

# オンラインショッピングサイトにおけるレビューを用いた 商品対の目的判定による商品推薦

## A Recommendation Method for Online Shopping using Purpose Detection on Pair Items

本田 達也† 北山 大輔† 角谷 和俊†

Tatsuya Honda Daisuke Kitayama Kazutoshi Sumiya

### 1. はじめに

近年、楽天市場[1] や Amazon[2] などのオンラインショッピングサイトが普及している。オンラインショッピングサイトでは閲覧商品に対して「この商品を買った人はこんな商品も買っています」というように、協調フィルタリング[3, 4] を用いた商品推薦が行われる。その際、閲覧商品と推薦商品を組み合わせる場合、各商品の組み合わせの目的が異なる場合が存在する。例えば、「インスタントカップ麺」を閲覧しているユーザに対して、「カレー味のインスタントカップ麺」や「パック米飯」、 「単三電池」といった商品が推薦される。ここで、「カレー味のインスタントカップ麺」や「パック米飯」をそれぞれ「インスタントカップ麺」との組で見た場合、その目的は「インスタント食品」と考えられる。しかし、「単三電池」と「インスタントカップ麺」の組み合わせの場合には「非常時用品」という目的と考えられる。このように、ある商品を他の商品と組み合わせる場合にその商品の目的が異なる場合が存在する。

また、我々は商品の組み合わせによる目的の差異に着目し、既存の商品推薦に代わる、組み合わせる時の目的に沿った商品の推薦が有用であると考えた。例えば、「インスタントカップ麺」を閲覧しており、推薦商品から「単三電池」を選択したユーザには、推薦されている「単三電池」よりも、「非常時用品」としてより適切な「太陽電池による充電機能付き単三電池」の方が望ましい推薦となる。本研究では、商品レビューに商品の目的が含まれると考え、商品レビューの類似度により商品対の目的を判定し、その目的に合致する代替商品を推薦する手法の提案を行う。

本稿の構成は以下の通りである。まず 2 章で本研究のアプローチおよび関連研究を説明し、3 章において、カテゴリ、商品およびレビューの特徴ベクトルの生成方法について述べる。4 章では、それぞれの特徴ベクトルを用いた目的の判定手法を議論し、5 章で推薦アルゴリズムを説明する。6 章でまとめと今後の課題を述べる。

### 2. 研究の概要

#### 2.1 アプローチ

本手法では、商品対の目的判定のためにオンラインショッピングサイトの商品カテゴリと商品、およびその商品レビューを用いる。ある商品カテゴリには特定の用途

の商品が分類されていると考えられるため、本研究では一つのカテゴリは一つの目的を表していると仮定する。商品レビューには、「非常時」や「備蓄用」という購入したユーザの使用目的を表すキーワードが出現していることがある。そこで、商品対の目的を判定する手法として、任意の商品対に付与されたレビューとカテゴリの類似度を算出することで商品対の目的を判定することが可能であると考えた。さらに、得られた目的を用いて代替商品推薦を行う(図 1)。以下に処理手順を説明する。なお、事前に商品カテゴリ、商品、レビューの特徴ベクトルは生成しておくものとする。本研究では  $\chi$  二乗値により特徴ベクトルを生成する。

- (1) 商品対となる商品  $i_x, i_y$  を入力し、それぞれのレビューを取得する。
- (2) 各レビューと各カテゴリの類似度を判定し、上位  $L$  件のカテゴリを目的カテゴリ  $P$  として抽出する。
- (3) 商品  $i_y$  と全商品との類似度を判定し、その上位  $M$  件の商品と目的カテゴリ  $P$  との類似度を算出、上位  $N$  件を推薦商品として提示する。

この処理により、選択した商品  $i_y$  に類似し、かつ商品対の目的  $P$  に合致する商品を推薦することが可能となる。

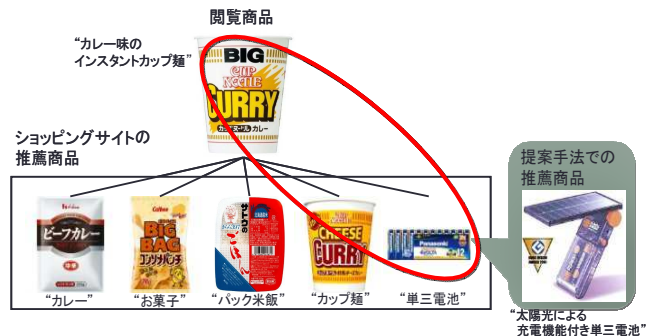


図 1 商品対の目的判定による代替商品推薦の概念

入力する商品対  $i_x, i_y$  として「インスタントカップ麺」と「単三電池」を選択した場合を例に処理の流れを説明する。まず、上記の手順に従い両商品のレビューを取得する。このとき、「インスタントカップ麺」および「単三電池」には表 1 のレビューが記述されているものとする。これらのレビューと全商品カテゴリについて後述する手法に基づき類似度を計算する。「非常時用品」のカテゴリの特徴語が「災害」、「停電」、「備蓄」、「買いだめ」である場合、商品レビュー  $r_{i_x}$  と  $r_{i_y}$  に類似するため、「非常時用品」が商品対  $i_x, i_y$  の目的を表すと判定する。

† 兵庫県立大学, University of Hyogo

インスタントカップ麺	単三電池
$r_1^x$ : とても美味しい	$r_1^y$ : 停電時のために備蓄
$r_2^x$ : お買い得でした	$r_2^y$ : アウトドアでよく使う
$r_3^x$ : 災害時用に買いため	$r_3^y$ : セットでお買い得

表1 商品レビューの例

判定した目的に基づく商品推薦では、まずユーザが選択した推薦商品“単三電池”と類似する“単一電池”，“マンガン電池”，“充電可能電池”などの商品を抽出する。そしてそれらの商品と「非常時用品」カテゴリとの類似度を求めることで“太陽電池による充電機能付き単三電池”のような商品を推薦する。

## 2.2 関連研究

本研究で用いているレビュー文に関しては、評判情報抽出[5,6]が盛んに行われているが、レビューの有用性、特に目的や理由など特定の情報を判断、抽出する技術についても取り組まれている。山澤ら[7]はAmazonレビュー文を収集し、数千の規模のレビューに対して、人手により「商品購入の判断に有用なレビューである」というラベルを付与し、SVMにより有用なレビューを自動判別し、82%の精度で分類した。この研究では、実験データの3~4割のレビューに有用ラベルが人手によって付与された。有用でないレビューには、本研究で仮定している目的情報は含まれていないと考えられ、これらのレビューを効率的に除去することで、本手法の精度も向上するものと考えられる。中山ら[8]はレビュー中に含まれる理由に関する記述とそれが商品選択に与える影響を調査している。この調査によると、映画などのレビューに書かれている感想になぜその感想に至ったのかという理由が述べられている方が参考情報になりやすいという結果が得られている。本研究で扱う商品の使用目的は、言い換えると購入理由である。価値のあるレビュー文には使用目的に関する文言が含まれやすいという可能性がある。

川中ら[9]はレビュー記事を用いた消費行動要因の抽出手法を提案している。すなわち、ある消費行動を起こした場合の理由に該当するキーワードをマイニングによって抽出している。平山ら[10]はレビュー中の手がかり表現に着目し、要因抽出とその可視化を行っている。このような技術により、本研究においても商品の利用目的を示すレビューおよびキーワードの抽出が可能になるものと考えられる。

推薦技術はコンテンツベースのフィルタリング[11, 12]と協調フィルタリング[3, 4]に大別できる。本研究はこのうち、コンテンツベースのフィルタリング技術に相当する。コンテンツベースのフィルタリングとして、加藤ら[13]は、遺伝的アルゴリズムによってユーザプロフィールを作成し、意外性と正確性の尺度によって両者を満たす推薦結果を生成する手法を提案した。これに対し本研究では、ユーザが選択した商品対からユーザプロフィールに相当する目的を抽出し、協調フィルタリングの結果では得られないが、ユーザの要求に合致する商品の推薦を行う。そのため、ユーザが行うことは閲覧商品に加えて、

推薦商品を選択することのみであり、入力負担を軽減することができる。

## 3. カテゴリ、商品、レビューの特徴ベクトル

### 3.1 商品カテゴリ、商品、商品レビューの関係

商品カテゴリ集合  $C$  と商品集合  $I$ 、商品レビュー集合  $R$  の関係について説明する。ある商品カテゴリは商品集合を含んでおり、各商品は商品レビュー集合を持つ。また、カテゴリはツリー構造であり、ルート以外のカテゴリは一つの親カテゴリをもつ。商品は末端カテゴリに所属し、その祖先カテゴリすべてに所属するものとする。すなわち、ルートカテゴリにはすべての商品  $I$  が所属する。以下の式によりこれらの関係を定義する。図2はこれらを模式的に表したものである。

$$C = \{c_1, c_2, \dots, c_n\} \quad (1)$$

$$c_a = \{i_1^a, i_2^a, \dots, i_m^a\} \quad (2)$$

$$i_b^a = \{r_1^b, r_2^b, \dots, r_l^b\} \quad (3)$$

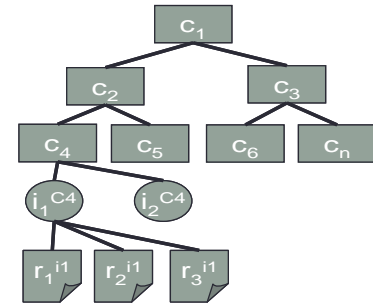


図2 想定するカテゴリ構造、商品、レビューの関係

### 3.2 カテゴリ特徴ベクトル

本研究で想定するデータの構造上、親カテゴリが持つ商品データの部分集合が子カテゴリの商品データとなる。そのため、カテゴリの特徴を表現するために、親カテゴリが持つ商品データのレビューの単語頻度を期待値として、子カテゴリが持つ商品データのレビューの単語頻度の  $\chi$  二乗値を計算する。これにより各単語の特徴量を算出し、特徴ベクトルとする。図3に算出時の観測値と期待値の関係を図示した。カテゴリ  $c_a$  の特徴ベクトルの生成は下記の手順で行う。

- (1) カテゴリ  $c_a$  に含まれる商品のレビューより  $c_a$  の各単語頻度を算出する。同様に親カテゴリ  $c_b$  の各単語頻度を算出する。
- (2)  $c_a$  の親カテゴリ  $c_b$  との組を用いて  $c_a$  の各単語の  $\chi$  二乗値を算出する。
- (3) 親カテゴリ  $c_b$  についても同様に各単語の  $\chi$  二乗値を算出し、 $c_a$  と  $c_b$  の平均をカテゴリ  $c_a$  の特徴ベクトルとする。

$$\chi_{k,c_a} = \frac{(O_{k,c_a} - E_{k,c_a})^2}{E_{k,c_a}} \quad (4)$$

$$E_{k,c_a} = |W_{c_a}| \times \frac{O_{k,c_b}}{|W_{c_b}|} \quad (5)$$

$$\chi'_{k,c_a} = \frac{\chi_{k,c_a} + \chi_{k,c_b}}{2} \quad (6)$$

式中の  $O_{k,c_a}$  はカテゴリ  $c_a$  におけるレビュー中のキーワード  $k$  の出現数であり、 $E_{k,c_a}$  はカテゴリ  $c_a$  におけるレビュー中のキーワード  $k$  が出現する期待値である。期待値は、カテゴリ  $c_a$  におけるレビュー中の全キーワード  $W_{c_a}$  の出現数に、親カテゴリ  $c_b$  におけるキーワード  $k$  の出現割合を乗算して求める。

ここで、単純に  $\chi$  二乗値を算出するだけでは、親カテゴリにおける子カテゴリの特徴が抽出されるに過ぎない。そこで、カテゴリを階層的に見た場合、各カテゴリは親カテゴリの特徴を含んでいるという考えに基づき親カテゴリの  $\chi$  二乗値と平均する処理を行う（手順(3)）。

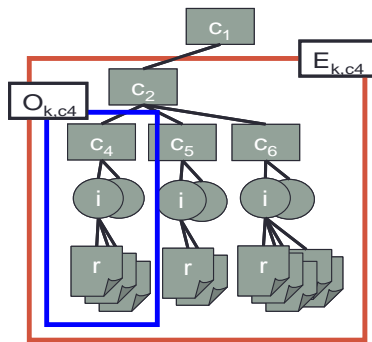


図3 特徴量算出時の観測値と期待値の関係

### 3.3 商品特徴ベクトル

商品の特徴ベクトルも、カテゴリの特徴ベクトルと同様の考え方で生成する。すなわち、ある商品  $i^b_a$  が所属するカテゴリ  $c_b$  に所属するレビューの単語頻度を期待値とした、商品  $i^b_a$  に所属するレビューの単語頻度の  $\chi$  二乗値を算出する。式(4)で算出する期待値  $E_{k,c_a}$  中ではカテゴリ  $c_b$  の単語頻度、 $O_{k,c_a}$  にアイテム  $i^b_a$  の単語頻度を代入して算出する。

### 3.4 レビュー特徴ベクトル

レビューの特徴ベクトルは、ある目的を表現するカテゴリ集合におけるそのレビューが表す目的の特徴を表現する。そのために、ある商品  $i^b_a$  が所属するカテゴリ  $c_b$  に所属するレビューの単語頻度を期待値としたレビュー  $i^b_a$  の単語頻度の  $\chi$  二乗値を算出する。すなわち、カテゴリ  $c_b$  内におけるレビュー  $i^b_a$  の特徴量を表現することで、レビュー  $i^b_a$  が所属カテゴリと異なる傾向を持つという特徴を抽出可能である。式(4)で算出する期待値  $E_{k,c_a}$  中ではカテゴリ  $c_b$  の単語頻度、 $O_{k,c_a}$  にアイテム  $i^b_a$  の単語頻度を代入して算出する。

## 4. 商品対の目的判定

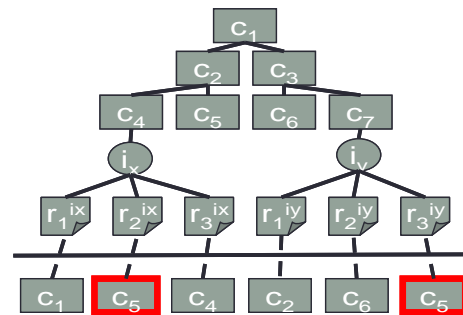
### 4.1 共通するカテゴリによる判定

我々は商品対の目的を判定する際に、3つの異なる手法を提案する。まず1つ目は、個々のレビューから目的抽出を行い、その目的が共通する場合に、それを商品対の目的とする手法である。これは、各商品にはそもそも複数の目的があり、組合せに意味があるならば、共通する目的を持つという考えである。具体的には下記の手順で行う。

- (1) 商品  $i_x$  の持つ各レビューの特徴ベクトルと各カテゴリの特徴ベクトルとの類似度を算出し、それぞれのレビューに対して、類似度が最大となるカテゴリをそのレビューの目的カテゴリとして対応付ける。この目的カテゴリ集合を  $P_x$  とする。商品  $i_y$  についても同様に対応付け、 $P_y$  とする。
- (2)  $P_x \cap P_y$  を商品対の目的  $P$  とする。

本手法を含め、以降の手法で用いている類似度はいずれもコサイン尺度である。

インスタントカップ麺の例の場合、「まとめ買い」カテゴリ（図4中の  $c_5$ ）において特徴量が高いキーワードが“大量”、“お買い得”であったとするならば、カップ麺のレビュー  $i^b_2$  と単三電池のレビュー  $i^b_3$  の類似度がそれぞれ高くなる。両商品ともに同じカテゴリが抽出されるため、この場合では目的カテゴリが「まとめ買い」と判定される。



判定されたカテゴリ

図4 共通するカテゴリによる判定例

この手法を用いた場合、明示的に目的  $P$  であるとわかるレビューが書かれている場合には、適切に判定することができ、偶発的に類似度が高くなってしまふ場合を減らすことが可能であると考えられる。しかし、商品  $i_x$  のレビュー  $i^b_a$  の目的カテゴリの候補が複数ある場合、すなわち類似度に有意な差が見られない場合には、目的  $P$  は  $i_x$  と  $i_y$  の共通の目的として判定されない可能性が高い。すなわち、適合率が高く、再現率が低い手法になると考えられる。

### 4.2 関係するカテゴリによる判定

2つ目の手法は、商品対の目的は両者の商品が既に割り当てられているカテゴリのうちのどちらか、またはその祖先カテゴリであるという考え方である。言い換えると、

ユーザの商品選択の目的の比重が閲覧商品か推薦商品のどちらかにあるという考えで、その中から目的を判定する。具体的には下記の手順で行う。

- (1) 商品  $i_x$  の各レビューの特徴ベクトルと商品  $i_y$  の所属するカテゴリ  $c_a$  からルートまでの祖先カテゴリそれぞれの特徴ベクトルとの類似度を計算する。類似度が最も高くなるカテゴリを目的  $P_x$  とする。
- (2)  $i_y$  についても同様に行い、 $P_y$  を得る。
- (3)  $P_x$  と  $P_y$  のうち、類似度が高いものを目的カテゴリ  $P$  とする。

インスタントカップ麺の例の場合、カップ麺の所属カテゴリが「インスタント食品」であり、その親カテゴリが「野外活動」（図 5 中の  $c_2$ ）であったとする。野外活動カテゴリにおいて特徴量が高いキーワードが“アウトドア”であったとすると、電池のレビュー  $r_2^y$  との類似度が高くなる。そのため、この場合では目的カテゴリは「野外活動」と判定される。

この手法を用いた場合、2 つの商品が構造的に比較的近いカテゴリに属する商品であった場合に、目的がどちらに偏っているのかを判定することが可能であると考えられる。しかしながら、祖先カテゴリに商品対の目的が存在することが少ないと考えられるため、適合率が高く、再現率が低い手法になると考えられる。

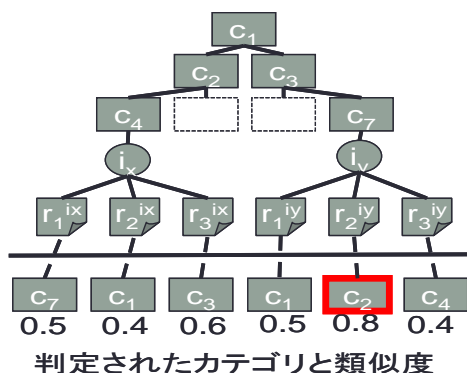


図 5 関係するカテゴリによる判定例

#### 4.3 レビューの合成による判定

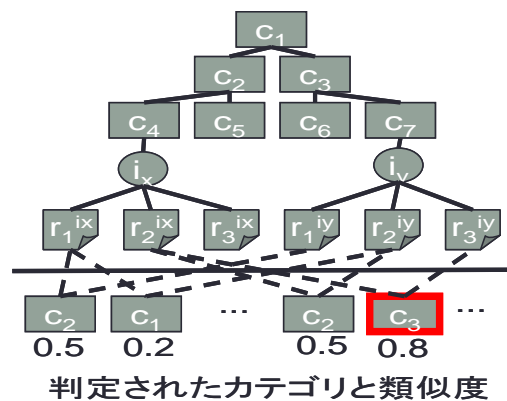
3 つ目は、商品を組み合わせたことによる意味を扱うために、各商品から得られるレビューを組み合わせて合成ベクトルを生成し、その合成ベクトルによって目的を抽出する手法である。1 つ目の共通するカテゴリによる手法と考え方は類似しているが、合成ベクトルを生成することで、個々の商品のみでの判定では得られない目的を抽出可能となる。具体的には下記の手順で判定を行う。

- (1) 商品  $i_x$ ,  $i_y$  より 1 つずつレビューを取り出し、合成ベクトルを生成する。
- (2) 合成ベクトルと各カテゴリ  $C$  との類似度を計算する。同様にすべての組み合わせで合成ベクトルを生成し、類似度を計算する。
- (3) 類似度の上位  $N$  件のカテゴリを目的  $P$  とする。

インスタントカップ麺の例の場合、「非常時用品」カテゴリ（図 6 中の  $c_3$ ）における特徴量が高いキーワードが

“災害”，“停電”，“備蓄”，“買いだめ”である場合、商品レビュー  $r_3^x$  と  $r_1^y$  を合成して作成する特徴ベクトルとの類似が高くなる。そのため、この場合では、非常時用品が商品対  $i_x$ ,  $i_y$  の目的カテゴリと判断される。

この手法を用いた場合、商品  $i_x$  のみでも得られない、また商品  $i_y$  のみでも得られない目的  $P$  を得られる可能性がある。しかしながら、無関係のカテゴリとの類似度が高くなることも予想され、適合率が低く、再現率が高い結果になると考えられる。



判定されたカテゴリと類似度

図 6 レビューの合成による判定例

#### 5. 目的判定を用いた推薦

我々はオンラインショッピングサイトにて閲覧商品に対して推薦される商品があった場合、その推薦商品よりもユーザの選択目的に合致した商品が存在すると考えた。例えば、ユーザが“インスタントカップ麺”を閲覧しており、推薦商品リスト中の“単三電池”を選んだ場合、この商品対の目的は防災用品であると考えられるため、より防災の目的に合致した“太陽電池による充電機能付き単三電池”を推薦する。これを実現するために、ユーザが選択した推薦商品に類似し、かつユーザが選択した商品対の目的に合致する商品を抽出する必要がある。そこで、以下の手順で推薦商品を決する。

- (1) 商品集合  $I$  より、ユーザの選択商品  $i_y$  の特徴ベクトルとの類似度が閾値以上の商品  $I'$  を抽出する。
- (2) 商品集合  $I'$  を商品対の目的  $P$  カテゴリの特徴ベクトルと類似する順に並べ、上位  $N$  件の商品を推薦商品とする。

ここで類似度はコサイン尺度を用いている。

インスタントカップ麺の例を図 7 に示す。(1) の処理によって、“アルカリの単三電池”，“太陽電池による充電機能付き単三電池”，“単一電池”というような商品が抽出される。次に(2) の処理により，“太陽電池による充電機能付き単三電池”，“アルカリの単三電池”，“単一電池”という順序に判定され、上位の“太陽電池による充電機能付き単三電池”が“単三電池”の代替商品として推薦される。



図7 代替商品の推薦例

## 6. まとめ

本稿では、ユーザが選択した商品対の目的の抽出手法について議論し、抽出した目的を用いた商品推薦アルゴリズムを説明した。今後、実データを用いて実装し、手法の検討を進める必要がある。具体的な検討事項は以下の通りである。

まず、商品の特徴ベクトルを生成する際に、親カテゴリの特徴量が大きい場合に、子カテゴリの特徴量がほとんど無視されてしまうことをすでに確認している。そのため親カテゴリの特徴をどの程度継承すべきなのかという妥当な重み付けを特定する必要がある。

次に、商品対の目的抽出に関しては、3つの手法の性質を実験により明らかにする必要がある。またレビューの文書の長さはまばらであるため、適切に類似度が計算されないことが予想される。そのため、個々のレビューを用いた場合、クラスタリングしていくつかのレビュー集合とした場合、すべてのレビューをひとまとめにして用いた場合を検討し、商品対の目的判定に用いるレビューの粒度を明らかにする必要がある。

最後に、商品推薦に関しては、協調フィルタリングで推薦された結果と提案手法による代替商品のどちらがユーザの満足度が高いのかを明らかにする必要がある。また、商品対の目的を利用したクラスタリングなど他のアプリケーションへの応用を行う予定である。

## 謝辞

本研究の一部は、平成24年度科研費若手研究(B)「ユーザ閲覧行動と周辺コンテンツの質的評価に基づく協調型画像撮影アシスト方式」(課題番号: 24700098)によるものです。ここに記して謝意を表すものとします。

## 参考文献

- [1] 楽天市場 : <http://www.rakuten.co.jp/>
- [2] Amazon : <http://www.amazon.co.jp/>
- [3] P.Resnick, N.Iacovou, M.Suchak, P.Bergstrom, and J.Riedl. "GroupLens: Open Architecture for Collaborative Filtering of Netnews" In Proc. of Conference on Computer Supported Cooperative Work, pp. 175-186, 1994
- [4] Badrul Sarwar, George Karypis, Joseph Konstan, John Riedl, Item-based collaborative filtering recommendation algorithms, In Proc. of the 10th international conference on World Wide Web, pp. 285-295, 2001
- [5] 宮崎林太郎, 森辰則, 製品レビュー文に基づく評判情報コーパスの作成とその特徴の分析, 情報処理学会研究報告 自然言語処理, Vol.2008 No.90, pp. 99-106, 2008年
- [6] 井上雄介, 太田学, ユーザの評価極性付きレビュー分析に基づく未評価レビューの極性分類, 情報処理学会研究報告 データベースシステム (DBS), Vol.2011-DBS-153 No.19, pp. 1-6, 2011
- [7] 山澤美由起, 吉村宏樹, 増市博, Amazon レビュー文の有用性判別実験, 情報処理学会研究報告 自然言語処理, Vol.2006 No.53, pp. 15-20, 2006
- [8] 中山記男, 神門典子, レビューにおける「理由」の分析～被験者実験より～, 情報処理学会研究報告 自然言語処理, Vol.2006 No.1, pp. 81-88, 2006
- [9] 川中翔, 宮田章裕, 東中竜一郎, 星出高秀, 藤村考, レビュー記事群を用いた消費行動要因のマイニング, 情報処理学会研究報告 データベースシステム (DBS), Vol.2010-DBS-151 No.5, pp.1-8, 2010
- [10] 平山 拓央, 湯本 高行, 新居 学, 高橋 豊: "商品レビューの傾向と要因のグラフ化", DEIM Forum, C3-5, 2012
- [11] 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).
- [12] 徳永健伸, 辻井潤一: 情報検索と言語処理: 言語と計算, 東京大学出版会 (1999).
- [13] 加藤 由花, 川口 賢二, 箱崎勝也: "オンラインショッピングを対象とした正確性と意外性のバランスを考慮したリコメンダシステム", 情報処理学会論文誌: データベース, Vol.46, No.SIG 13(TOD 27), pp.53-64, 2005