

要因語辞書と出現位置を用いた消費行動要因分析

川中 翔^{†1} 宮田 章裕^{†1} 東中 竜一郎^{†2}
星出 高秀^{†1} 藤村 考^{†1}

レビュー記事を入力として、様々な消費行動を引き起こした要因の推定とその影響力の算出を行う手法を提案する。本研究では、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、条件を要因と定義し、個別の要因を抽象的に表現する語を要因語と定義する。提案手法では、要因語を獲得するためにコーパスから生成する要因語辞書および記事中の語の出現位置を用いる。要因語辞書は、出現文脈に依存せず高確率で要因を表す語のリストであり、文脈からでは推定が難しい要因箇所に含まれる要因語を確実に要因と推定するために用いる。語の出現位置は、文脈情報の一つであり、要因語辞書に含まれない出現頻度の低い要因語を、適切に獲得するために用いる。実レビュー記事を用いた検証実験において、提案手法は比較手法に比べ高い適合率で要因語を獲得できることを示した。

Analysis of Consumer Decision Factors utilizing Factor Dictionary and Word Positions

SHO KAWANAKA,^{†1} AKIHIRO MIYATA,^{†1}
RYUICHIRO HIGASHINAKA,^{†2} TAKAHIDE HOSHIDE^{†1}
and KO FUJIMURA^{†1}

We propose a method for analyzing factors for consumer decisions. 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. Our method uses *Factor Dictionary*, that consists of words meeting the predefined requirements, to improve the precision of extracting *factor words* that are frequently used. We also employ the word position information within the document to obtain rare *factor words*. Compared to baseline methods, our method offers better performance in extracting *factor words*.

1. はじめに

企業にとって、顧客を知り、顧客にあった商品・サービスを提供することは重要である。顧客を知るための活動は一般にマーケティングリサーチ¹⁾ *1とよばれ、マーケティングにおいて重要なプロセスを担っている³⁾⁴⁾。マーケティングリサーチにおいて、“なぜ顧客は商品を購入したか”などの問いへの解となる消費行動要因^{*2}の分析は企業にとって重要である。自社に関係する消費行動要因を適切に知ることは、効果的なマーケティング活動に有用である。

現状、消費行動要因を獲得する代表的な方法として、質問調査^{*3}が存在する。質問調査は、顧客の行動ログデータ分析などの手法に比べ、既定の質問事項に対する回答が得られる一方で次の問題がある。(1) 実施コストが高い。質問調査は顧客への質問の回答というタスクの依頼である。ゆえにその実施コストは対象となる顧客数や顧客を拘束する時間に比例する。さらに、質問調査では(2) 回答が恣意的な方向へ誘導される傾向がある。質問調査は回答するシチュエーションや目的、また質問項目に回答内容が依存し、率直な意見を得られない可能性が多分に存在する。マーケティングリサーチにおいて、率直な意見を得られなかった例が報告されている⁵⁾。

上記問題に対し本研究では、ソーシャルメディア上の人々の書き込みから、消費行動要因を獲得する手法を提案する。レビューサイト、ブログをはじめとするソーシャルメディアには、人々が、要因を含め様々な情報を自発的に発信している。既に存在するこれらの情報を活用することで、顧客に質問への回答を依頼することなく、人々の生の声に基づいた大量の消費行動要因の獲得が可能である。

従来にも、Web上のリソースを用いて、ブランド(本稿では企業、店、商品などの消費行動要因を獲得する対象をブランドとまとめて表記する)の消費行動要因獲得に関連する研究がなされている。評判分析の枠組みにおいてはブランドの属性ごとの極性値を獲得する研究が存在し、因果関係分析の枠組みにおいては言語パターンを手がかりに事象間の因果関係を

†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

*1 日本のマーケティングリサーチの市場規模...1766 億円 (JMRA 調べ)²⁾

*2 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

*3 アンケートやインタビューが代表的な手法である。

獲得する研究が存在する。しかしながら前者には、獲得できる属性が評価表現との共起に依存するという課題(3.1節)がある。後者では、言語パターンを伴わない表現は獲得できないという課題(3.2節)があり、また人々の消費行動と要因の間の因果関係は研究の対象として着目されていない。

上記問題に対し、本稿では、コーパスから生成する要因語辞書および記事中の語の出現位置を用いる手法を提案する(4章)。要因語辞書は、出現文脈に依存せず高確率で要因を表す語のリストである。文脈からでは推定が難しい要因箇所に含まれる要因語を確実に要因と推定するために用いる。(4.2節, 4.3.1節)。語の出現位置は、文脈情報の一つであり、要因語辞書に含まれない出現頻度の低い要因語を、適切に獲得するために用いる。(4.3.2節)

以降、第2章にて研究目標を述べ、第3章では関連研究について記す。第4章では提案手法を説明する。第5章では評価実験を報告する。第6章ではまとめを述べる。

2. 研究目標

ブランドについてのレビュー記事を入力として、消費行動に影響を与えた消費行動要因(以降、要因と表記する)集合と各要因の影響力の高さを出力する技術の確立を目指す。ここで、ブランドについての消費行動に対するレビュー記事とは、ブランドの(商品やサービスの)購買について消費者が記述した記事を指す。また、要因の影響力の高さを、多くの人に影響を与えた要因ほど高く評価する指標とする。なお、本稿においては、レビュー記事中で事例が多い要因ほど多くの人に影響を与えているとする。

本技術の確立は、企業のマーケティングにとって有用である。本技術の応用として、様々な消費行動の要因を把握するためのシステムを作成することが可能となる(図1)。図1のようなシステムを通して、マーケティングが自社ブランドの消費行動に、高い影響力を与える要因を知ることができれば、効果的なマーケティング戦略の検討に有用である。さらに他社ブランドの要因と自社ブランドの要因を比較することで、他社の存在を考慮したマーケティング戦略の検討に有用である。上記のようなマーケティングシーンへの活用のために、獲得する要因の単位は次の要件を設定する。

1) 類似する要因事例を集約できること

異なる消費者によって書かれた類似する事例を集約できることが重要である。要因の集約ができないと、マーケティングは影響力が高い要因を知るために、一つ一つの事例に目を通し、類似する要因をまとめる必要が生じ負担がかかる。集約が適切に行われていると、要因の影響力の高さの集計を容易に行うことができ、他社と自社を比較する上でも有効である。

2) 要因事例を抽象化するラベルを付与できること

要因の傾向をマーケティングが見て把握するためには、理解を助けるために、各事例に抽象化された要因を表すラベルが付与されていることが望ましい。

4章に上記の目標を達成するための提案手法を述べる。

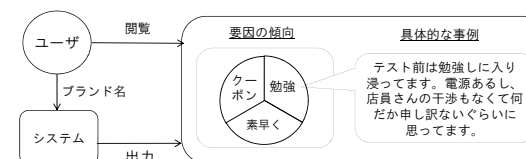


図1 システムイメージ

3. 関連研究

従来にも、Web上のリソースを用いて、ブランドの要因の獲得に関連する手法が提案されている。本章では従来研究と本研究との違いを述べる。

3.1 評判分析における評価属性抽出

ブランドの評判情報を獲得するために評価属性を用いる方法が従来提案されている⁶⁾⁷⁾。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、最終的に各評価属性とその極性値のペアの集合が評判情報として出力される。

本研究の目的の達成にあたり、評価属性を用いた方法を要因獲得への適用が考えられるが、技術課題1)獲得できる評価属性の評価表現との共起への依存により難しい。すなわち評価表現を伴わずに暗黙的に属性となる語をブランドの良い運として利用できない。例えば、レストランを訪れる時の“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が“ベビーカー同伴”という状態で頻繁に利用される場合、“ベビーカー同伴”という表現はレストランAの要因を表す語となるべきであるが、従来手法ではその獲得に適していない。“ベビーカー同伴”は消費者の状態を表す表現であって、評価する指標を添えてレビュー記事に記載される場合は稀だからである。

3.2 NLP(自然言語処理)における因果関係分析

NLPにおける因果関係分析⁸⁾⁹⁾¹⁰⁾(以降、因果関係分析と表記する)は、ブランドの要因の獲得に関連する。因果関係分析は、Webやニュース記事などの大規模な電子化文書集合から、一般的な常識や社会的事象に関する原因表現と結果表現の対を自動的に抽出することを目的としている。因果関係分析では、“ので”、“ため”のような因果関係を獲得するための“手がかり表現”を用いて、その周辺に存在する原因と結果の対を獲得する。例えば、“晴れたので観光客が多かった”という表現があるとき、〈原因:晴れ、結果:観光客が多い〉という因果関係を獲得することができる。

因果関係分析を、ブランドの要因の獲得に適用すると、たとえば“安いで行きました”という表現がレビュー記事中にあるとき、接続標識“ので”の前にある“安い”という表現は“行きました”という消費行動の実施の要因であると自動で判断する。さらに他のレビュー記事も分析し、消費行動Aについてのレビュー記事100件のうち、“安い”が要因として選ばれる回数が30回であるとき、消費行動Aの要因“安い”の影響力の高さは $30/100 = 0.3$ などとすることができる。しかしながら、レビュー記事の要因箇所の推定では、技術課題2)因果関係を表す接続標識の欠落による再現率の低下が障壁となる。因果関係分析で一般に分析対象となる新聞記事などと異なり、消費者は口語に近い形で記述するため、因果関係が記述される場合に必ずしも因果関係を表す接続標識を伴わない。事前の観察により“行く”(原形以外も含む)に係る“ので”という語は20レビュー記事に1回程度しか出現しない。上記性質により、対象となるレビュー記事を、因果関係を表す接続標識とのパターンマッチのみに基づいた方法では多様な要因事例を適切に獲得することができない。獲得する要因事例を増やすために、“行く”以外の行動表現の原因を獲得する方法も考えられるが、現状、消費行動を表す表現のリストは存在していないため自動では行えない。

3.3 本研究の位置づけ

本研究では、技術課題1に対応するために、評価表現に依存しないアプローチを採る(評価表現から評価属性を獲得することをしない)。また、技術課題2に対応するために、2つの方法を組み合わせる。方法の1つめは、要因箇所の推定にあたり、自動獲得する要因語辞書を用い、必ずしも手がかり表現を伴わない表現についても要因箇所か否かを推定する辞書ベースの方法である。方法の2つめは、要因箇所の推定にあたり、語の出現位置を用いる文脈情報ベースの方法である。詳細は4章に記す。

4. 提案手法

消費行動を引き起こした要因を表す要因語の獲得と、要因語の影響力の算出にあたり、コーパスから生成する要因語辞書および記事中の語の出現位置を用いる手法を提案する。要因語辞書は、出現文脈に依存せず高確率で要因を表す語のリストである。文脈からでは推定が難しい要因箇所に含まれる要因語を確実に要因と推定するために用いる。(4.2節, 4.3.1節)。語の出現位置は、文脈情報の一つであり、要因語辞書に含まれない出現頻度の低い要因語を、適切に獲得するために用いる。出現位置を用いる理由は、我々の過去の研究の文脈情報を用いた要因箇所推定¹¹⁾において、出現位置がF値向上に高い寄与を示したためである。(4.3.2節)。

要因語辞書と語の出現位置という2つの特徴を用いた方法を組み合わせることで、“ペーカカー”、“二次会”のような異なるブランド間で共通する一般的な要因語(主に要因語辞書により確実に獲得)およびブランド固有の商品名などの1ブランドに固有の要因語(主に語の出現位置により獲得)を適切に獲得することを意図している。

提案手法の構成を図2に示す。提案手法は任意の消費行動についてのレビュー記事を入力とし、要因語集合と、各要因語の影響力の高さを出力とする。要因語は要因を抽象的に表す語であり、4.1節で定義を与える。提案手法は前処理、要因語辞書作成、要因語影響力評価の3処理から構成される。本章では各処理の詳細を説明する。

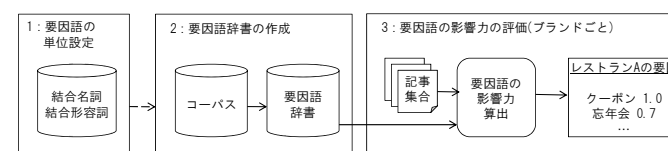


図2 提案手法の構成

4.1 要因語の単位設定

要因語の単位設定では2章で述べた集約の要件と抽象ラベルの要件を満たす要因の単位を具体的に決めるルールを与える。提案手法では、両要件を考慮し、獲得を目指す要因の単位は名詞1語、形容詞1語、名詞1語と形容詞1語のペアの3種のいずれかとする(以後これらを要因語と表記する)。要因語の単位を上記3種類とした理由は、事前の実際のレビュー記事の観察によりわかった下記の性質を考慮したためである。

- 名詞と形容詞で多くの要因が表現されている。
動詞等を獲得の単位としないのは、名詞や形容詞で多くの要因がカバーできることが観察により予想できることに加え、まずは対象を絞り問題の性質を単純にするためである。
- 要因の単位の語数を増やすと多くの要因は事例が一つとなり、マーケッターが全体の傾向を掴むことが困難になる。
例として、3つ組(例:“朝”, “メニュー”, “良い”)や4つ組を要因の最小単位とすると多くの要因は1回の消費行動においてしか観測されない。要因の数多くが事例が一つとなると、要因の影響力の高低がほぼ全ての要因で同一となり、マーケッターは全体の傾向を掴むことが難しくなる。

- 単語によって1語での意味の一意性に差がある。
本稿では要因語の意味の一意性を、1語で要因を一意に想像可能である度合いと定義する。このとき標記の性質が問題となる。例として、“安い”という表現は概ね価格の安さを表すのに使われているが、“多い”という表現は“量”や“人”など異なる複数の対象を評価するのに使われている。“多い”のような意味の一意性が低い語は、何を表しているのか情報が足らず、マーケッターにとって要因の理解する妨げになる。

上記3つの性質に対し、提案手法では要因の効果的な集約と意味の一意性保証の両立を目指し次の方針を採る。

- 1語で一意性が高い語は、1語で要因の最小単位とする。
 - 1語で一意性が低い語は、他の1語と組み合わせることで要因の最小単位とする。
- さらに、1語で一意性が高い語と一意性が低い要因語の区別のために次の仮説をもつ。
- 仮説 1:名詞を主語として伴う確率が高い形容詞(結合形容詞とする)は、主語名詞とペアではじめて高い一意性が保証される。
 - 仮説 2:形容詞を述語として伴う確率が高い名詞(結合名詞とする)は、述語形容詞とペアではじめて高い一意性が保証される。

提案手法では、上記仮説に基づき、予めコーパスを用いて、結合形容詞、結合名詞それぞれのリストを取得する。なお、本稿ではコーパスには、Webよりランダムに収集したレストラン訪問についてのレビュー記事400000件を指し、以降も同様とする。本研究では、結合形容詞、結合名詞作成のための確率 $p_{combine}$ を定める。(w は名詞もしくは形容詞とする。)

$$p_{combine}(w) = \frac{ptf(w)}{tf(w)} \quad (1)$$

$tf(w)$ は入力名詞(もしくは形容詞)のコーパス全体における出現回数である。 $ptf(w)$ は、

入力名詞(もしくは形容詞)が、次の条件のいずれかを満たす回数である。(矢印の前後は連続する語を表す。なお、パターン中に出現する副詞は削除する)。

- 条件 1:(形容詞) (名詞)
- 条件 2:(名詞) [が | は | も] (形容詞)

$p_{combine}$ が閾値より高い名詞・形容詞は、それぞれ結合名詞・結合形容詞とする。結合形容詞は、名詞を伴わずに用いられる場合は要因語として評価せず、名詞を伴って出現する場合は、その名詞とのペアで要因語とする。結合名詞も同様に、形容詞を伴わずに用いられている場合は要因語として評価せず、形容詞を伴って出現する場合は、その形容詞とのペアで要因語とする。

4.2 要因語辞書作成

要因語の影響力を評価する準備として、コーパスからの要因語辞書の作成を行う。要因語辞書は、全要因語の部分集合であり、明示的に消費行動の要因として用いられる回数が多い表現の集合を指す。また本稿では、要因語辞書の要素となる語を要因辞書語と表記する。要因辞書語は明示的に要因を表す文脈でレビュー記事に記述されていなくても、単に出現しているだけで、暗黙的に要因を示している傾向が高いと評価し、影響力の高さを測る際に重みを与える。例えば、仮に“クーポン”という語が要因辞書語であるとき、次のようにレビュー記事中の文において“クーポン”という語が“の”などの手がかり表現を伴わなくても、消費行動に影響を与えた要因語とする。

例文：“クーポン get!!行ってきました。”

提案手法では、要因辞書語の獲得にあたり、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獲得する。

- 言語パターン:(要因辞書語) (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。消費行動表現は自動獲得する(4.2.1節)。

接続標識は、消費行動表現の付近にある語のうち、要因辞書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

上記の消費行動を含む言語パターンを採用した理由は、消費行動を実施した表現に、直接的につながる表現には消費行動を実施した経緯や、条件、関連情報が多く含まれ、それらはブランドの要因となると考えたためである。

以下、4.2.1 節では、消費行動表現を獲得する処理について、4.2.2 節では、獲得した要因辞書語について述べる。

4.2.1 消費行動表現の獲得

多くの要因語を獲得するために、消費行動表現を自動獲得する。提案手法では、事前の検討により、“レストラン訪問”を対象とする消費行動表現に次の傾向があることを確認した。

- レビュー記事の冒頭に出現する傾向が高い
- 直前に出現する助詞の中では、“に”、“へ”の頻度が高い

上記の性質を持つ語は、消費行動表現であるとの仮説を設定し、コーパス中における条件 1, 条件 2, 条件 3 を全て満たす語 w_a を消費行動表現として獲得する。(なお、 w_a の品詞は動詞:自立(可能動詞を除く)もしくは名詞:サ変接続である場合に限る)。

- 条件 1: $\frac{Pos_{a1}(w_a)}{|D|} > \lambda_1$
- 条件 2: $\frac{Pos_{a1}(w_a)}{df(w_a, D)} > \lambda_2$
- 条件 3: 直前に出現する助詞のうち最も頻度が高い語が“に”もしくは“へ”である。

$$Pos_{a1}(w_a) = \sum_{d_k \in D} dp(w_a, d_k) \quad (2)$$

$$dp(w_a, d_k) = (0.5 - dq(w_a, d_k)) \quad (3)$$

$$dq(w_a, d_k) = \frac{w_a \text{ が } d_k \text{ 内で最も早く出現した位置 (冒頭からの語数)}}{d_k \text{ の語数}} \quad (4)$$

D はコーパス全体を表す。 $|D|$ はコーパスの記事数である。 $df(w_a, D)$ は D における w_a の出現記事数を示す。 $dp(w_a, d_k)$ は w_a がレビュー記事の中央の語より前に出現する際に正の値を返し、中央の語より後ろに出現する際に負の値を返す。語 w_a の冒頭に出現する傾向を評価するために用いる。 $dq(w_a, d_k)$ は w_a のレビュー記事 d_k における相対的な出現位置を表す。

上記は予備実験において実施し、様々な閾値を試し獲得した結果、 $\lambda_1 = 0.005$, $\lambda_2 = 0.083$ が、誤りを少なく、多くの表現を獲得できることがわかったので実験で用いる(実験者をレビュー記事を実際に見て、獲得した各語が消費行動表現として適切かを判定した)。結果として表 1 の表現を獲得した。なお、表 1 中の“する”という語は不適切な語であると人手で評価し、4.2.2 節の要因語の獲得の際は用いないとする。

4.2.2 獲得した要因辞書語

要因辞書語の獲得は次の手順で行う。

- (1) コーパスから、4.2 節で設定した言語パターンに合致する各語の回数を集計

表 1 獲得した消費行動表現

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

- (2) 各語について (1) で求めた回数が、閾値 λ_3 より高い語を要因辞書語として出力上記までは予備実験で実施し、 $\lambda_3 = 2$ で 737 の要因辞書語を獲得した。獲得した要因辞書語の一例を表 2 に示す。要因辞書語の評価は評価実験 (5 章) において行う。

表 2 獲得した要因辞書語の一例

安い, ランチ, 二人, 友人, 家族, 車, 2 次会, 宴会, 上司, デート, 友達, 同僚, ディナー, 良い評判, 接待, 大勢 思いつき, プランチ, 結婚式, 接待, 忘年会, 誕生日祝い, 夕食, 居酒屋感覚, 紹介, 記念日, グループ, カップル, 2 名, 送別会, 仕事, テイクアウト, タクシー, 二次会, モーニング, 宴会, プライベート, 合コン, 平日ランチ, 会食

4.3 要因語影響力評価

任意のブランド b_j についてのレビュー記事集合が入力として与えられたときの、各要因語 w_i の影響力の高さ $E(w_i, b_j)$ を算出する。

$E(w_i, b_j)$ は次に示す、 $DicDF$ (要因語辞書ベースの指標)と $PosDF$ (文脈情報ベースの指標)の線形結合式であり、次節以降で関数 $DicDF$ と $PosDF$ を説明する。

$$E(w_i, b_j) = \lambda_4 \cdot DicDF(w_i, b_j) + \lambda_5 \cdot PosDF(w_i, b_j) \quad (5)$$

4.3.1 $DicDF$

$DicDF$ は“ベビーカー”、“二次会”のような一般的な要因語を確実に獲得するために有効である。 $DicDF$ は要因語 w_i が要因語辞書 D_{dic} に含まれているか否かを評価し、要因語辞書に含まれている語に限り df をスコアとして与える。辞書に含まれていない語のスコアは 0 とする。

$$DicDF(w_i, b_j) = \begin{cases} df(w_i, b_j) & (w_i \in D_{dic}) \\ 0 & (else) \end{cases} \quad (6)$$

4.3.2 $PosDF$

$PosDF$ は要因語の各文書における出現文脈から、要因として用いられているかを評価する指標である。出現文脈からの要因らしさの推定に関し、要因語の出現位置が適合率向上に大きく寄与することをわれわれは報告した¹¹⁾(レビュー記事では執筆者は消費行動を行うまでの経緯を冒頭に書くことが多く、要因がその中に含まれていることが多かったためだと

考えられる)． $PosDF$ を用いることで，要因語辞書の要素とならない出現頻度が低い要因を表す語を獲得することができる．(例えばブランド固有の商品名のような 1 ブランドに固有な語が該当する．)

本稿ではその知見を利用し，要因語の出現位置を考慮した素性 $PosDF$ を，要因語の影響力の高さを文脈情報から測る指標として用いる． $PosDF$ は大規模コーパスに数多く登場しないような要因語も， b_j における影響力の高さを測るために用いる． $PosDF$ の式は次の通りで，各単語が冒頭に登場するほど，またその回数が多いほど値が高くなる (λ_6 はパラメータである)．

$$PosDF(w_i, b_j) = \sum_{d_k \in D(b_j)} dp_{os}(w_i, d_k) \quad (7)$$

$$dp_{os}(w_i, d_k) = \begin{cases} dp(w_i, d_k) - \lambda_6 & (dp(w_i, d_k) > \lambda_6) \\ 0 & (else) \end{cases} \quad (8)$$

5. 評価実験

提案手法の有効性を評価するための Web より収集した実際のレビュー記事を用いた実験を実施した．検証・分析する内容は次の 2 つである．

- 要因辞書語の適切性の検証とその性質分類
- ブランドの要因語の影響力ごとの適切性の検証

5.1 要因辞書語の適切性の検証と性質分類

5.1.1 検証の概要

要因語辞書 (4.2 節) の適切性の評価および性質を分析するために，要因辞書語の分類を人手で行った．今回，分析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¹²⁾ 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，Belk の分類法¹³⁾ を採用する．Belk の分類法では，消費行動要因は，(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 つの消費者場面 (Consumer Situations) に分類される．(1) 物理的環境は，消費者場면을現す有形の特性全般を表す．(2) 社会的環境は他者の存在を意味し，(3) 時間は時間に関連する情報を意味する．(4) の課題は消費者が持つ特定の目標や目的であり，(5) の先行状態は，消費者がその場に持ち込んだ一時的な状態である．本研究では，Belk らの分類に

加え，要因の単位として不適切な語の存在を確かめるため (6) ノイズという観点を加える．(6) ノイズには，語のラベルを見たときに，一意に意味が分からない語や，明らかに要因を表していない語が該当する．

5.1.2 検証結果

分類は，実験者 1 名が実際のレビュー記事を参照しつつ，Belk の定義と上記ノイズの定義に従い実施した．分類の結果を表 3 に示す．まず $\lambda_3 = 2$ で獲得した 737 語中 574 語を要因語辞書として適切 ((6) ノイズ以外) と判定したので，適合率は $574 / 737 = 0.78$ である．また図 3 に， λ_3 を変化させた場合の適合率と獲得語数を示す． $\lambda_3 = 3$ では 421 語を獲得し適合率は 0.79， $\lambda_3 = 4$ では 348 語を獲得し適合率は 0.80， $\lambda_3 = 5$ では 348 語を獲得し適合率は 0.82 となっており， λ_3 の値が低くなるに従い，適合率が下がる傾向がある．

各消費者場面ごとの分類では (2) の社会的環境についての語が 240 で最も多かった．“5 人”，“ふたり”などの同行人数，“カップル”，“上司”，“子供”などの同行者との関係を表す語などが抽出された．(5) の先行状態についての語も，“買い物”，“出張”などの前後の行動を表す語や，“思いつき”，“試し”などのユーザの心境を表す語が抽出できた．(4) の課題についても 88 語と数多くの語が獲得できた．店舗へ行く目的として，“忘年会”，“デート”，“パーティー”など食事自体以外についての語が数多く抽出されており，店舗の利用シーンを正しく分析する上で有用であると考えられる．(1) の物理的環境についての語は“近い”，“近く”，“銀座”などの場所に関する語や，“すごい行列”などの店舗の状態を表す語が抽出できた．(3) の時間についての語は，“夜”，“ランチ”などの一般によく使われる語に加え，“ランチ”，“週末ランチ”，“日曜ランチ”などより時間帯を限定した語などが多く獲得できた．

表 3 要因辞書語の分類

番号	消費者場面	数	要因辞書語の例
(1)	物理的環境	59	近い, バー, 銀座, すごい行列, ラーメン屋
(2)	社会的環境	240	友達, カップル, 5 人, 子供, 2 名
(3)	時間	30	ランチ, 夜, モーニング, 週末ランチ, ブランチ
(4)	課題	88	忘年会, デート, 合コン, 女子会, お忍び
(5)	先行状態	156	紹介, ノリ, 興味津々, 出張, 思いつき
(6)	ノイズ	89	人数, 店, 目的, 用途, 自分

今回は人手で分類したが，消費者場면을自動で分類できるようになれば，要因語の活用上有効であり，今後の課題である．

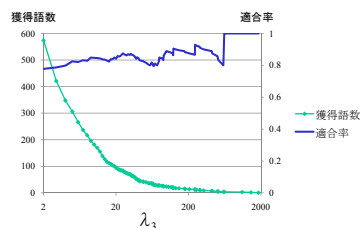


図 3 要因語辞書作成

5.2 ブランドの要因語の適切性

5.2.1 実験方法

提案手法 (4.3 節) によって獲得するブランドの要因語の適切性を検証する。検証は次の手順で行う。

- (1) 知名度が高い消費行動 b_x についてのレビュー記事集合 $R(b_x)$ を Web より収集する。
- (2) $R(b_x)$ から、提案手法、比較手法によって各要因語とその影響力の高さを獲得する。
- (3) 各手法で獲得した要因語の影響力上位 n 語、影響力中位 n 語、影響力下位 n 語について、実験者により適切か否かの 2 値評価を実施、結果を比較する。

各要素は次のように設定する。 b_x は実験者が評価しやすいように知名度が高いハンバーガー店 A への訪問という消費行動を設定し、Web より収集した 356 件のレビュー記事を用いる。影響力上位 n 語は獲得された要因語のうち最上位から影響力の降順に n 語をサンプルとして用いる (下位 n 語も同様に最下位から昇順に n 語を用いる)。影響力中位 n 語は影響力の中央値を持つ要因語を中心とする周辺 30 語をサンプルとして用いる。

提案手法と比較手法の詳細を次に示す。

- 提案手法 1: 提案手法
- 提案手法 2: 提案手法 ($PosDF$ のみを利用, $\lambda_4 = 0$)
- 比較手法 1: $R(b_x)$ での各語の df を影響力の高さとする方法
- 比較手法 2: $R(b_x)$ での次の言語パターンをみだす各語の回数を影響力の高さとする方法 (各語) (因果関係を表す接続標識) (否定形を除く“行く”)
- 比較手法 3: $R(b_x)$ での次の言語パターンをみだす各語の回数を影響力の高さとする方法 (各語) (因果関係を表す接続標識) (否定形を除く動詞)

なお、比較手法における要因語の品詞は名詞と形容詞とする。提案手法 1 の閾値 λ_4 と λ_5 は、指標 $PosDF$ 、と指標 $DicDF$ の最大値を取る要因語が 1.0 となるよう設定する。また

提案手法 1 の $DicDF$ では、5.1 節の実験において不適切と評価された語も要因語辞書の要素として用いている。 λ_6 は事前の観察により、消費者が明示的に記述する要因表現はレビュー記事の冒頭 15% に集中したため、0.15 と設定する。形態素解析器には Mecab¹⁴⁾ を用いる。適合率評価のためのパラメータ n には 30 を与える。

5.2.2 実験結果

各手法の適合率を表 4 に示す。提案手法 1、提案手法 2 は比較手法 1、比較手法 3 に比べ概ね高い適合率を示した。比較手法 2 では、要因語は 18 語だけ獲得された。獲得した 18 語は高い適合率を示した。

表 4 各手法の適合率

範囲	上位 30 語	中位 30 語	下位 30 語
提案手法 1	0.73	0.73	0.70
提案手法 2	0.80	0.73	0.70
比較手法 1	0.50	0.43	0.63
比較手法 2(*)	0.78(*)	-	-
比較手法 3	0.63	0.53	0.63

(*) 比較手法 2 では要因語は 18 語しか獲得されなかった。

5.2.3 考察

提案手法は、比較手法 1、比較手法 3 に比べ概ね適合率が上回った。各手法が獲得した語を実際に調べたところ、“ない”、“多い”のような一意に要因を表現していない語が適合率の差に影響を与えていることが分かった。これらの語が比較手法では多く獲得され、提案手法では比較的少なかった。一方で、提案手法においても“前”、“店内”のような一意に要因を表現していない一般語が一定数選出されており、それらを排除するためのルール改善が課題である。また比較手法 1、比較手法 3 では“ひどい”、“狭い”に代表されるネガティブな語が上位 30 語に確認された。これらは実験者による評価では、不適切と判断されており、これらの語を獲得してしまったことが適合率の低下に影響を与えたと考えられる。

提案手法 3 ($PosDF$) は提案手法 1 ($DicDF$, $PosDF$) に比べ上位 30 語で高い適合率を記録した。また提案手法 1 の適合率は 5.1 節における要因語辞書の適切性の平均値に比べ低い結果となった。これらに共通する原因として、“前”、“店内”のような、誤って要因語辞書に登録されている出現頻度の高い一般語が、適合率を下げていることが考えられる。

提案手法 2 では、抽出された要因語を観察したところ、ブランド固有の商品名や、ブランド固有のサービス名などの要因語辞書に出現しないブランド固有の要因語を獲得できてい

ることがわかった．これらの要因語は *PosDF* を採用した意図通りに働いた例である．

比較手法 2 は，要因語を 18 語しか抽出できず，3.2 節で指摘した問題通りの結果となった．

考察のまとめとして，今回の実験では，提案手法は，比較手法 1，比較手法 3 に比べ，影響力の高低にかかわらず高い適合率を示した．また比較手法 2 は，獲得した 18 語については高い適合率を記録したが，数多くの要因語を獲得できなかった．提案手法は，数多くの要因語について比較的高い適合率を示し，4 章で示した提案手法のアプローチが有効に機能したと考えられる．一方で再現性の評価は充分にできていないので，検証を適切に行うための実験の設計も含めてその評価は今後の課題である．また，本研究の応用として図 4 のようなサービスが実現可能である．

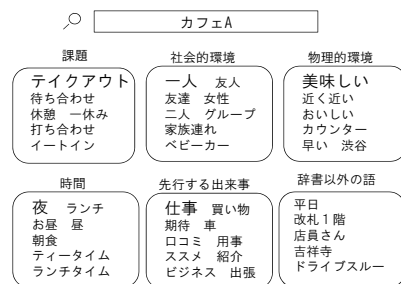


図 4 ブランドごとの影響力が高い要因語の表示

6. おわりに

本稿では，レビュー記事を入力として，ブランドに対する各要因語と各影響力の高さを出力するという新しい問題に取り組み，コーパスから生成する要因語辞書と要因語の出現位置を用いる手法を提案した．要因語辞書は，出現文脈に依存せず高確率で要因を表す語のリストであり，文脈からでは推定が難しい要因箇所に含まれる要因語を確実に要因と推定するために用いた．語の出現位置は，文脈情報の一つであり，要因語辞書に含まれない出現頻度の低い要因語を，適切に獲得するために用いた．実験において，提案手法は比較手法に比べ要因語を高い適合率で獲得できることを示した．また要因語辞書の要素である 737 の要因語について，その適切性を評価し，77.8%の語が要因語として適切であることが分かった．さらに要因語の性質を明らかにするために，消費者場面に関する 6 観点から人手で分類した．

今後は要因語の自動分類や再現率の評価が課題である．

参 考 文 献

- 1) Kotler.P et al: Marketing Management(12th Edition), Prentice Hall (2006)
- 2) 田下憲雄.: マーケティング・リサーチ業界の現状と将来展望, デジタル・マーケティング NEXT2009 (2009)
- 3) Ries. A et al.: ポジショニング戦略, 海と月社 (2008)
- 4) 魚谷 雅彦.: こころを動かすマーケティング コカ・コーラのブランド価値はこうしてつくられる, ダイヤモンド社 (2009)
- 5) 藤井大輔.: 「R25」のつくりかた, 日経プレミアムシリーズ (2009)
- 6) 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, NL-163, pp. 178 (2004)
- 7) Bing.L.et al:Opinion Observer: Analyzing and comparing opinions on the web. In Proceedings of WWW 2005, pp.342-351 (2005)
- 8) Higashinaka, R. et al. : An unsupervised method for learning generation dictionaries for spoken dialogue systems by mining user reviews, ACM Transactions on Speech and Language Processing, Volume 4, Issue 4, Article 8 (2007)
- 9) Khoo, C.S.G.: Extracting Causal Knowledge from a Medical Database Using Graphical Patterns, In Proceedings of 38th Annual Meeting of the ACL, pp.336-343 (2000)
- 10) 坂地泰紀 et al.: 構文パターンを用いた因果関係の抽出, 言語処理学会第 14 回年次大会論文集, pp.1144-1147 (2008)
- 11) 川中翔 et al.: レビュー記事群を用いた消費行動要因のマイニング, 第 151 回データベースシステム研究会 (2010)
- 12) Assael, H.: Consumer Behavior and Marketing Action, Kent Publishing (1987)
- 13) Belk, R.W.:Situational variables and consumer behavior, Journal of Consumer Research (1975)
- 14) Kudo, T et al.: Applying conditional random fields to Japanese morphological analysis, In Proceedings of ESCA Eurospeech, pp.2707-2710 (1997)