

トピックモデルを用いた消費者場面毎の競合分析

Competitor Analysis of Consumer Situations utilizing Topic Model

川中翔*¹ 宮田章裕*¹ 東中竜一郎*² 星出高秀*¹ 藤村考*¹
 Sho Kawanaka Akihiro Miyata Ryuichiro Higashinaka Takahide Hoshide Ko Fujimura

*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

*²日本電信電話株式会社 NTT サイバースペース研究所
 NTT Cyber Space Laboratories, NTT Corporation

We propose a method for analyzing competitive relations of brands. Our method calculates competitive scores of brand pairs utilizing similarities of *factor word* distributions. Here, we define *factors* as events or conditions that provide people with the opportunity to consume products or services, and *factor words* as words that represent each factors abstractly. Each brand pair has five variant competitive scores based on Belk's *consumer situations* framework. Furthermore, our method extracts specific competitive topics and competitive words of pairs. Experimental results show brand pairs which have high competitive score tend to be correct competitive relations.

1. はじめに

企業にとって競合関係にある対象の分析は重要である。競合企業や競合ブランドを適切に把握することで、市場の状況に則した戦略を採ることができる。競合関係は、マーケティング分野において、3C分析で顧客、自社と並ぶ分析対象となるなど、競合分析はマーケティングリサーチにおいて重要なタスクの一つである。また近年、グローバル化の進行に伴い、競合関係は様々な分野に跨る複雑な構造を示しており [Friedman 05]、競合関係の正確な把握は難しくなっている。

そこで本稿では、Web上の人々の書き込みから、ブランド(本稿では企業、店、商品などをブランドと表記する)間の競合関係を推定する手法を提案する。Web上には、一般の消費者による様々なブランドに対する口コミが発信されている。既に大量に存在する消費者の口コミを分析することで、効果的なブランド間の競合関係解析手法の確立を目指す。

提案手法では、任意のブランドに対して、消費者の消費行動シーンにおける主要な代替選択肢と成りうるブランドを競合ブランドと定義する。提案手法ではまず、各ブランドを対象とする消費行動について消費者が記述したレビュー記事群をWebから収集する。次に、レビュー記事を解析し、各ブランドの消費行動要因(ブランドを対象とする消費行動実施の決断にポジティブな影響を与えた事象や状態、条件)の特徴を表すベクトル(特徴ベクトル)を作成する。この際、消費者場面(Consumer Situations)[Belk 75]の分類に基づく5つの特徴ベクトルを作成する。続いて、各ブランド間の5つの特徴ベクトルの類似性を分析することで、ブランド間の消費者場面ごとの競合度を出力する。さらに具体的な競合理由を表す競合トピックと、競合要因語を出力する。

本稿の構成は次の通りである。2章では関連研究について述べる。3章では提案手法について説明する。4章では評価実験を報告し、5章にシステムイメージを描き、6章ではまとめを述べ締めくくる。

2. 関連研究

2.1 ブランド共起文を手がかりとする競合分析

従来に、ブランド共起文を用いて、Webから競合関係を解析する試みがなされている [Bao 08]。ブランド共起文とは、*Brand A versus Brand B*のような複数のブランド名が並列的に記述されている文である。[Bao 08]では、ブランド共起文などのブランドの共起性を手がかりとして、競合ブランドを獲得する。

ブランド共起文を用いた方法では並列的に語られないブランド同士は競合関係として獲得することができない。

2.2 ブランド毎の特徴データを比較する競合分析

Webから競合関係を解析する別の方法として、ブランドごとに個々にその特徴を表すデータ(特徴データ)を獲得し、ブランド間の特徴データの類似度や優劣から競合関係を推定する方法が考えられる。

特徴データを用いる第一の方法として、レビュー記事の出現語の類似度が高いブランド同士を競合とする方法が考えられる。しかしながら、全ての出現語を用いると、競合関係の解析に適さない単語が混入し精度が下がる問題がある。

特徴データを用いる第二の方法として、評価属性 [Bing 05][立石 04]を用いる方法が考えられる。評価属性はブランドを評価するための属性である。評価属性を用いる方法では、“良い”、“素晴らしい”などの評価表現(典型的には形容詞)を手がかりとしてその主語を獲得する方法などにより、評価属性を表す語(“デザイン”、“バッテリー”など)を獲得する。さらに、獲得した各評価属性に伴って出現する評価表現の極性(典型的には正負の実数値)を獲得し、ブランド毎の各(評価属性, 極性値)を特徴データとする。獲得された各評価属性の極性値を複数のブランド間で比較することで、“ある評価属性ではブランドAはブランドBに比べ勝っている”などのブランド間の関係についての知識を獲得する。実際に [Bing 05]では、獲得する評価属性を用いて、複数のブランドを比較するシステムを提案している。

しかしながら上記方法では、競合を解析するための評価属性が、評価表現との共起に依存するという問題がある。すなわち評価表現を伴わずに暗黙的に、競合を発見するための手がかりとなる属性を利用することができない。例えば、レストラ

連絡先: 川中翔, 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, 神奈川県横須賀市光の丘 1-1, 046-859-3675, 046-859-5552, kawanaka.sho@lab.ntt.co.jp

ンを訪れる時の“ベビーカー同伴”という消費者の状態についての情報が該当する。レストラン A とレストラン B がそれぞれ頻りに“ベビーカー同伴”という状態で頻りに用いられる場合、レストラン A とレストラン B は、競合関係の候補になり得るが、評価属性を用いる方法ではその競合関係の獲得に適していない。ベビーカー同伴等の状態は、一般に優劣を比較するための指標として消費者は用いないからである。

また、単語ベースの情報が利用され、そのまま用いると評価属性の数が増えた際にユーザが競合傾向を理解するための負担が高くなる。[立石 04] では、評価属性の分類をシステムの機能に含めているが、その方法はユーザの自由な操作に委ねており、専門知識が必要となる。

3. 提案手法

3.1 アプローチ

本研究では、関連研究のうちの後者の、ブランド毎に特徴データを獲得し、ブランド間で特徴データを比較することで競合関係を分析するアプローチを採る(競合ブランドの獲得にあたりブランド名の共起性に依存する問題を回避する)。既存の有力な特徴データを比較する方法である、評価属性を用いる方法には次の 2 つの課題がある。

1. 評価表現を伴わない競合属性^{*1}を利用できない
2. 競合属性を整理する方法が示されていない

上記課題に対し、本研究は次のアプローチを採る。

まず競合属性は、評価表現ではなく、消費行動表現(消費行動を実施したことを表す表現)を手がかりに獲得する(3.3 節)。消費行動表現を手がかりに用いることで、既存方法では獲得できなかった種類の競合属性の獲得が可能とする。なお、既存研究で獲得できた競合属性を、本研究で獲得できるかは保証されないが、互いに補完し合う技術という位置付けなので問題としない。また本研究では、競合属性(要因語)を消費者場面 [Belk 75] に基づき分類する(3.4 節)ことで競合属性を整理する。

3.2 提案手法の構成

提案手法は Web 上から収集するレビュー記事を入力として、各ブランド間の消費者場面毎の競合度を出力する。消費者場面は、競合関係の種類を表す属性であり、3.4 節に定義する。消費者場を用い、単に競合度のスコアを出すだけではなく、消費者場面毎の競合度スコアを出すことで、競合している観点をユーザは把握することができる。さらに競合ブランド同士が競合している理由を具体的に表す、競合トピックと競合要因語を出力(3.6 節)し、ユーザの具体的な理解を促進する。

提案手法は次節以降に示すとおり、要因語を獲得する処理、要因語を分類する処理、各ブランド間の競合度を算出する処理、競合トピックと競合要因語を出力する 4 処理から構成される。要因語は、個々の消費行動要因(1 章)を表す語であり、関連研究における評価属性に相当する。なお、本稿では事前にレビューサイトより収集したレストラン訪問についてのレビュー記事を例に説明する。また本稿では、収集した全てのレビュー記事を“コーパス”と表記する。コーパスの記事数は 400000、ブランド数は 44074 である。

表 1: 獲得した要因辞書語

| |
|---|
| 安い, ランチ, 二人, 結婚式, 家族, 2 次会, 宴会, 上司 デート, 友達, 同僚, ディナー, 評判: 良い, 接待, 大勢, 夕食, 居酒屋感覚, 紹介, 記念日, ベビーカー, カップル |
|---|

3.3 要因語の獲得

競合を分析するための手がかりとなる要因語の集合をコーパスより獲得する。本稿では消費行動要因を、消費行動実施の決断にポジティブな影響を与えた事象や状態、条件と定義し、要因語は消費行動要因を表す語である。要因語は、“ベビーカー同伴”などの評価表現を伴わない消費者の状態を表す語なども競合関係分析の情報として用いることができる。

各ブランドはレビュー記事中における要因語の出現頻度分布によって特徴付けられる(3.5 節)。要因語の獲得は我々の過去の提案 [川中 11] に順ずる。本章では、その処理を示す。

提案手法では、コーパス中において、次に示す消費行動表現を含む言語パターンに合致する語を要因語として獲得する(なお、下記表記において矢印の両端の語はこの順序で接続してのものとする)。

- 言語パターン L : (要因語) (接続標識) (消費行動表現)

消費行動表現とは、“来店”、“訪問”、“行く”などのレストラン訪問についてのレビュー記事において、消費行動を実施したことを表す表現である。消費行動表現を用いることで、消費行動が実施された時の、利用者の状態や目的を獲得し、競合分析のための特徴として活用することを意図している。消費行動表現は、[川中 11] で自動獲得した次の 11 語を与える。

- 行く, 訪問, 伺う, 着く, 並ぶ, 通す, 向かう, 訪れる, 連れる, 来店, 到着

接続標識は、消費行動表現の付近にある単語のうち、要因語になり得るものだけを獲得するための標識として用いられる単語である。本稿では接続標識は、人手で次の語群を与える。

- ので, ため, ということで, という訳で, で, と

コーパス中における、言語パターン L を満たす各語の回数を集計し、その回数が閾値 θ より高いものを要因語として出力する。上記までは予備実験で実施し、 $\theta = 2$ において、737 の要因語を獲得した。一例を表 1 に示す。

3.4 要因語辞書の分類

要因語をその性質に応じて人手で分類する。今回、対象となるレストラン訪問という消費行動は、Assael の関与水準 [Assael 87] が低い購買行動に分類されるため、消費者の購買行動の決断は、自らの信念などに大きく依存せず、状況的な要素がより強い影響を与える。ゆえに要因語の分類は、購買における状況や場面を考慮した分類法が適切と考えたため [Belk 75] の分類法を採用する。[Belk 75] は購買行動における消費者場面を (1) 物理的環境, (2) 社会的環境, (3) 時間, (4) 課題, (5) 先行状態の 5 つに分類した。また、本稿では、Belk らの分類に加え (6) ノイズという観点を加える。(6) ノイズには分類上不適切な語が該当する。実際の分類は、実験者 1 名がレビュー記事を参照しつつ、[Belk 75] の定義に従い実施した。分類の結果を表 2 に示す。本稿では、(1) から (5) の各クラスに分類された単語群を用い競合度を算出する(3.5 節)。

*1 競合分野を表す語である。関連研究では評価属性が該当する。

表 2: 要因語の分類

| 観点 | 数 | 要因語の例 |
|-----------|-----|----------------------|
| (1) 物理的環境 | 59 | 近い, 近く, 銀座, 行列:すごい |
| (2) 社会的環境 | 240 | 友達, カップル, 子供, 2名 |
| (3) 時間 | 30 | ランチ, 夜, モーニング, ブランチ |
| (4) 課題 | 88 | 忘年会, デート, 合コン, 女子会 |
| (5) 先行状態 | 156 | リクエスト, 紹介, 興味津々, 出張 |
| (6) ノイズ | 164 | 人数, 感覚, 目的, 用途, お客さん |

3.5 競合度の算出

本処理では, 各入力ブランド b に対する他ブランド c の, 消費者場面クラス n における競合度 $R(b, c | n)$ および $R(b, c, \Lambda)$ を算出する. $R(b, c | n)$ は, b と c が特定の消費者場面 n (例: 課題) で競合している度合いを表し, $R(b, c, \Lambda)$ はブランド間の総合的な競合度を表す. なお, 予め各ブランド b について, 消費者場面クラス n 毎に要因語出現頻度ベクトル θ_{bn} を作成しておくものとする.

3.5.1 トピックモデルの作成

消費行動要因が似ているブランド同士は競合であるという仮説に基づき, 各消費者場面クラス n 毎に, 各 b , 他の各 c との消費行動要因の類似度の高さによって競合度を測る. 単純に要因語の出現頻度ベクトルの類似度を用いると, データ量が不十分な場合に, データの偏りに大きく影響を受け良くないので, 次元を圧縮して類似度を測る. 次元圧縮の手法には $pLSI$ を用いる. $pLSI$ は文書などの対象集合における潜在的なトピックを検出するために用いられる技術であり, 一種の次元圧縮の方法と捉えることができる. $pLSI$ を用い, $\theta_n = \{\theta_{bn}\}_{b=1}$ が与えられたときに, 潜在的な各トピック z および各確率 $p(z | n), p(b | z, n), p(w | z, n)$ を算出する (w は各要因語である). 各確率の算出には EM アルゴリズムを用いる. なお z のトピック数 K はパラメータとして与える必要があり, 本稿では 10 とする.

3.5.2 消費者場面毎の競合度の算出

3.5.1 節の処理により, 各消費者場面クラス n 毎に, 各ブランド b の各トピックへの分布ベクトル ξ_{bn} が定められる. このとき, 次のスコア $R(b, c | n)$ を他ブランド c との消費者場面 n における競合度として算出する.

$$R(b, c | n) = \sum_z p(z | b, n) * \log \frac{p(z | c, n)}{p(z | b, n)} \quad (1)$$

$R(b, c | n)$ では, KL-divergence を用い, ブランド間のトピック分布の類似度を計算している. 本処理では, 全ての b, c, n の組み合わせについて上記 R スコアを算出する.

3.5.3 競合度の算出

消費者場面ごとの競合度を用いてブランド間の総合的な競合度を算出する (次式). なお, λ_n は各消費者場面クラスの重みであり, $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5\} \in \Lambda$ である.

$$R(b, c, \Lambda) = \sum_n \lambda_n \cdot R(b, c | n) \quad (2)$$

3.6 競合トピックと競合要因語の推定

ブランド同士 (b, c のペアとする) の競合理由を具体的に把握するために, 消費者場面ごとに, ブランド間で共通して帰属確率が高いトピック (競合トピック) および, 競合トピックと

表 3: 要因語の設問

| | |
|----|---|
| 3 | 一語だけで, 競合要因を表現する語 |
| 2 | 3 もしくは 1 のどちらともいえない |
| 1 | 他の語と合わせてはじめて競合要因を表現する語 |
| 0 | 上記 3,2,1 のいずれかに該当するが 両ブランドの競合要因としては不適切な語 |
| -1 | どのような競合要因を示唆しているか分からない語 |

関連する要因語のうち特に両ブランドに共通する消費行動要因を表す要因語 (競合要因語) を推定する. 消費者場面 n における競合トピックは次式 T で選ばれるトピック z とする.

$$T(b, c | n) = \forall_z \max(\min(p(z | b, n), p(z | c, n))) \quad (3)$$

$\forall_z \max(func(z))$ は全ての z における $func(z)$ の最大値を出力し, $\min(k_1, k_2)$ は k_1, k_2 のうち小さい値を出力する.

競合要因語のスコア W は次の式によって推定する (b, c 別々に獲得する. その際, 下記式の b は c と読み替えられる. b, c 毎に, スコア W の高い要因語上位 α (閾値) 件を競合要因語としてユーザに表示する). なお, 本節では w は $p(z | w, n) > 1/K$ を満たす要因語とする.

$$W(w | b, n, z) = p(z | w, n) * p_{score}(w | b) \quad (4)$$

$p(z | w, n)$ はトピック z と要因語 w の関係の強さを表し, $p_{score}(w | b)$ は w がコーパス全体に比べて b で出現し易いかをカイ 2 乗値によって評価するスコアとする.

4. 評価実験

4.1 概要

提案手法の有効性を確認するために, 被験者 2 名を用いた検証実験を行った. 検証実験では (1) 獲得した競合ブランドが適切か否か, (2) 競合ブランドの競合要因語が適切か否か, の 2 項目を評価した.

実験手順は次の通りである. (1) 予め用意した各 4 ブランドについて, システムは, 競合度が上位 5 件のブランドと下位 5 件のブランドを取得する (パラメータ $\lambda_n = 1.0$ とする). またシステムは, 上位 5 件のブランドについて各競合トピック (T 値上位 2 トピック) と各競合要因語 (1 トピックにつき最大 4 語) を取得する. (2) 被験者は, 各競合度上位 5 件, 下位 5 件のブランドについて設問文への選択肢から最も適切と感じるものを選択する. (設問文: 両ブランドは消費者の消費行動シーンにおける主要な代替選択肢である. 選択肢: 5: 非常に同意できる, 4: 同意できる, 3: どちらともいえない, 2: 同意できない, 1: 非常に同意できない) (3) 被験者は, 各競合要因語について, 選択肢 (表 3) から最も適切と感じるものを選択する.

なお, 検証実験では, 競合の候補となるブランドは, 記事数が 100 以上ある 513 のブランドに限定して実施した. また記事数と, 被験者のブランドに対する知識量が結果に影響を与えることを考慮し, 条件が異なる 4 ブランドを評価対象として採用した (知識量は被験者へのインタビューにより確認した). 採用した 4 ブランドは, カレーうどん屋 A (知識量少, 記事数 364), ハンバーガー店 B (知識量多, 記事数 356) ふく料理屋 C (知識量少, 記事数 113) ファミリーレストラン D (知識量多, 記事数 112) の通りである. また被験者がアンケートに答える際には, Web 等を利用し, 自由に調べて良いという条件とした.

表 4: 競合ブランドに対する選択肢数の分布 (2 被験者の平均値)

| 選択肢 | 5 | 4 | 3 | 2 | 1 |
|-------|-----|-----|-----|-----|------|
| 競合度上位 | 4.0 | 5.5 | 4.5 | 4.0 | 2.0 |
| 競合度下位 | 0.0 | 0.0 | 0.5 | 0.5 | 19.0 |

表 5: 競合要因語に対する選択肢数の分布 (2 被験者の平均値)

| 選択肢 | 3 | 2 | 1 | 0 | -1 |
|--------|------|------|------|-----|------|
| 競合要因語数 | 18.5 | 44.5 | 44.5 | 3.0 | 16.5 |

4.2 実験結果と考察

被験者アンケートの結果を表 4(競合ブランドへの被験者の選択肢数の分布), 表 5(競合要因語への被験者の選択肢数の分布)に示す。表 4 の通り, 競合度上位のブランドは, 競合度下位のブランドに比べて, 被験者が競合ブランドであると判断する傾向が高かった。提案手法が算出する競合度が有効に働いていることを示唆している。一方で競合度上位のブランドの中にも, 競合ブランドではないと評価する例もみられた。被験者のアンケート結果によると, 被験者は価格帯が明らかに違うブランド同士を互いに競合ブランドではないと判断する傾向があり, 価格帯が異なるブランド同士の多くが競合ブランドではないと判定されている。この問題が起きた原因として, 異なる価格帯のブランドに対するレビューにおいて, 同一の要因語が使用されていることが考えられる。今後はこの問題への対処が課題の一つである。

表 5 の通り, 競合要因語は選択肢 1 が選ばれるケースが多くあり, 多くの語が他の語と結びついてはじめて競合要因を表す語であることがわかった。“夕食”, “仕事”などが例である。今後はこれらの語を有効に活用するために, 他の語と組み合わせさせてユーザに提示するなどの提示法の工夫が課題である。

また, 一定数の選択肢 0 の語が選ばれた。選択肢 0 に選ばれた語にはラーメン屋とカレーうどん屋における“餃子目当て”のような片方のブランド(ラーメン屋)にとって適切で, もう片方のブランド(カレーうどん屋)にとって不適切な語が該当した。これは共に帰属確率が高い競合トピックがあったとしても, 含まれる全ての要因語が, 両ブランドにとって必ずしも適切ではないことを示している。今後は最適なトピック数の決定などのアルゴリズムの改善が課題である。

また一定数の選択肢-1 の語も存在した。選択肢-1 に選ばれた語の例には, “久しぶり”のような, ユーザの消費者場面を表すが, 想定されるシーンが幅広すぎて, 競合要因語として不適切と判断された例が多くみられた。今後は, これら不適切な要因語のフィルタリングが検討すべき課題である。

5. システムイメージ

獲得した競合関係情報をユーザに可視化するシステムイメージを示す。提案システムの実装イメージを図 1 に示す(実際は実ブランド名が表示されるが図 1 では伏せている)。提案システムでは①の入力欄にユーザが入力したブランドの競合ブランドが②の出力領域に表示される。競合ブランドは競合度 $R(b, c, \Lambda)$ の降順にソートされ表示される。③には, 各競合ブランドが競合している消費者場面を可視化するため, 各 $R(b, c | n)$ を偏差値化(入力ブランド b と全ての他各ブランド c との $R(b, c | n)$

が要素)した値をレーダーチャート形式で表示している。④には, 競合トピックの要因語や, 各ブランドの頻出語など, 競合要因を具体的に表す情報を示すことができる。また, ⑤のスライダーを操作することによって, λ_n の重みを変更した上で再度競合ブランドを競合度の降順に表示することができる。⑤により, “社会的環境と課題が似ている競合ブランド”などの, ユーザが求めるタイプの競合関係の獲得に柔軟に対応することを意図している。

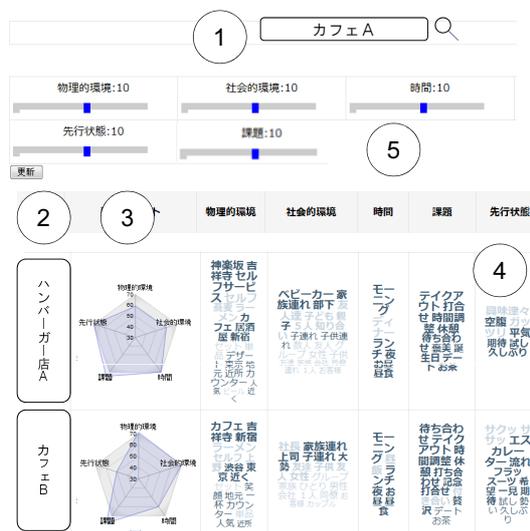


図 1: 可視化システム

6. おわりに

本稿では, 獲得した要因語とトピックモデルを用いて, 消費者場面毎に競合ブランドを発見する手法を提案した。評価実験において, 提案手法は競合ブランドを適切に獲得できていることを示したが一方で競合要因語の適切性などに改善の余地も確認され, 今後の課題である。

参考文献

[Bao 08] Bao, S. et al: Competitor Mining with the Web, IEEE Transaction on Knowledge and Data Mining, Vol.20, No.10, pp.1297-1310 (2008)

[Jindal 06] Jindal.N et al: Identifying Comparative Sentences in Text Documents, SIGIR-06 (2006)

[Belk 75] Belk, R.W.: Situational variables and consumer behavior, Journal of Consumer Research, Vol.2, No.3, pp.157-164 (1975)

[Friedman 05] Friedman.T.L: The World Is Flat: A Brief History of the Twenty-First Century. Farrar, Straus and Giroux (2005)

[立石 04] 立石健二 et al: Web 文書集合からの意見情報抽出と着眼点に基づく要約生成, 情報処理学会研究報告 2004(93), pp.1-8 (2004)

[Bing 05] Bing.L.et al: Opinion Observer: Analyzing and comparing opinions on the web. In Proceedings of WWW 2005, pp.342-351 (2005)

[Assael 87] Assael, H.: Consumer Behavior and Marketing Action, Kent Publishing (1987)

[川中 11] 川中翔.et al: 要因語辞書と出現位置を用いた消費行動要因分析, 第 80 回グループウェアとネットワークサービス研究発表会 (2011)