来事の望ましさを判定を目的とした語彙知識獲得，第 10 回言語処理学会年次大会 (2004).
- 13) 東山昌彦，乾健太郎，松本裕治：述語の選択選好性に着目した名詞評価極性の獲得，言語処理学会第 14 回年次大会論文集，pp.584-587 (2008).
- 14) De Saeger Stijn，鳥澤健太郎：トラブルを見つける，言語処理学会，第 14 年次大会

論文集, pp.1073–1076 (2008).

- 15) 乾健太郎, 原 一夫: 経験マイニング: Web テキストからの個人の経験の抽出と分類, NLP2008 (2008).
- 16) 益岡隆志: 日本語モダリティ探求, くろしお出版 (2007).
- 17) Fuchi, T. and Takagi, S.: Japanese morphological analyzer using word co-occurrence – JTAG, *Proc. COLING-ACL '98: 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, Montreal, Vol.1, pp.409–413 (1998).
- 18) goo ランキング, available from (<http://ranking.goo.ne.jp/>).

(平成 23 年 3 月 20 日受付)

(平成 23 年 9 月 12 日採録)

推 薦 文

本論文は, ユーザが記述したテキストから, ユーザの不満を解決する商品を発見することを目的としている. ユーザの実体験が記述されたブログのテキストを大量に収集し, その中から不満表現の抽出と, 商品と不満との関係を抽出する実験を行い, その結果, 「解消」「防止」等の不満をなくす, あるいは防ぐといった意味を持つ動作である「改善動作」との

共起情報を利用することで, 精度良く抽出できるという知見を得た. この知見から, ユーザのいくつかの不満に対して妥当な商品を発見することが可能となり, 有益な研究であると考えられる. (グループウェアとネットワークサービス研究会主査 小林 稔)



坂井 俊之 (正会員)

日本電信電話株式会社 NTT サイバーソリューション研究所研究員. 2007 年東北大学大学院工学研究科修士課程修了. 同年日本電信電話(株)入社. 現在, ソーシャルメディアからのデータマイニングの研究に従事.



藤村 考 (正会員)

日本電信電話株式会社 NTT サイバーソリューション研究所主幹研究員. 1989 年北海道大学大学院工学研究科情報工学専攻博士課程修了. 同年日本電信電話(株)入社. 現在, ソーシャルメディアからのデータマイニング, 大規模データの可視化, 検索インタフェースの研究に従事. 工学博士. 電子情報通信学会, 日本データベース学会各会員.