Vol.2010-DBS-151 No.5
2010/11/12

レビュー記事群を用いた消費行動要因のマイニング

川中 翔[†1] 宮田章裕[†1] 東中竜一郎[†2]
星出高秀[†1] 藤村考[†1]

レビュー記事を用いて，要因という観点から人々の消費行動を整理する枠組みとそのために必要な手法を提案する．本研究では，消費行動を選択するに至った主要な原因となる事象や状態を要因と定義する．提案手法は要因が書かれている箇所を推定するために，記事内の位置情報，消費行動表現，消費行動後に頻出する表現などを利用することを主な特徴とする．実験により，従来手法に比べて高精度に要因箇所を推定できることが確認できた．

Mining of Factors Affecting Consumer Decisions using Review Pages

Sho Kawanaka,[†1] Akihiro Miyata,[†1]
Ryuichiro Higashinaka,[†2] Takahide Hoshide[†1]
and Ko Fujimura [†1]

We propose a method for analyzing relations between consumer decisions and factors that affect decisions. In our research, we define consumer *factors* as events and conditions which mainly cause consumer decisions. Our method classifies keywords into factors and others by utilizing clues such as positions, expressions of consumption and frequently-appearing expressions after expressions of cosumptions. Experimental results show our method can accurately classify into factors and others than a simple baseline method.

†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. 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)2)⋆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[4)5)]．マーケティングリサーチのうち，顧客のコメントから “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[⋆2]の要因[⋆3]を分析する手段として質問調査[⋆4]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他の手法に比べ，顧客に明示的に要因を問うことができる利点がある一方で次のような問題がある．(1) 顧客を一定時間拘束するため実施コストが高く，大規模もしくは継続的な調査を行うことは容易ではない．(2) 回答が恣意的な方向へ誘導される可能性がある (率直な回答が得られなかった例が報告されている[6)])．

上記問題に対して，本研究では，顧客が説明する要因情報の情報源について，質問調査からではなく，ソーシャルメディア上の人々の書き込みを収集するアプローチを採ることで解決を図る．ソーシャルメディアには，人々が消費行動要因を含め様々な観点から自発的に情報を発信している．既に存在するそ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と考える．

消費行動要因の獲得には，ソーシャルメディア上の記事における，要因と消費行動という因果関係が記述された箇所を推定する必要がある．しかしながら，ソーシャルメディア上の一般ユーザが書いた経験記事[⋆5]には，因果関係を記述する場合においても，必ずしも因果関係を表す接続標識が使われない．そのため既存に利用されている因果関係を表す接続標識を推定の特徴に用いても充分な性能が保証されないという課題がある．そこで本研究では，経験記事中の要因箇所を推定するために，記事内の位置情報などの経験記事の構成を解釈する手がかりを利用する手法を提案する (5 章)．さらに本研究では，上記推定手法に加え，レビューサイトにおける経験記事の属性情報を利用することで，要因という観点から人々の消費行動を整理する枠組みを提案する．

本稿の構成は下記の通りである．2 章では消費行動と要因の定義について述べ，3 章では

⋆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)[3)]
⋆2 消費行動の例…‘‘携帯機種 X を購入する’’ ‘‘ハンバーガー店 A を訪れ食事をする’’ など
⋆3 顧客がその消費行動を選択した主要な原因
⋆4 アンケートやインタビューが代表的な手法である．
⋆5 経験記事とは人々が自らの経験について記した記事を指す

サービスイメージと研究目標，アプローチについて説明する．4 章では研究課題について議論し，5 章では提案手法について述べる．6 章では実験について報告し，7 章で関連研究，8 章でまとめを述べ締めくくる．

2. 消費行動とその要因

2.1 消費行動の定義

本稿では大辞林[7] の定義に基づき，消費行動を**生産された財貨・サービスを使うこと**と定義する．消費行動の例を次に示す．

- 購買：携帯機種 X を買う
- レストラン来店：ハンバーガー店 A に行き，食事をする

2.2 消費行動要因の定義

本稿では，消費行動要因を**人々が消費行動を選択するに至った主要な原因となる事象や状態**と定義する．以降では消費行動要因を単に要因とよぶ．要因は消費行動より時間的に前の事象や状態である．人々が書いた記事中の要因表現の例を次の例文の下線部に示す．
例：友人に勧められてたし，クーポンをもらったのでハンバーガー店 A に行きました．

3. 目標とアプローチ

3.1 目　　標

本研究では，マーケッターをユーザとして想定した要因を切り口にした消費行動分析システムの作成を目指し，本稿ではそれに必要な技術の確立を目標とする．システムの概要を表 1 に，システムイメージを図 1 に示す．

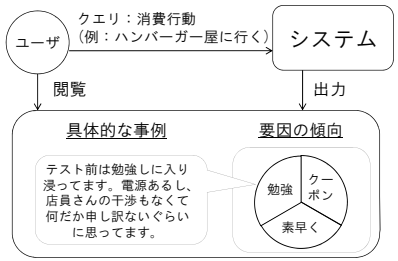

図 1　システムイメージ

表 1　システムの概要

ユーザ	任意の消費行動の，要因の情報 (傾向や具体例など) について情報要求を持つユーザー (マーケッター)
入力	任意の消費行動を表すクエリ
出力	クエリによって表現される消費行動についての要因の傾向と， 消費行動とその要因についての経験が記述された記事．また記事は重要度の高いものから出力される．

利用シーンについて例とともに示す．ユーザとして，人々が “ハンバーガー屋に行く” 要因について把握したいと考えるマーケッターを想定する．そのときマーケッターは，クエリ “ハンバーガー屋に行く” をシステムに入力すると，人々が “ハンバーガー屋に行く” 要因の傾向を知ることができる．また要因が書かれた具体的な人々の経験記事を閲覧することができる．このようにシステムを利用することで，マーケッターは消費行動について俯瞰的視点と実際の人々の書き込みの両方から理解することが可能になる．

前述のシステムを実現するために，本研究ではソーシャルメディア上の記事から消費行動と要因についての情報を獲得する方法を採る．ソーシャルメディア上の情報を活用することで，質問調査における問題 (1 章) を避けて消費行動要因の獲得が可能と考えるからである．必要となる技術について定式化した形で下記に示す．

人々が自らの消費行動について記述した記事集合 $\mathbf{D} = \{d_i\}$ があるとき，まず各 d_i が何の消費行動について書かれた記事か表す消費行動クラス集合 $\mathbf{A} = \{a_i\}$ の定義および各 d_i の各消費行動クラスへの帰属度を与える関数 $\Psi_{da} : \mathbf{D} \times \mathbf{A} \longrightarrow \{0, 1\}$ の作成が必要となる．また同様に，各 d_i に記述された消費行動が何の要因に基づいたか表す要因クラス集合 $\mathbf{F} = \{f_i\}$ の定義および，各 d_i の各要因クラスへの帰属度を与える関数 $\Psi_{df} : \mathbf{D} \times \mathbf{F} \longrightarrow \mathbf{R}$ の作成が必要となる（なお，本稿では実数全体を $\mathbf{R}$ と表記する）．最後に各消費行動の要因の傾向を示すためには，消費行動クラスと要因クラスの間の関係を表すスコアの算出が必要である．すなわち，消費行動要因関係関数 $\Psi_{fa} : \mathbf{F} \times \mathbf{A} \longrightarrow \mathbf{R}$ の作成が必要となる．

本稿では特に，適切な Ψ_{df} を作成するために必要となる，経験記事中の要因が書かれている箇所を推定する手法を提案する (5 章)．

3.2 アプローチ：レビューサイトの利用

我々は 3.1 節で述べたシステムを実現するために，ソーシャルメディアの中でも特にレビューサイトに着目する．

3.2.1 レビューサイトの特徴と採用理由

我々はソーシャルメディアを次のように分類する．

- レビューサイト
  - 各レビュー記事が 1 つ以上の消費行動と対応するサイト.
- 非レビューサイト
  - blog, twitter[⋆1]などの，各記事が特定の消費行動と必ずしも対応しないサイト.

レビューサイトの特徴として，消費行動に対する人々の経験に関する記述が集積していること，各レビュー記事が一つ以上の消費行動に対応していることが挙げられる．後者の例として，“ハンバーガー屋 A 店” のレビュー記事があるとき，そのレビュー記事に対応する消費行動は “ハンバーガー屋 A 店に行く” であることが，本文を参照することなく分かる．

このように各記事が何の消費行動について書かれたものか本文から判別する必要がないという利点の活用を意図し，本研究ではレビューサイトを情報源として用いる．

3.2.2 レビューサイトのデータ構造

一般にレビューサイトでは，各レビュー記事が消費行動と対応しており，各レビュー記事はカテゴリやタグにより意味付けされている．ゆえに消費行動クラス集合 $\mathbf{A} = \{a_i\}$，消費行動クラス関数 $\Psi_{da} : \mathbf{D} \times \mathbf{A} \longrightarrow \{0, 1\}$ はレビューサイトにより与えられる．a_i の例は “ハンバーガー屋 A 店に行く” などである．

4. 研 究 課 題

前章で述べた目標を達成するためには，経験記事の要因箇所を推定，適切に分類する必要がある．既存の要因表現抽出に関する研究では，要因箇所の推定のために，要因を表す表現と結果を表す表現の間などの周辺に存在する “手がかり表現” を用いた手法[8)9)10)11)12)13)14)15)16)]が存在する．手がかり表現として典型的なものは，“ので”, “ため” のような因果関係を表す場合に用いられる接続標識である．このとき，ソーシャルメディア上の記事を分類するタスクにおいては，既存手法の適用を考える上で次の 2 つの課題がある．

技術課題 1: 一般ユーザが書いた記事における因果関係を表す接続標識の欠落

技術課題 2: 一般ユーザが書いた記事における同一単語の要因性のゆれ

4.1 因果関係を表す接続標識の欠落

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，一般ユーザ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．

例えば，要因クラス：“気兼ねなく電源使えるので”，消費行動クラス：“ハンバーガーショップ A に行く” である一連の経験は，次の下線部のように記述することができる．

例 1）気兼ねなく電源が使えるのでハンバーガーショップ A に行きました。

一方でソーシャルメディア上においては，人々は思い思いの表現を用いる．上記と同様の経験を，下記のような形で記述され得る．

例 2）電源いっぱしあるしハンバーガーショップ A をチョイス。

例 3）周りきにせず電源を使えるって素晴らしいと思ってハンバーガーショップ A へ Go!!

例 2 や例 3 には因果関係を表す接続標識が無く，要因箇所を推定するには単純な手がかり表現のみを特徴とする手法では難しい．

4.2 記事における同一単語の要因性のゆれ

要因箇所を推定する処理を省き，文書分類のアプローチから記事を要因について分類する方法が考えられる．例えば，予め要因クラス毎に特徴語ベクトルを作成し，記事の特徴語の出現頻度との類似度などを特徴として分類を行う方法がある．しかしながら，一般ユーザが書いた記事には，要因となるような消費行動前の事象や状態のみならず，消費行動をした後の消費行動に対する評価や消費行動を行っている最中の情報などが存在する．同一の単語であっても，文脈によって要因として用いられているものとそうでないものがある．例を示す．次の 2 文において “ガイドブック” という単語が例 1 では要因として用いられ，例 2 では要因として用いられていない．このように文書分類のアプローチを単純には適用できない．

例 1)ガイドブックをただただ信じて○○に行ってきました

例 2)○○に行きました。肉のボリュームも多く。思ったよりも良かったです。
ガイドブックに載るべきだと思います。また行きたい。

5. 提 案 手 法

レビューサイトを情報源として用いた消費行動と要因の関係を与える手法を提案する．提案手法は，経験記事中の要因箇所を推定する手法を主な特徴とする．

5.1 問 題 設 定

レビューサイトを情報源として利用することで，消費行動クラス $\mathbf{A} = \{a_i\}$，消費行動分類関数 $\Psi_{da} : \mathbf{D} \times \mathbf{A} \longrightarrow \mathbf{R}$ は既知となる (3.2.1 節)．問題設定は次のようになり，本稿では特に問題 1 を解く手法について提案する．

問題 1）経験記事中の要因表現が記述されている箇所の推定

問題 2）経験記事集合 $\mathbf{D} = \{d_i\}$ があるとき，要因クラス集合 $\mathbf{F} = \{f_i\}$ の定義および，要因分類関数 $\Psi_{df} : \mathbf{D} \times \mathbf{F} \longrightarrow \mathbf{R}$ の作成

⋆1 http://twitter.com/

問題 3)　消費行動要因関係関数 $\Psi_{fa} : \mathbf{F} \times \mathbf{A} \longrightarrow \mathbf{R}$ の作成

5.2　要因箇所推定のための着眼点：経験記事の構成

要因箇所推定のために経験記事の構成に着目する．事前に実際の経験記事を観察したところ，人々は消費行動を行う前の場面についてまず記述し，その後消費行動中の経験や，消費行動後の評価を記述するケースが多く見受けられた．さらに消費行動前後の表現 “行きました” などの行動表現により区切られたり，消費行動前後の記述では頻出単語が異なる傾向が見られた．また要因は 2 章の定義より，消費行動前の事象や状態に限定されるため，消費行動前についての記述に多く登場すると考える．本稿では，上記観察結果と要因の性質に基づき，要因箇所推定のために，経験記事の構成を解釈する手がかりを用いることが有効と考える．詳細についてレストラン訪問についての経験記事を例に説明する．

5.2.1　記事中の出現位置

人々が自らの経験を記す際に，経験記事の構成と記事中の出現位置について関係があるとの仮定に基づき，要因箇所推定のために記事中の出現位置を用いる．

5.2.2　消費行動表現

消費行動を行ったことを表す行動表現を，消費行動前の要因箇所を推定するための手がかり表現として用いる．行動表現の前後には，行動を選択する前の情報が記述されやすいと考えるからである．下記の例 1 では，“行きました” という行動表現とそれに係る “ため” という接続標識が手がかり表現である．例 2 では，接続表現は存在しないが “利用しました。” という行動表現に係る表現が要因表現となっている．特に例 2 のような接続標識を伴わない行動表現も要因箇所推定のための手がかりとして利用できると考える．

例 1)クーポンが気になったため，行きました。
例 2)彼女とのデートで利用しました。

5.2.3　該当箇所が要因でないことを表す，負の手がかり表現

要因でない箇所を推定するために，負の手がかり表現を用いる．負の手がかり表現とは，消費行動決断後の記述に特に頻出すると考えられる表現である．例 1 の “美味しかったです。” という表現は食後 (消費行動後) の評価にあたる表現であり, 例 2 の “店内” という表現は，入店決断 (消費行動) 後の言及について記しているため負の手がかり表現である．

例 1)blog を読んで訪問．美味しかったです。次は友達と行きたい．
例 2)店内に入ると，想像以上に贅沢な空間でした．

5.3　提 案 手 法

提案手法は 5.3.1 節から 5.3.3 節に述べる 3 ステップから成る．5.2 節で述べた推定のための手がかりは，5.3.1 節に記す分類器の素性に反映される．提案手法は，**経験記事の構成を考慮した素性を用いることで，特定の表層的パターンに依存しない要因箇所の評価が可能である** (技術課題 1 の解決)．さらに，要因箇所を判定する処理と，分類する処理を分けることで，要因でない表現をできる限り排除した分類が可能になる (技術課題 2 の解決)．

5.3.1　要因文脈スコア付与処理

概要と目的:入力である，各記事 d の各単語 w_k に対して，要因を表す文脈で用いられている度合いを表す要因文脈スコアを与える処理である．例として，次のような経験記事があるとき次の記事中においては，“ガイドブック” という単語は高い要因文脈スコアが付与されることが望ましく，“安い”，“ボリューム” という単語は低いスコアが望ましい. 記事の筆者がお店を訪問した要因はガイドブックであると見て取れるからである[⋆1]．このように提案手法は経験記事毎に文脈を分析することで各語の各記事中における要因らしさを評価する．
例:ガイドブックを見て○○に行きました. 思ったより安くてボリュームがありました.

付与方法: 要因性スコアの付与には機械学習ベースのアプローチを用いる．まずトレーニングデータとして，記事集合 $\mathbf{D}$ の部分集合 $\{d^t\}$ を用意する．各 d^t の各単語 w_k に対して，人手で正解ラベル $\{FACTOR, NOTFACTOR\}$ を付与する．

上記トレーニングデータを利用し，未分類のデータを分類するための分類器を作成する．すなわち，未分類データである各記事の各単語を，ラベル $\{FACTOR, NOTFACTOR\}$ で表現される 2 クラスに分類する分類器の作成のおこなう．なお，分類器は尤度を出力可能なものとし，クラス $FACTOR$ への尤度を各単語の要因文脈スコアとする．分類器の素性には，着眼点をベースに表 2 に示す特徴を用いる．表 2 の ID1〜3 の素性が 5.2.2 節の特徴に対応し，ID7〜12 の素性が 5.2.3 節の特徴，ID13 の素性が 5.2.1 節の特徴に対応する．ID4〜6 の素性は因果関係を表す接続標識に基づく素性である．

5.3.2　要因クラススコア付与処理

要因クラススコア付与は，各単語への要因文脈スコア付与済みのドキュメント集合を $\mathbf{D}^+ = \{d_i^+\}$ にするとき，各 d_i^+ に対して，各要因クラスへの帰属度を与える処理であり，要因分類関数 $\Psi_{df+} : \mathbf{D}^+ \times \mathbf{F} \longrightarrow \mathbf{R}$ の動作に対応する．関数 $\Psi_{df+}(d_i^+, f_i)$ は，記事 d_i^+ の要因クラス f_i への帰属度を返す関数であり，今回は次の式により実装する．

$$\Psi_{df+}(d_i^+, f_i) = \sum_{w_k \in d_i^+} (fclass(w_k, f_i) * fscore(w_k)) \quad (1)$$

⋆1 別の記事においては，‘‘安い’’，‘‘ボリューム’’ という単語に高いスコアが付くことは充分あり得る

表 2 分類器の素性

ID	素性名	式	利用する特徴	詳細
1	depend_action	$\Sigma depend\_close(w_k, ActionEx)$		係り受け距離
2	action	$\Sigma close(w_k, ActionEx)$	行動表現からの近さ	前方単語距離
3	action_lat	$\Sigma close\_lat(w_k, ActionEx)$		後方単語距離
4	depend_factor	$\Sigma depend\_close(w_k, FactorEx)$		係り受け距離
5	factor	$\Sigma close(w_k, FactorEx)$	要因表現からの近さ	前方単語距離
6	factor_lat	$\Sigma close\_lat(w_k, FactorEx)$		後方単語距離
7	depend_not	$\Sigma depend\_close(w_k, NotFactorEx1)$		係り受け距離
8	not	$\Sigma close(w_k, NotFactorEx1)$	非要因表現 1 からの近さ	前方単語距離
9	not_lat	$\Sigma close\_lat(w_k, NotFactorEx1)$		後方単語距離
10	depend_not2	$\Sigma depend\_close(w_k, NotFactorEx2)$		係り受け距離
11	not2	$\Sigma close(w_k, NotFactorEx2)$	非要因表現 2 からの近さ	前方単語距離
12	not2_lat	$\Sigma close\_lat(w_k, NotFactorEx2)$		後方単語距離
13	position	$position(w_k)$	ドキュメント中の位置	

素性について “レストラン訪問” についてのレビュー記事を利用する場合を想定して記す．行動表現は “行く”，“入店” などの対象のジャンルに応じた消費行動表現である．要因表現は行動表現に係る文節で用いられる “ので”，“ため” などの接続標識である．非要因表現は消費行動中，消費行動後の描写に特に頻出する表現を意図する．非要因表現 1 は “美味しかった．” などの文末から n 語以内で用いられる形容詞，非要因表現 2 は “店内”，“メニュー”，“注文” などの特徴語を想定する．$close(w_k, w_i)$ は w_k と w_i のドキュメント上の出現位置の差の逆数である．(Fw_i より w_k が前方に限る) $close(w_k, w_i)$ は w_k と w_i のドキュメント上の出現位置の差の逆数である．(ただし w_k より w_k が前方に限る) $depend\_close(w_k, w_i)$ は w_k と w_i のドキュメント上のそれぞれの単語が含まれる出現文節の出現位置差の逆数である (なお両方の文節が係り受け関係にないときは 0 とする)．

関数 $fscore(w_k)$ は文書 d_i における単語 w_k の要因文脈スコアである．関数 $fclass : \{\langle w_k, f_i\rangle\} \longrightarrow \mathbf{R}$ は単語 w_k が要因クラス f_i の特徴語である度合いを示す値とする．$\mathbf{F}$ および $fclass$ の作成は要因クラスを定義づける処理である．ただし本稿においては提案の範囲外とし，実験においては，予め複数の観点から作成した要因辞書を利用する手法などを用いる (6.2.1 節)．

5.3.3 消費行動要因関係算出処理

5.3.2 節までで $\Psi_{df} : \mathbf{D} \times \mathbf{F} \longrightarrow \mathbf{R}$ が作成済みである．また $\Psi_{da} : \mathbf{D} \times \mathbf{A} \longrightarrow \{0, 1\}$ は既知である．それらを組み合わせ $\Psi_{fa} : \mathbf{F} \times \mathbf{A} \longrightarrow \mathbf{R}$ を作成する．Ψ_{fa} には，Ψ_{fa1} と Ψ_{fa2} の 2 つの異なる指標を用いる．Ψ_{fa1} は消費行動 a_i に頻出する要因クラスを高く評価する関数，Ψ_{fa2} は消費行動に特に顕著な要因を抜き出すことを意図し全体の傾向と比較して特に消費行動 a_i に頻出する要因クラスを高く評価する関数であり，次式で表現する．

$$\Psi_{fa1}(f_i, a_i) = \sum_{d_i \in \mathbf{D}} (\Psi_{df}(d_i, f_i) * \Psi_{da}(d_i, a_i)) \quad (2)$$

$$\Psi_{fa2}(f_i, a_i) = chi2(\Psi_{fa1}(f_i, a_i), |a_i|, \sum_{a_j \in \mathbf{A}} \Psi_{fa1}(f_i, a_j), \sum_{a_j \in \mathbf{A}} |a_j|) \quad (3)$$

$chi2(a, b, c, d)$ は a,b,c,d を 2×2 分割表の要素とするときのカイ二乗値を返す関数である ($(a/b) < (c/d)$ の時は 0)．$|a_i|$ は消費行動クラス a_i について書かれた記事数である．

6. 実　　験

提案手法の有効性検証のために実データを用いた実験を実施した．実験は要因文脈スコア付与についての評価実験と要因と消費行動の関係抽出に対する定性分析の 2 つから成る．データセットは Web に公開されているレストランについてのレビュー記事を収集し用いた．

6.1 要因文脈スコア付与の妥当性を測る評価実験

6.1.1 実 験 方 法

予めランダムに選んだ記事 200 件の各語 (名詞，動詞，形容詞，副詞) に対してラベル $FACTOR$ もしくは $NOTFACTOR$ を研究者 1 名が人手で付与した．付与基準は，2 章の定義に従い，消費行動前の消費行動に至る原因となる事象や状態を直接的に単語およびそれを補足する関連語に $FACTOR$ ラベル，それ以外に $NOTFACTOR$ ラベルを付与した．

評価実験では，それらの一部をトレーニングデータとして分類器を学習させ，残りのテストデータを用い分類性能を評価する．学習のための素性は前章で提案したものを用いる．素性で用いる具体的な表現集合は，研究者 1 名の記事集合の観察により表 3 の通り選出した．

表 3 素性に用いる具体的な表現集合

表現集合名	具体的な表現
ActionEx	“行く”，“来る”，“寄る”，“来店”，“訪問”， “尋ねる”，“伺う”，“見つける”，“訪れる”，“入る”，“行う”
FactorEx	“ので”，“ため” のうち ActionEx に係るもの
NotFactorEx1	“文末（“。”）から 4 語以内にある形容詞
NotFactorEx2	“注文”，“店内”，“店員”，“雰囲気”，“メニュー”

記事中の単語数は $FACTOR$:705 単語，$NOTFACTOR$:5185 単語であった．学習手法には AdaboostM1[17] を用い．弱学習器には決定株，Iteration は 10，Seed は 1，Weight Threshold は 100 とした．なお分類器は分類対象データに，0 以上 1 以下の実数を尤度として与える．評価方法には 10 点交差法を用いた．提案手法と比較手法を下記に示す．

- OurMethod:全ての素性を利用
- Baseline:素性として FactorEx(因果関係を表す接続標識) を用いたもののみ利用

6.1.2 実 験 結 果

表 4 に，実験によって求められた各手法の F 値を示す．尤度閾値 0.1, 0.25, 0.4, 0.6 においては OurMethod が高く，尤度閾値 0.75 では Baseline の方が高い結果となった．また尤度閾値 0.9 では両方とも 0 となっている．最も高い F 値は OurMethod が尤度閾値 0.1 で記録した 0.409 である．また分類器 OurMethod の詳細について，表 5 に示す．分類器 OurMethod は position, depend_action, not2_lat, action_lat の 4 つの素性を分類に用いた．

表 4 分類性能 (F 値)

尤度閾値	0.1	0.25	0.4	0.5	0.6	0.75	0.9
OurMethod	0.409	0.408	0.407	0.365	0.171	0.00	0.00
Baseline	0.220	0.119	0.119	0.119	0.119	0.119	0.00

尤度閾値 x は，分類器の出力尤度が x より高いときに，分類器の予測を $FACTOR$ であるとみなす値である．

表 5 作成された分類器

名称	重み	素性	閾値	low	high
cf1	1.99	position	0.26	N	N
cf2	0.99	position	0.31	F	N
cf3	0.46	depend_action	0.06	N	F
cf4	0.62	not2_lat	0.01	F	N
cf5	0.21	action_lat	0.01	F	N
cf6	0.29	not2_lat	0.10	N	N
cf7	0.09	not2_lat	0.10	N	F
cf8	0.08	not2_lat	0.10	N	N
cf9	0.08	not2_lat	0.10	N	F
cf10	0.08	not2_lat	0.10	N	N

OurMethod は cf1〜cf10 の決定株をブースティングにより結合した分類器である．表の見方の例として，決定株 cf2 は，入力単語について素性 position が 0.31 より低い (low) ときに，$FACTOR$ と判定し，そうでないとき (high) に $NOTFACTOR$ と判定していることを表す．またブースティングにおける cf2 の重みは 0.99 である．

6.1.3 考　　察

実験において，OurMethod は最も高い F 値を記録した．また OurMethod は，記事中の位置を評価する position, 行動表現との距離を評価する dependaction, actionlat，非要因表現との距離を評価する not2_lat をそれぞれ分類に用いる素性として選択した．提案手法において着眼した点に基づく素性がそれぞれ用いられている．

また提案手法の問題点を探るために，正例を誤予測した例を観察し分析を行った．(観察対象は人手で与えたラベルは $FACTOR$ であるが，機械が判定した $FACTOR$ である尤度が 0.1 以下であった予測結果である.) 観察から次に示すような誤判定につながる典型的なケースがみられた．

(1) 要因について言及している長さが平均的な場合と異なるケース
今回は記事中の出現位置を，要因箇所を判定する素性として用いある程度有効に機能した．一方で，記事によって要因について言及する長さにはばらつきがあり，記事の後方まで要因について言及している場合などは尤度を高めることができなかった．

(2) “立ち寄る”，“いく”，“行列に並ぶ” など未設定の後が行動表現として使われたケース

(1) は周囲の数多くの出現単語傾向から要因箇所である正否の判定を行う手法，(2) は手がかりとなる表現の自動獲得手法などが，性能向上に有効であると考え，今後の課題である．

6.2 要因と消費行動の関係抽出

6.2.1 実 験 方 法

提案手法の有用性分析のために，提案手法を約 100 万件のレビュー記事に適用し，要因と消費行動の関係を抽出した (消費行動要因関係関数 $\Psi_{fa} : \mathbf{F} \times \mathbf{A} \longrightarrow \mathbf{R}$ を適用した)．本章ではその性質と問題点について論じる．適用における必要要素は次のように用意した．

1) 要因文脈スコア付与処理に用いる分類器

実験 1 で作成した OurMethod を利用する．

2) 要因クラス F の作成

今回 **F** の作成はユニークな 1 単語を 1 クラスとする単純な手法により実験を行う．**F** に用いる単語の選出については，2 つの方法を別に行った．方法 1 は予め人手で要因に適する語を選出する方法である．表 6 に示す語が実際に選出した語の集合である．選出は，予め記事集合における，行動表現に係る接続標識の前に頻出する語を観察した上で，1 語で要因に対応する語の選出を意図し，人手で行った．方法 2 は品詞が名詞，形容詞，副詞である出現回数 100 回 (全記事において) 以上の全単語を用いる方法である．以下では，方法 1 により作成された **F** を $\mathbf{F_1}$，方法 2 により作成された **F** を $\mathbf{F_2}$ と表記する．

3) 消費行動クラス A の作成

各記事 d_i に付与されている属性情報のうち，店舗のチェーン店名を用いる.**A** はチェーン店名の集合である．例として，ハンバーガーチェーン A 横須賀中央店についての記事の，消

費行動クラスは，“ハンバーガーチェーン A へ行く” とすることができる．いずれの場合も 5.3.2 節に述べた $fclass(w_k, f_i)$ の値は，$f_i \in \mathbf{F}$ であるとき 1，そうでないときは 0 とする．

表 6 $\mathbf{F_1}$ に用いる単語

種類	具体的な単語
シーン	出張, 帰り, 会社, 休日, デート, 途中, 休憩, 旅行, 用事, 職場, 家, 仕事場, 休み, 勤め先, 昼, 夜, お昼, 朝, モーニング, 通勤
宣伝	クーポン, チラシ, キャンペーン, セール, 割引, 配布
評判	テレビ, 番組, 雑誌, 特集, 芸能人, 口コミ, 評判, 話題, 人気, 推薦, 勧め, 紹介, 噂, 有名, ガイドブック, 近所, 雑誌, サイト, 評価, ランキング, 広告, 絶賛
距離	近い, 遠い, 近所
感性	美味しい, 新しい, がっつり, 軽く
その他	天気, 行列, 混む, オープン, 晴れ, 曇り
人	友達, 友人, 先輩, 社長, 彼女, 妻, 親, 両親, 子供, 先生, 連れる, 誘う, 招く, 招待する, 共に, 薦める, 勧める

6.2.2 実験結果

提案手法を実データに適用した結果例を表 7 と表 8 に示す (なお，実験結果では具体的なチェーン店名は伏せて示す)．表 7 では, 要因クラス集合を $\mathbf{F_1}$，$\mathbf{F_2}$ とした場合それぞれについて，また消費行動要因関数を Ψ_{fa1}, Ψ_{fa2} のそれぞれの場合について，ai=“ハンバーガーチェーン店 A に行く” という消費行動上位の要因を示す．表 8 も同様である．

表 7 ai=“ハンバーガーチェーン A に行く” の要因クラス上位 (939 記事)

$\mathbf{F_1}$				$\mathbf{F_2}$			
Ψ_{fa1}		Ψ_{fa2}		Ψ_{fa1}		Ψ_{fa2}	
朝	50.5	朝	504.1	マック	164.6	マック	45183.4
混む	34.8	クーポン	416.4	駅	83.7	スルー	3264.3
クーポン	24.4	混む	67.6	利用	63.2	月見	1078.6
子供	20.9	キャンペーン	57.5	時間	53.7	ドライブ	928.8
安い	17.1	休憩	46.0	ない	50.8	メガ	530.2
お昼	14.0	子供	41.6	朝	50.5	朝	504.1
夜	14.0	美味しい	25.0	多い	49.5	フード	472.1
美味しい	13.3	有名	13.1	店内	49.4	クーポン	416.4
近い	12.4	休日	9.5	階	48.1	ポテト	373.9
帰り	12.2	モーニング	7.2	人	43.9	コート	322.4
昼	11.5	友人	5.1	コーヒー	42.1	シェイク	290.9

表 8 ai=“ラーメンチェーン B に行く” の要因クラス上位 (1248 記事)

$\mathbf{F_1}$				$\mathbf{F_2}$			
Ψ_{fa1}		Ψ_{fa2}		Ψ_{fa1}		Ψ_{fa2}	
行列	69.3	行列	270.6	ラーメン	366.2	赤丸	45945.4
有名	62.6	有名	61.8	味	189.7	白丸	38179.3
美味しい	62.0	社長	11.8	赤丸	132.1	新味	9302.3
夜	25.0	安い	9.4	麺	123.4	ラーメン	1386.3
人気	21.2	友人	6.1	白丸	103.9	とんこつ	1272.2
高い	17.8	紹介	5.5	とんこつ	81.8	替え玉	1029.3
お昼	17.0	口コミ	4.2	好き	79.1	大名	845.7
混む	16.5	リーズナブル	3.7	ない	74.8	骨	659.2
帰り	13.9	朝	2.9	人	74.6	高菜	641.54
オープン	12.81	人気	2.8	スープ	74.5	もやし	576.9
昼	12.31	評判	2.7	いつも	71.8	こってり	451.9

表 9 特徴的な要因

モーニング	$\longrightarrow$	喫茶店チェーン C	5714.87
行列	$\longrightarrow$	ドーナツチェーン D	1101.28
安い	$\longrightarrow$	ファミリーレストラン E	909.77
休憩	$\longrightarrow$	喫茶店チェーン F	846.62
キャンペーン	$\longrightarrow$	アイスクリームチェーン G	815.09

6.2.3 考　　察

表 7 の要因クラス集合 $\mathbf{F_1}$ の，Ψ_{fa1} において上位に登場する “安い”，“夜”, “昼” などの単語は Ψ_{fa2} において登場しない．これらの単語は他の消費行動においても上位となるため，消費行動 a_i において特に顕著な単語ではないため順位が下がっている．このように Ψ_{fa1} は，単に消費行動 a_i に多く出現する要因を評価し，Ψ_{fa2} は他の消費行動と比較して顕著要因と，異なる観点から評価しており，消費行動の特徴を理解する上で重要と考える．

表 7 の “クーポン” や “朝” という単語はそれぞれクーポンを利用したことや，朝に来店したことが想像可能といえる．一方で表 8 の “行列” という単語は，行列を見て入店したのか，他の店の行列を敬遠して入店したのか判断がつかず，単語とクラスとが一意に対応していない．この問題は $\mathbf{F_2}$ においてより顕著となり，“階”，“時間”，“人” など何の要因を表しているのか判断がつかない単語が数多く登場している．また別の問題として，“お昼” と “昼” など 1 クラスにまとめられるべき単語の集合もある．このように，提案手法では要因の最小単位とは言い難い語が検出される場合があり，さらに予め要因クラスの単語をフィルタリングしない ($\mathbf{F_2}$) と，その傾向が高まることが本実験により分かった．一方で，フィルタリングをしないことで，チェーン店に固有な表現も抽出されることが分かった．今後は，

要因の最小単位を的確に捉え，それら幅広く抽出するための **F** の設計が主な課題である．

表 9 は全ての **A**, $\mathbf{F_1}$ の要素の組み合わせから Ψ_{fa2} の値が高い，f_i と a_i のペアが上位から並べたものである．このように他店に対して際立って重要な要因を抽出することができる．

7. 関連研究

本研究で着目した問題設定・手法は情報抽出における因果関係分析に関連する．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因果関係知識を自動的に抽出することを目的としている．因果関係分析においては，(1) 特定の修辞的要件を満たす原因表現と結果表現のペアの抽出や，そのための素性となる言語パターンの抽出手法が提案されている[8)9)10)11)12)13)14)15)16)]．また (2) 記事の主題特定器の作成と，記事からは原因表現のみを抽出する手法[16)]，(3) 個々の因果関係をつなぎ合わせることで因果関係ネットワークを構築する手法[10)13)] が提案されている．(4) 因果関係を，医療分野に特化した分類を行う手法[14)] や，“事態” と “行為” に基づいた分類を行う手法[9)]，重要な原因と結果の発見を目指す手法[13)] なども提案されている．これらに対し，本研究では人々の自らの経験についての因果関係を獲得することを目的としている．経験記事の口語的な表現に対応し適切な分類を行うために要因箇所を推定するために経験記事の構成を考慮した手がかりを利用すること，消費行動を特定するためレビューサイトにおける経験記事の属性情報を利用することを特徴とする点で異なっている．

また本研究は評判分析[18)19)] に関連する．評判分析は，人々が消費行動を体験する後の意見や評価を主なターゲットとしているのに対して，本研究では人々が消費行動を体験する前の事象や状態から表現される要因を求めることを目的としている．

また倉島らの経験マイニングについての研究[20)] に本研究は関連する．経験を分類するために，語の共起と統計指標を用いている点を本研究では参考にしている．

8. おわりに

本稿ではレビュー記事群を情報源として用いた，要因という観点から人々の消費行動を整理する枠組みとそのために必要な手法を提案した．提案手法は，要因が記事中に書かれている箇所を推定するために経験記事の構成を考慮した手がかりを利用することを主な特徴とし，評価実験により従来手法より高精度に要因箇所を推定できることが分かった．また実データへ提案手法を適用した結果，要因クラス集合の適切な設計が今後の主な課題と分かった．今後，効果的なマーケティングリサーチ手法を確立できるよう課題を解決していきたい．

参考文献

1) Kotler.P et al: Marketing Management(12th Edition), Prentice Hall (2006)
2) http://ja.wikipedia.org/wiki/マーケティングリサーチ
3) 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)
4) Ries. A et al.: ポジショニング戦略, 海と月社 (2008)
5) 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)
6) 藤井大輔.: 「R25」のつくりかた, 日経プレミアムシリーズ (2009)
7) 三省堂.:大辞林第二版 (1995)
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, Vol.4, Issue.4, Article.8 (2007)
9) 乾孝司 et al.: 接続標識「ため」に基づく文書集合からの因果関係知識の自動獲得 情報処理学会論文誌, Vol.45, No.3, pp.919-933 (2004)
10) 佐藤岳文 et al.: Web マイニングを用いた因果ネットワークの自動構築手法の開発, 社会技術研究論文集 Vol.4, pp.66-74 (2006)
11) Torisawa, K.: Acquiring Inference Rules with Temporal Constraints by Using Japanese Coordinated Sentences and Noun-Verb Co-occurrences (2006)
12) 鳥澤健太郎.: 「常識的」推論規則のコーパスからの自動抽出, 言語処理学会第 9 年次大会, pp.318-321, (2003)
13) 青野壮志 et al.: 要因検索による因果関係ネットワークの構築と因果知識の獲得, DEIM Forum (2010)
14) Khoo, C.S.G.: Extracting Causal Knowledge from a Medical Database Using Graphical Patterns, In Proc. 38th Annual Meeting of the ACL, pp.336-343 (2000)
15) 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)
16) 坂井浩之 et al.:交通事故事例に含まれる事故原因表現の新聞記事からの抽出, 自然言語処理, Vol.13, No.4, pp.99-124 (2006)
17) Freund, Y. et al.:A decision-theoretic generalization of on-line learning and an application to boosting, Journal of Computer and System Sciences, Vol.55, pp.119-139 (1997)
18) 鍛治伸裕 .: テキストからの評判分析と機械学習, SIG-FPAI(2009).
19) 乾孝司 et al.:テキストを対象とした評価情報の分析に関する研究動向, 自然言語処理, Vol.13, No.3, pp.201-241(2006)
20) 倉島健 et al.: 大規模テキストからの経験マイニング, 電子情報通信学会論文誌 Vol.J92-D, No.3, pp.301-310(2008)