

価値観モデルを利用した マイクロブログ発言者の社会的類型の推定

谷田泰郎 馬場彩子 河本裕輔 藤井絵美子
シナジーマーケティング株式会社

{ tanida.yasuo , baba.ayako, kawamoto.yusuke, fujii.emiko }@synergy101.jp

1 はじめに

インターネットやスマートフォンに象徴される情報検索の進化は、消費者が自らスピーディーにかつ能動的にサービスを選好することを可能にした。また、デフレーション状態が長引く日本の経済環境は、消費者の心理状態に大きな影を落としながらも、サービス供給者が発信する価格情報やブランド価値のみを要素とする従来の選好リストを破壊し、マイクロブログなどを通じて消費者同士が情報を送受信することで、サービス需要のための新たな選好リストを構築しつつある。そのようなサービス需給構造の変化の過程で、マスメディアを中心にした従来型のアウトバウンドマーケティング施策に加えて、インターネット上での消費者の能動的な行動を仮定したインバウンドマーケティング施策を併用して実施する企業が増えている。企業が施策の最適化を行うためには、性別、年齢、職業などのデモグラフィック属性だけではなく、消費者心理を理解し、その行動を予測することが必要となっている。

そもそも、多くの企業がメール会員やEC会員など既存の顧客だけではなく未来の顧客候補の情報を保有しており、同時に会員ID、購買履歴と紐づいたメールやメールクリック情報、商品レビュー、ツイート情報、サービス内容説明、回遊履歴など様々な情報を持っており、その情報の多くが自然言語であることから情報抽出のための自然言語処理は欠かせない技術となっている。我々は、抽出した情報を人ごとにトピック化し、サービスに対するベネフィットや購買行動、情報源との関係を明らかにし、これをデモグラフィック属性やサイコグラフィック属性（本稿では価値観成分と呼ぶ）と紐づけることでより精緻に人を理解することを目指している。このようなモデル化を行うことで、マーケティング施策に利用しやすいセグメントの構築や消費者の期待効用の測定に活用できるだけでなく、共通のセグメントを利用してサービス提供者の枠組みを超えた分析を行うことで集合知を構築して社会に還元することが可能となる。

一方、ソシオ・センサーとしての位置を確立しつつあるマイクロブログにも発言者の価値観が潜在している。

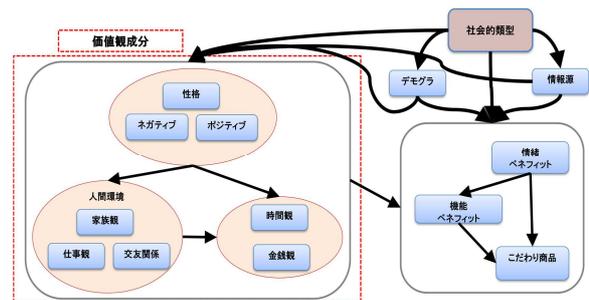
例えば、荒牧は発言者の使用語彙数を推定している[1]が、語彙の種類や表現方法、すなわち言語能力は発言者の価値観そのものだと考える。

このような背景の下で、心理的な属性である価値観のみに着目し、価値観成分の保有状態を社会的類型としてモデル化する。さらに、マイクロブログ発言者のテキスト情報から価値観成分と関係性の強い言語的成分を抽出し、社会的類型との関係をモデル化して評価することでマイクロブログ発言者の社会的類型の推定可能性を確かめる。

2 実験の全体像

2012年に消費者の行動に関する価値観調査を始め、デモグラ、価値観、その他消費行動に関する1,000問程度の質問項目からなる定量調査（Web調査）を行い、約1.1万人分のサンプルを取得した。このような社会調査情報とサービス提供企業が持つ消費者の行動履歴や購買履歴との関係のモデル化（図1）を進めている。

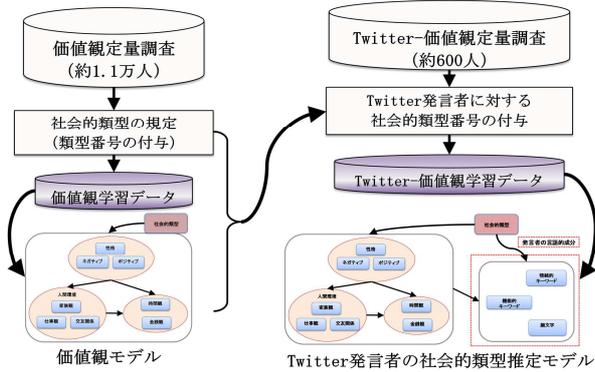
図1：消費者行動予測モデルの仮説例



本実験（図2）では、価値観定量調査の質問項目のうち、価値観に関する質問項目から得られた価値観成分のみを利用して、その類型化を行うことで社会的類型を規定し、価値観モデルを構築した。また、ツイート内容の提供が可能な被験者に対して、価値観に関する質問項目をさらに絞り込んだ小セットの質問項目を用意して定量調査を行い、600人分のサンプルを得た。そのサンプルに対して、価値観成分の計算を行い、価値観モデルによって得られた社会的類型番号の解と紐づけ、ツイート内容から得られた言語的成分と組み合わせで教師デー

タを作成した後、Twitter発言者の社会的タイプの推定モデルを構築し評価した。

図2：実験の全体像



3 価値観による社会的類型

ヒトの持つ潜在的な情報の伝達単位は、ヒトからヒトに複製されながら変化を続けている。このような情報の伝達単位はコミュニケーションの伝達単位であり、コミュニケーションを可能にする複雑な脳を持つヒトのような生物に限定され、動物学的な遺伝子と区別されるコミュニケーションにおける自己複製子は、meme(ミーム)と呼ばれる概念[2][3]に近いものである。然しながら、我々の目的は、脳の中に定住しているミームを確認することではなく、コミュニケーションの伝達単位を脳の外側で観察できる表現型効果、外界での帰結としてとらえ、マーケティングコミュニケーションのツールとしての精度を向上させることである。本研究が目指す範囲は、消費行動に限定したものであり、その視点から価値観を情報の伝達単位と考え、以下に記述する定性調査と定量調査を実施した。

3.1 価値観調査

消費行動に限定した価値観の予備調査として、性別・年代のセルのみを考慮して、デプスインタビュー、グループインタビューによる定性調査を実施した。調査の背景には「Activities, Interests, Opinions(AIO)」(どのようなことに時間を使っているか、どのようなことに興味・関心を持っているか、政治、社会問題など、さまざまな出来事をどう感じているか、という3つの側面(+デモグラフィック属性)について質問することで、生活全般に関するライフスタイル、あるいは、特定の生活領域や製品カテゴリに関するライフスタイルを測定しようとするアプローチ)があった[4]。AIOの採択理由は、「行動」にヒトの本質が現れるという観察から得られる証拠を信じるという我々の考え方に近く、また、行動を聴取することで、潜在的な関心や価値観を抽出し、情報の伝達単位を成分としてモデル化するという我々独自の一連のプロセス(行動履歴データと定量調査デー

タを混合して行動予測モデルを作成する)との相性も良かったからである。

定性調査で得られた仮説に基づいて、定量調査を予備調査と本調査に分けて実施した(対象は、学生を除く20歳以上、男女比は1:1、他の条件は自然発生)。聴取した項目は、性格が関与しそうな行動・ポジティブ・ネガティブに関する価値観、家族・人間関係・仕事に関する価値観、お金・時間に関する価値観の他、お気に入りの商品や購入経路や場所、情報源、消費の際の機能的あるいは情緒的ベネフィット、SNS関連、デモグラなどである。予備調査では、定性調査から得られた知見を基に膨らませた質問項目をできるだけ削除せずに聴取し(約1,000人強の被験者)、予備調査で得られた回答の分析結果(回答の反応数、相関分析・主成分分析の結果、その他定性的な指標)により本調査の質問項目を絞り込んだ。予備調査と本調査は、それぞれ別の調査会社(別のパネル)に約2カ月の期間を開けて依頼した。調査結果は、共通質問項目の回答の分布の相関(Pearsonの相関係数)が、価値観に関する質問項目(303項目)で0.99、こだわり商品に対する情緒ベネフィット(34項目の情緒的なイメージを徴収した)で0.95あり、特筆すべき信頼性があった。予備調査と本調査の価値観に関する共通質問項目のみを混合して主成分分析を行った結果得られた価値観成分は63個あった(表1)。

表1：価値観成分の抽出結果

価値観フレーム	成分数	成分の内容(ネーミング)
基本的な性格	11	好奇心旺盛 デリケート マイペース 協調型 勤勉 上昇志向 短気 正義感 ルーズ・不精 無気力 文系的
ポジティブ価値観	8	自己愛 自己成長 アウトドア スポーツ 恋愛 趣味 ギャンブル ひとり時間
ネガティブ価値観	3	否定・批判 非常識 期待はずれ
友人関係	8	ストレス 親友中心 ネットワーク重視 社会的 大人数派 消極的(独身) 仕事人脈中心 ノンストレス
家族関係	7	結婚願望 不仲 責任感(主婦軸) 責任感(扶養軸) 良好(別居家族) 不十分 良好(同居家族)
仕事に対する価値観	7	満足 ストレス プライベート重視 キャリアアップ 転職願望 堅実 社会的意義
お金に対する価値観	8	ギリギリゆとり 貯蓄志向 家族優先 慎重派 自己投資 堅実生活 常識的
時間に対する価値観	11	ゆとり 余裕がない 充実 仲間優先 家族優先 趣味優先 インドア派 アウトドア派 家事分担 退屈 自己投資

3.2 Societas (ソシエタス)

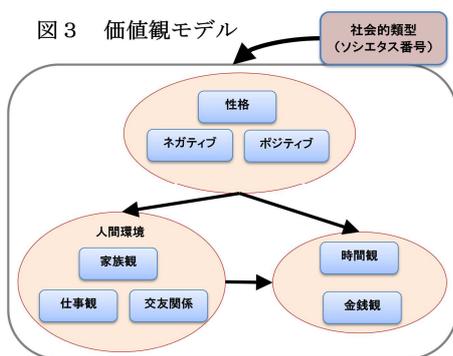
被験者ごとの価値観の成分得点の分布を基にクラスタリングを行うことで社会的タイプの規定を試みた。価値観成分から社会的タイプを規定する理由は、マーケティング施策を企画する際のターゲティングやセグメンテーションにデモグラフィック属性だけでは叶えられないインスピレーションを与えること、各種マーケットのシミュレーションを行う際の複製元、すなわち消費者のエージェントの雛型を作成すること、企業の枠組みを超えた共通の物差しを作成すること、という目的があったからである。本稿では、このような社会的タイプをSocietas

(ソシエタス)と呼ぶことにする。Societas (ソシエタス)とは、「社会、共通の目的を持つ団体、同盟、連邦」の意のラテン語であるが「顧客をライフスタイル、価値観、感情、行動で類型化したもの」の意味で用いる。ソシエタスは、現時点では、定量調査で得られたデータ、すなわち、前述の価値観成分のクラスターリングによって得られるものである。ソシエタスを求めるためのクラスターリングに関しては、マーケターにとっての説明力を重視して、Ward法による結果を採用した。表2に規定した社会的類型である12個のソシエタスを示す。

表2：ソシエタス番号（価値観成分の社会的類型）

#1-1	受け身な隠者タイプ
#1-2	受け身な清閑タイプ
#2-1	家族大好き悠々タイプ
#2-2	家庭的な真面目タイプ
#3-1	アウトロータイプ
#3-2	こだわりインドア派タイプ
#4-1	自分中心的なアクティブタイプ
#4-2	好奇心旺盛なバランス人間タイプ
#5-1	家族想いの多忙ワーカータイプ
#5-2	社交的な堅実ホームメーカータイプ
#6-1	繊細な個人主義タイプ
#6-2	好奇心旺盛な人生謳歌タイプ

3.3 価値観モデル



被験者ごとに表2の価値観成分(表1)とソシエタス番号(表2)を与え、それを教師データとして学習することで

ベイジアンネットワークによる確率モデルを構築した(図3)。学習に利用したデータ件数は、予備調査及び本調査から得られた11,410件である。ソシエタス番号はハードクラスターリングによって求めており、11,410人の被験者すべてが12種類のソシエタスのどこかに100%属している(真性ソシエタス保有者と呼ぶことにする、実際は確率モデル化すると同時に真性は殆どいなくなる)と仮定した。価値観成分はすべて「あり/なし」の2値に離散化した。ソシエタス番号は12水準の変数である。構造探索は図3に示すような制約条件(矢印の先が親になることを許容する、円で囲まれた部分のカテゴリ内の変数同士はフリー探索、すなわち双方が親になる可能性がある)で、全体のAIC(赤池情報量)が最大になるようにモデルを構築した(双方向リンク、ループの回避に関しても情報量損失が最小になるよう考慮した)。

4 Twitter発言者と価値観

マイクロブログに関しては多くの先行研究があり、ブログ発言者の属性推定に関しても様々な先行研究がある[5]。また、マイクロブログ発言者に対して社会調査(Web調査)を併用した結果の分析報告もある[6]。本研究においても、発言者の価値観を推定するためには、定量調査を併用して価値観と発言内容を紐づける必要があると考え、その実現可能性を確認するため、定量調査を実施することで基礎データを取得した。また、発言の取得のし易さ、発言の即時性や匿名性により、消費行動におけるヒトの本音や性質が現れやすいことなどの理由から、分析するマイクロブログの対象としてTwitterを採用した。

4.1 Twitter発言者に対する価値観調査

TwitterIDを公開できること、過去の1,000ツイート以上の書き込みを取得できること、データ分析利用を許容することを条件に、TwitterIDの取得と合わせて、価値観に関する定量調査(Web調査)を実施した。通常の価値観調査の質問項目の約6分の1に当たる60問に絞り込んだ小セットの定量調査を行い、600人分を回収した。然しながら、TwitterIDが自己申告であったため、全くツイートを取得できなかった人も含め、100K未満のテキストしか取得できなかった被験者のデータは信頼性の観点から排除し、結果562人分のデータを採用した。価値観成分に関しては、表1の成分うち小セットの質問項目と関連性が強い22成分の得点を前述のモデル構築時と同じ成分得点係数により計算した。

4.2 発言テキストと価値観

562人の発言者に対して、発言テキストから発言者ごとの使用語彙を得た。得られた約49万語の異なり語彙のうち、使用発言者数が10%以上90%未満(57人~505人)であった語彙約13,000語強を価値観成分との関係の分析対象とした。その理由は、発言者の反応率が高過ぎるものは、被験者の特徴を表さないと考えられ、逆に低過ぎる語彙は、発言者の特徴を強く表すものも含まれると考えられるが、条件付き確率モデルを利用する場合には学習データが疎になる可能性が高いからである。

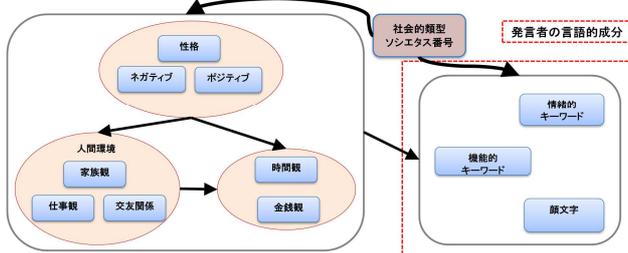
次に、分析対象とした語彙のうち、22個の各価値観成分に対して、その価値観成分を持っている発言者の語彙出現確率 $P(W_i=1|X_j=1)$ と価値観成分を持っていない発言者の語彙出現確率 $P(W_i=1|X_j=0)$ の差(絶対値)が大きい候補100語彙を抽出して精査した(Wは語彙、Xは価値観成分)。語彙の出現を条件として成分の確率分布差により候補を抽出すると言う考え方もあるが、価値観の存在が使用語彙を決めると言う仮説からこの方法を採用

した。各成分と関係性が強かった語彙には、例えば、性格の「協調型」では、「ありがと、うれしい」のような感情表現や「(^-^)」のような顔文字、「お昼、ごはん、茶、チーズ、お寿司、只今、お祝い、リクエスト」など人付き合いの良さを象徴する語彙が見られた。また、同じく性格の「マイペース」では、「www, Ω」のような感情を表す記号、「やばい、めんどい、萌え」などの特徴的な語彙が見られ、「デリケート」では同じ顔文字でも「><」のような少しネガティブなものや「きゃー、ぎゃー」などの感情を表す表現が見られた。一方、同じ性格でも「好奇心旺盛」では、感情表現や顔文字、記号が殆ど見られず「問う、歴史、米国、研究、継続、刺激、提案、警告、市場、維持」などのように抽象度の高い表現が多かった。最終的には、価値観成分ごとに100候補の語彙を30まで絞り、主成分分析により得られた第1成分、第2成分を発言者の言語的成分とした。

4.3 Twitter-価値観モデル

図3に示した価値観モデルの22成分版を全成分の価値観モデルと同様の方法で作成し、Twitter発言者の価値観成分を入力にしてソシエタス番号の推論を行った。ベイジアンネットワークにおける推論アルゴリズムは、Loopy-BPを採用した[7]。ソシエタス番号の推論結果の確率分布に応じてウェイトバックをかけ（例えば、ソシエタス番号#1-2である確率が、17%であれば、1の位を四捨五入して20%とし、そのデータを2件複製、100%であれば10件複製）教師データとした。作成したモデルの探索条件を図4に示す。その他のベイジアンネットワークの作成条件や探索条件は、価値観モデルの作成条件に準じた。発言者の言語的成分は「あり/なし」の2水準に離散化した。

図4 Twitter発言者の社会的類型推定モデル



5 評価結果

表3 : モデルの評価結果 (推論アルゴリズム:Loopy-BP)

	テスト方法	価値観成分数	証拠	1位正解率	3位以上正解率
価値観モデル	closed	63	価値観成分	54.2%	80.5%
	open	22		36.3%	67.8%
twitter-価値観モデル	closed		言語的成分	33.8%	65.5%
	open		20.1%	43.9%	
					32.6%
				19.1%	41.7%

評価結果 (表3) の見方は以下のとおりである。

- 正解率は、価値観成分もしくは言語的成分を証拠として与えたときの推論結果と教師データの正解ソシエタス番号の一致率であり、3位以上正解率は正解ソシエタス番号が3位までに含まれる割合。
- ランダム正解率は、1位で8.3%、3位以上で25%。
- openテストは10分割の交差確認テストにより実施。

また、評価結果は、以下のことを示唆している。

- 価値観モデルのclosedの結果は、ソシエタス番号の事後確率の揺れの大きさを示している。また、双方のモデル共に、openとclosedの精度差が小さく、交差確認テストの内容を確認すると、open実験でのモデル構造の揺れも精度の分散も小さかった。証拠部分のデータのスパース性やモデル構築方法ではなく、教師データのソシエタス番号、社会的類型のクラスタリング手法の問題。
- 3分の1の価値観成分で約7割の精度担保が可能、3割の精度低下を重視すれば他の成分が必要。
- Twitter-価値観モデルの上限は、価値観成分を与えたclosedテストの結果(32.6%)であり、価値観成分なしの言語的成分のみでの推論では、1位正解率で上限の約6割、3位以上正解率で約7割の精度が担保できる。少なくとも下限のランダム正解率(下限)よりも上。言語的成分による推論は有効。

6 おわりに

社会的類型のクラスタリング、価値観成分と言語的成分の紐づけ手法など、基礎的な方法論に関する課題も多いが、少なくとも、Twitterのようなマイクロブログの発言者のテキスト情報を証拠とした発言者の社会的類型の推定に関する有効性を確認できた。また、本稿では利用していないが、デモグラフィック属性、行動履歴、サービスに対するベネフィット、あるいは、フォロワー数やフォロー数、1ツイートあたりの語彙数や推定使用語彙数、その他プロフィールから取得できる情報など、他に利用可能な証拠も多い。それらの利用手法を確立し、マーケティング活動にとって有用な消費者行動予測モデルを量産して行くことが重要である。

参考文献

- [1] 荒牧英治, 増川佐知子, 森田瑞樹, 保田祥(2012). 日本人のオンライン・コミュニケーション上での平均使用語彙数は 8,000 語である. 情報処理学会研究報告自然言語処理(NL), 2012-NL-208(9)
- [2] R・ドーキンス(1987). 日高敏隆他訳. 利己的な遺伝子. 紀伊国屋書店
- [3] R・ドーキンス(1976). 日高敏隆他訳. 延長された表現型. 紀伊国屋書店
- [4] 池尾恭一, 青木幸弘, 南知恵子, 井上哲浩(2010). マーケティング. 有斐閣
- [5] 奥村学(2012). マイクロブログマイニングの現在. 電子情報通信学会第3回集合知シンポジウム
- [6] 小林哲郎(2012). ソーシャルメディアと分断する社会的リアリティ. 人工知能学会誌 Vor27 No. 1
- [7] 本村陽一(2006). ベイジアンネットワーク技術. 東京電機大学出版局