

社会的影響と広告の影響とを考慮した購買行動モデルに基づく購買要因の推定

田中 佑典^{1,a)} 倉島 健¹ 藤原 靖宏² 岩田 具治³ 澤田 宏¹

概要: シングルソースデータからユーザの購買行動を引き起こした要因を推定するための購買行動モデルを提案する。シングルソースデータとは商品の購買履歴と広告閲覧履歴とを各ユーザ ID に紐づけて収集したものである。1) 個人の嗜好, 2) 他者から受ける影響, 3) メディア広告から受ける影響の3つの要因に基づく確率過程にしたがって起こるものとしてユーザの購買行動をモデル化する。提案モデルは購買行動が起こった時刻において各要因がどの程度影響しているかを分析することにより、購買行動を引き起こした要因を推定することができる。実シングルソースデータを用いた評価実験では、提案モデルが比較手法よりも高精度にユーザの購買行動を予測可能なことを確認した。この結果は、提案モデルがユーザの購買行動をより正しく説明できることを示しており、提案モデルにより推定された購買要因が妥当であることを示唆するものである。さらに、企業による TV 広告がどの程度購買につながったかという観点で広告効果を分析した結果について報告する。

1. はじめに

近年、TV や Web、ソーシャルメディアなどの普及により、ユーザは日常的に多くの情報に接触しながら生活している。ユーザはこれらの情報に影響され、購買行動における意思決定を行うことも多い。例えば、友人に推薦された商品を購入するといった場合や、TV で宣伝された商品を購入するといった場合などが挙げられる。情報接触がユーザの購買行動に与える影響について分析する技術は、マーケティングが広告戦略の最適化 [9] やバイラルマーケティング [5] を行う際に非常に有用である。

マーケティング調査会社は、実施した広告戦略の効果を定量的に評価するために、ユーザの商品購入履歴および広告閲覧履歴を継続的に収集している。このデータは“シングルソース”（同一のユーザ）から収集された購買および広告閲覧履歴であり、シングルソースデータと呼ばれる。シングルソースデータを用いて、メディア広告の効果を推定する手法がマーケティングサイエンスの分野において多

く提案されている [6], [14]。しかし、これらの手法の多くは、ユーザの購買行動が広告接触のみに基づいて起こるということを仮定しており、現代のユーザを取り巻く社会環境に適合していない。Web やソーシャルメディアが普及している現代の社会環境においては、他ユーザによる推薦のような社会的影響による購買行動を考慮することがますます重要となってきている。既存手法はこのような社会的影響による購買行動を考慮しないため、広告接触がユーザに与える影響を正確に推定することはできない。

本研究の目的は、シングルソースデータに基づいて、購買行動を引き起こす要因を推定し、マーケティングにとって有用な知見を抽出することである。しかし、購買行動の要因を正確に推定することは難しい問題である。なぜなら、購買および広告閲覧時刻の順序関係や近さが、購買を引き起こす直接の因果関係を示しているとは限らないからである。例えば、購買の直前に友人から商品の推薦を受けたり、TV 広告を閲覧したからといって、それらが購買の要因であるとは限らない。また、これらのような外部からの影響に依存せず、ユーザ個人の嗜好に基づいて購買がなされることも考慮しなければならない。

本稿では、購買を引き起こす要因を効果的に推定するために、ユーザ間および企業とユーザ間の関係の強さをモデル化する。これは、ユーザ間および企業とユーザ間の関係の強さに応じて、これらの間を伝わる影響の大きさが異なると考えられるためである。例えば、友人や家族による商

¹ 日本電信電話株式会社 NTT サービスエボリューション研究所
NTT Service Evolution Laboratories, Yokosuka-shi, Kanagawa, 239-0847 Japan

² 日本電信電話株式会社 NTT ソフトウェアイノベーションセンター
NTT Software Innovation Center, Musashino-shi, Tokyo, 180-8585 Japan

³ 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所
NTT Communication Science Laboratories, Soraku-gun, Kyoto, 619-0237 Japan

a) tanaka.y@lab.ntt.co.jp

品推薦は見知らぬ人からの推薦よりもユーザの購買意欲に大きな影響を与えることが想定される。同様に、ユーザにとってなじみのある企業による広告は、そうでない企業によるものと比べて影響は大きいと考えられる。このようなことを念頭において、本稿では、シングルソースデータに基づきユーザの購買行動を引き起こす要因を推定するための新たな確率モデルを提案する。提案モデルでは、ユーザの購買行動が、1) 個人の嗜好、2) 他者から受ける影響、3) メディア広告から受ける影響の3つの要因に基づく確率過程にしたがって起こるものと仮定する。そして、ユーザ間および企業とユーザ間に存在する影響関係をモデルに組み込み、ユーザの購買意欲をモデル化する。提案モデルは、確率的EMアルゴリズムを用いて各購買行動を引き起こした要因を推定可能である。また、メディア広告による影響だけでなく、他者からの社会的影響および個人の嗜好をも同時に推定することができるという特徴を持つ。これにより、各ユーザが購買行動を行う際に、これら3つの要因にどの程度影響されているかについて定量的に評価することが可能となる。これを実現することはマーケティングにとって重要な意味を持つ。例えば、過去に実施した広告戦略がどの程度商品の購買につながったかという観点で広告戦略の評価を行うことができる。

評価実験では、購買およびTV広告閲覧履歴からなる実シングルソースデータを用いて提案モデルの有効性を示す。購買予測精度を評価指標として、提案モデルが比較手法よりも高い性能をもつことを確認した。この結果は、提案モデルがユーザの購買行動をより正しく説明できることを示しており、提案モデルにより推定された購買要因の妥当性を示唆するものである。また、購買行動を引き起こした要因を定量的に評価し、新商品に対する広告効果や、購買行動に影響を与えるユーザや企業を分析した結果を報告する。

2. 関連研究

ソーシャルメディア上におけるユーザの選択行動履歴（話題の共有や商品の購買など）から、ユーザが構築するソーシャルネットワーク構造を推定するための手法としてカスケードモデル [1], [8] が提案されている。カスケードモデルは、バイラルマーケティング [5] やインフルエンサー抽出 [12] を行う際によく利用される。カスケードモデルを用いれば、ユーザネットワークの構造が未知であるという条件の下で、最尤推定に基づきユーザの選択行動を尤もらしく説明するようなネットワーク構造を推定可能である。しかし、カスケードモデルは、データの観測期間内において各ユーザが話題や商品を一人につき一回だけ選択するという仮定をする。この仮定は、商品を複数回選択することが容易に想定される購買行動をモデル化する場合には適切ではない。このような仮定を置かない、より柔軟

なモデルとして点過程の枠組みに基づく手法が提案されている。点過程 [11] とは、連続時間軸上においてイベントの発生時刻をモデル化するための確率過程である。時刻と付加情報（本稿ではユーザ）のペアで表されるイベントの発生時刻をモデル化するための手法としてマーク付き点過程 [4] がある。マーク付き点過程に基づき、ソーシャルメディア上においてユーザが様々なアイテムを選択した履歴から、ユーザ間の影響関係を推定可能な shared cascade Poisson process (SCPP) [3] が提案されている。しかし、SCPPは購買行動をモデル化する際に重要な要因である広告の影響を考慮していない。提案モデルはSCPPを拡張し、購買履歴と広告閲覧履歴とを同時に分析することを可能とした。これにより、提案モデルを用いれば、広告がユーザの購買行動に与えた影響を推定し定量的に評価することができる。さらに、ユーザ間における社会的影響および商品に対する個人の嗜好についても推定することが可能である。

マーケティングサイエンスの分野では、企業による広告量と商品の購買量に基づき、メディアによる広告の効果を分析するための手法が多く提案されている。これらの手法の多くは、広告が商品の購買量に対して与える影響をモデル化し、広告量と購買量との相関をマクロな視点で分析するものである [6], [14]。しかし、これらの手法では、各ユーザに対して広告がどの程度影響を与えたかといった個人に紐付いた分析はできない。各ユーザに対して広告ストック（広告の影響の蓄積量）を定義し、シングルソースデータに基づいて各ユーザに対する広告効果を推定する手法 [15] が提案されている。しかし、この手法は、ユーザの購買行動が広告の影響によってのみ引き起こされると仮定しているため、購買を引き起こす他の要因を考慮できない。提案モデルは3つの要因を考慮し、従来手法と異なり購買行動に影響を与える他の要因を加味した上で、広告効果を推定することができる。

提案モデルのように購買を引き起こす要因として複数の要因を考慮する手法が提案されている [2], [13]。しかし、これらの手法では、広告や社会的影響を全ユーザに共通に与えられる要因として扱う。そのため、各ユーザに対する各要因の影響の大きさの違いを考慮することはできない。例えば、複数のユーザから同一の商品を推薦されたとしたとき、どのユーザの影響を受けて購買に至ったかという要因を特定することはできない。したがって、これらの手法では、ユーザ毎に異なる他ユーザや企業との影響関係を考慮して購買要因の推定を行うことはできない。

3. シングルソースデータ

本章では、シングルソースデータの詳細について述べる。シングルソースデータとは、“シングルソース”（同一のユーザ）から収集された購買および広告閲覧履歴である。シングルソースデータは、広告効果の測定を目的とし

表 1 シングルソースデータに関する記号とその定義

記号	定義
I	アイテム集合
U	ユーザ集合
M	企業集合
i	アイテム, $i \in I$
u_{ij}^p	アイテム i の j 番目の購買行動を行ったユーザ $u_{ij}^p \in U$
u_{ik}^v	アイテム i の k 番目の広告閲覧を行ったユーザ $u_{ik}^v \in U$
m_i	アイテム i を販売する企業, $m_i \in M$
t_{ij}^p	アイテム i の j 番目の購買行動がなされた時刻
t_{ik}^v	アイテム i の k 番目の広告閲覧がなされた時刻
D	購買イベント集合
C	TV 広告閲覧イベント集合
T	観測期間
J_i	時刻 $t \in [0, T]$ におけるアイテム i の総購買数
K_i	時刻 $t \in [0, T]$ におけるアイテム i の広告の総閲覧数

て, TiVO Research Inc*1 や株式会社インテージ*2 などのマーケティング調査会社によって収集されている。

第 5 章の評価実験では, 株式会社インテージによって収集された全国個人消費者パネル調査 (SCI) およびインテージシングルソースパネル (i-SSP) を用いた。SCI は日常の買い物時における購買履歴を記録したデータである。ユーザが買い物から帰ってきた時点で, ユーザ自身がオンライン上のデータ収集システムに購買情報を入力することでデータが蓄積される。i-SSP はユーザの自宅における TV 広告の閲覧履歴を記録したデータである。TV 広告の閲覧履歴は, 自宅の TV に接続された専用の記録装置を用いて自動的に収集される。

本稿で用いるシングルソースデータに関する記号とその定義を表 1 に示す。アイテム i の j 番目の購買イベントを (t_{ij}^p, u_{ij}^p) とし, ユーザ u_{ij}^p が時刻 t_{ij}^p においてアイテム i を購入したことを表す。 J_i をアイテム i の総購買数としたとき, 時刻 0 から時刻 T までの観測期間内になされたアイテム i の購買イベント集合を $D_i = \{(t_{ij}^p, u_{ij}^p)\}_{j=1}^{J_i}$ とする。また, アイテム集合を I とし, 全購買イベント集合を $D = \{D_i\}_{i \in I}$ とする。購買イベントは時刻順に並んでおり, $j < j'$ のとき $t_{ij}^p \leq t_{ij'}^p$ を満たすものとする。ここで, この仮定はデータの一般性を失うものではない。次に, アイテム i の k 番目の広告閲覧イベントを (t_{ik}^v, u_{ik}^v) とし, ユーザ u_{ik}^v が時刻 t_{ik}^v においてアイテム i の広告を閲覧したことを表す。 K_i をアイテム i の広告の総閲覧数としたとき, 観測期間内におけるアイテム i の広告閲覧イベント集合を $C_i = \{(t_{ik}^v, u_{ik}^v)\}_{k=1}^{K_i}$ とする。また, 全広告閲覧イベント集合を $C = \{C_i\}_{i \in I}$ とする。購買イベントと同様に, $k < k'$ のとき $t_{ik}^v \leq t_{ik'}^v$ を満たすものとする。ここで, ユーザ集合を U とし, 購買イベント集合と広告閲覧イベント集合とにおいて共通であることに注意が必要である。また, 企業集合を M としたとき, アイテム i を販売する企業 $m_i \in M$ が与えられる。

*1 <http://www.tivoresearch.com>

*2 <http://www.intage.co.jp>

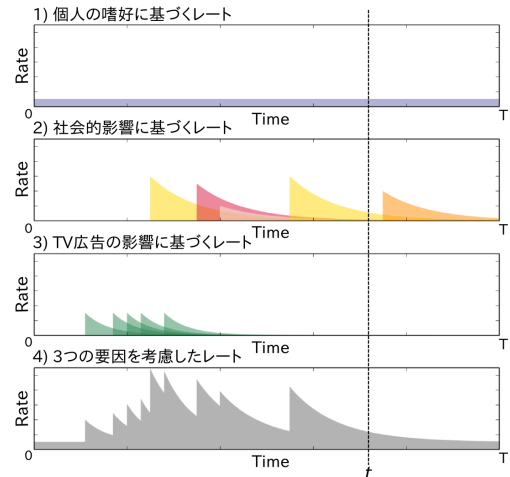


図 1 時刻 t におけるユーザ u のアイテム i の購買レート。1) 個人の嗜好に基づくレート。2) 他ユーザからの社会的影響に基づくレート。ユーザ毎の影響の違いを色の違いで表す。3) TV 広告の影響に基づくレート。4) 時刻 t 以前の購買および広告閲覧イベントに基づく 3 つの要因を考慮したレート。

4. 提案法

4.1 モデル

提案モデルでは, 購買イベントがマーク付き点過程 [4] にしたがって生成されると仮定する。各アイテムに対して時刻とユーザのペア (t, u) で表される購買イベントの生成過程をモデル化する。マーク付き点過程では, 非負の値をとる購買レート $\lambda_i(t, u)$ によってイベントの生起のしやすさが決まる。ここで $\lambda_i(t, u)$ は, 単位時間当たりユーザ u が時刻 t にアイテム i を購入する確率を表す。1) 個人の嗜好, 2) 他者から受ける影響, 3) メディア広告から受ける影響の 3 つの要因に基づく購買レートを設計する。 $f_i(u)$ をユーザ u がアイテム i を個人の嗜好に基づいて購入するレートとする。次に, $g_{(i,t',u')}(t, u)$ をユーザ u' が時刻 t' にアイテム i を購入したことによる影響で, ユーザ u が時刻 t にアイテム i を購入するレートとする。さらに, $h_{(i,t',m_i)}(t, u)$ をユーザ u が時刻 t' にアイテム i の TV 広告を閲覧したことによる影響で, ユーザ u が時刻 t にアイテム i を購入するレートとする。各購買レートに関する詳細は後ほど述べる。提案モデルでは, ユーザの購買行動が, 1) 個人の嗜好, 2) 過去他ユーザの購買イベント, 3) 過去の TV 広告閲覧イベントによって引き起こされると仮定する。3 つの要因を考慮した購買レートを,

$$\lambda_i(t, u) = f_i(u) + \sum_{(t', u') \in D_{it \setminus u}} g_{(i,t',u')}(t, u) + \sum_{(t', u') \in C_{itu}} h_{(i,t',m_i)}(t, u) \quad (1)$$

と表す。ここで, $D_{it \setminus u} = \{(t', u') | (t', u') \in D_i \text{ and } t' < t \wedge u' \neq u\}$ は, ユーザ u の購買を除いて時刻 t より以前に起こったアイテム i の購買イベント集合を表す。また,

$C_{itu} = \{(t', u') | (t', u') \in C_i \text{ and } t' < t \wedge u' = u\}$ は、時刻 t より以前に起こったユーザ u の TV 広告閲覧イベント集合を表す。

個人の嗜好に基づく購買レート

ユーザは各アイテムに対して異なる嗜好を持つと仮定する。また、図 1(a) に示すように、ユーザのアイテムに対する嗜好は観測期間内において一定であると仮定する。個人の嗜好に基づきアイテム i をユーザ u が購入するレートを、

$$f_i(u) = \alpha_i \phi_{iu} \quad (2)$$

と表す。ここで、 $\alpha_i \geq 0$ は、過去の購買および広告閲覧イベントに依らずアイテム i が購入されるレートであり、アイテム i の人気度を表す。また、 $\phi_{iu} \geq 0$ は、アイテム i を購入するユーザが、ユーザ u である確率であり、 $\sum_{u \in U} \phi_{iu} = 1$ を満たすものとする。

他ユーザからの社会的影響に基づく購買レート

ユーザの購買行動が過去に他ユーザの購買行動によって引き起こされると仮定する。例として、友人が購入した商品を推薦され同じものを購入するという行動が挙げられる。図 1(b) に他ユーザからの社会的影響に基づく購買レートを図示する。黄、赤、白、オレンジで表される 4 人の異なるユーザからの影響を表す。図 1(b) に示すように、他ユーザからの影響の大きさはユーザ毎に異なるものとする。例えば、友人や家族のような親しいユーザからの影響は見知らぬ人よりも大きいことが考えられる。他ユーザからの影響による購買レートを、

$$g_{(i,t',u')}(t, u) = \begin{cases} \alpha_{u'} \theta_{u'u} e^{-\gamma^p(t-t')} & \text{if } t' < t \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

と表す。ここで、 $\alpha_{u'} \geq 0$ はユーザ u' の影響力を表す。 $\theta_{u'u} \geq 0$ は、ユーザ u' からユーザ u への関係の強さを表し、 $\sum_{u \in U \setminus u'} \theta_{u'u} = 1$ を満たすものとする。ただし、 $\setminus u'$ は集合からユーザ u' を除くことを表す。そして、これらのパラメータの積 $\alpha_{u'} \theta_{u'u}$ は、ユーザ u' からユーザ u への影響の大きさを表す。また、 $e^{-\gamma^p(t-t')}$ は、パラメータを $\gamma^p \geq 0$ とし、時間と共に影響が減衰することをモデル化している。アイテムの流行は時間と共に移り変わるため、他ユーザからの影響が減衰するという仮定は自然である。この仮定は、カスケードモデルにおいてユーザ間の影響関係を推定する際にもよく用いられる [1], [8]。式 (3) において、各購買イベントは過去の購買には影響を与えないので $t' \geq t$ のとき購買レートは 0 となる。

TV 広告の影響に基づく購買レート

ユーザの購買行動が過去の TV 広告閲覧によって引き起こされると仮定する。また、TV 広告を通じた企業からの影響はユーザによって異なるものとする。例えば、ユーザにとってなじみのある企業の TV 広告は、そうでない企業よりも大きな影響を与えることが考えられる。図 1(c) に TV

表 2 提案モデルのパラメータ

パラメータ	定義
α_i	アイテム i の人気度, $\alpha_i \geq 0$
α_u	ユーザ u の影響力, $\alpha_u \geq 0$
α_m	企業 m の影響力, $\alpha_m \geq 0$
ϕ_{iu}	ユーザ u のアイテム i に対する嗜好 $\phi_{iu} \geq 0, \sum_{u \in U} \phi_{iu} = 1$
$\theta_{u'u}$	ユーザ u' からユーザ u への関係の強さ $\theta_{u'u} \geq 0, \sum_{u \in U \setminus u'} \theta_{u'u} = 1$
ω_{mu}	企業 m からユーザ u への関係の強さ $\omega_{mu} \geq 0, \sum_{u \in U} \omega_{mu} = 1$
γ^p, γ^v	減衰パラメータ, $\gamma^p \geq 0, \gamma^v \geq 0$

広告の影響に基づく購買レートを図示する。TV 広告の影響に基づく購買レートを、

$$h_{(i,t',m_i)}(t, u) = \begin{cases} \alpha_{m_i} \omega_{m_i u} e^{-\gamma^v(t-t')} & \text{if } t' < t \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

と表す。ここで、TV 広告の影響は企業 m_i からユーザ u への一方向であり、ユーザ間の影響のように双方向ではないことに注意が必要である。 $\alpha_{m_i} \geq 0$ は企業 m_i の影響力を表す。 $\omega_{m_i u} \geq 0$ は、企業 m_i からユーザ u への関係の強さを表し、 $\sum_{u \in U} \omega_{m_i u} = 1$ を満たすものとする。そして、これらのパラメータの積 $\alpha_{m_i} \omega_{m_i u}$ は、企業 m_i からユーザ u への影響の大きさを表す。また、 $e^{-\gamma^v(t-t')}$ は、パラメータを $\gamma^v \geq 0$ とし、TV 広告の影響が時間と共に減衰することを表す。文献 [14] にみられるように、広告の影響は指数関数を用いてモデル化されることが多く本研究でもそれを採用した。

式 (2)~式 (4) を式 (1) に代入し、3つの要因を考慮した購買レートを得る (図 1(d))。提案モデルのパラメータを表 2 にまとめる。アイテムの人気度を $\{\alpha_i\}_{i \in I}$ 、ユーザの影響力を $\{\alpha_u\}_{u \in U}$ 、企業の影響力を $\{\alpha_m\}_{m \in M}$ と表す。また、ユーザ間の関係の強さを $\Theta = \{\theta_{uu'}\}_{u \in U}$ 、 $\theta_u = \{\theta_{uu'}\}_{u' \in U \setminus u}$ 、企業とユーザ間の関係の強さを $\Omega = \{\omega_m\}_{m \in M}$ 、 $\omega_m = \{\omega_{mu}\}_{u \in U}$ と表す。さらに、ユーザのアイテムに対する嗜好を $\Phi = \{\phi_i\}_{i \in I}$ 、 $\phi_i = \{\phi_{iu}\}_{u \in U}$ 、減衰パラメータを $\gamma = \{\gamma^p, \gamma^v\}$ と表す。次節では、これらのモデルパラメータの推定方法について説明する。

4.2 推定

購買イベント集合 D と TV 広告閲覧イベント集合 C とが与えられたとき、表 2 に示すモデルパラメータを推定することを考える。以下では、確率的 EM アルゴリズムに基づき推定を行う方法について述べる。アイテム i に関する購買イベント集合 D_i が、購買レートを $\lambda_i(t, u)$ とするマーク付き点過程にしたがって生成されたとしたとき、全購買イベント集合 D の尤度関数は、

$$P(D|C, \alpha, \Theta, \Omega, \Phi, \gamma) = \prod_{i \in I} \exp \left(- \int_0^T \sum_{u \in U} \lambda_i(t, u) dt \right) \prod_{j=1}^{J_i} \lambda_i(t_{ij}^p, u_{ij}^p) \quad (5)$$

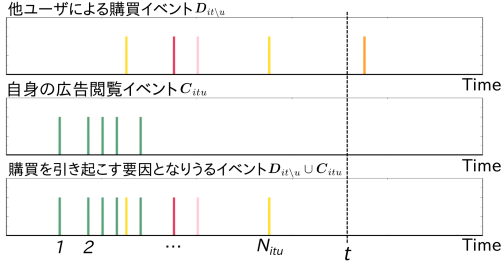


図 2 時刻 t において購買を引き起こす要因となりうる購買および広告閲覧イベントの集合。

と表すことができる。ここで、式 (5) の指数で表される因子は、観測期間において購買イベントが起こらない確率を表す。また、式中の積分は解析的に計算ができ、

$$\int_0^T \sum_{u \in U} \lambda_i(t, u) dt = \alpha_i T + \frac{1}{\gamma^p} \sum_{j=1}^{J_i} \alpha_{u_{ij}} \left(1 - e^{-\gamma^p (T - t_{ij}^p)}\right) + \frac{\alpha_{m_i}}{\gamma^v} \sum_{k=1}^{K_i} \left(1 - e^{-\gamma^v (T - t_{ik}^v)}\right) \quad (6)$$

となる。

点過程における重ね合わせの定理 [11] より、複数の独立な点過程の重ね合わせは、各レートを持つ独立な点過程に分解することができる。この定理を利用して、パラメータ推定を効率的に行えることが知られている [3]。アイテム i の j 番目の購買イベントを引き起こす要因となったイベントのインデックスを表す潜在変数 $z_{ij} \in \{0, \dots, N_{itu}\}$ を導入する。ここで、 $z_{ij} = 0$ は、購買イベントが過去の購買および広告閲覧イベントに依らず、個人の嗜好に基づいて起こったことを表す。また、図 2 に示すように時刻 t において購買を引き起こす要因となりうる購買および広告閲覧イベントの集合を $D_{it \setminus u} \cup C_{itu}$ としたとき、この集合に含まれるイベント数を N_{itu} とする。潜在変数 z を用いることにより、式 (1) で表される購買レートは $\lambda_i(t, u) = \sum_z \lambda_i(t, u, z)$ と書き直すことができる。 $\lambda_i(t, u, z)$ は、

$$\lambda_i(t, u, z) = \begin{cases} f_i(u) & \text{if } z = 0 \\ g(i, t_{iz}, d_{iz})(t, u) & \text{if } z \neq 0 \wedge d_{iz} \in U \\ h(i, t_{iz}, d_{iz})(t, u) & \text{if } z \neq 0 \wedge d_{iz} \in M \end{cases} \quad (7)$$

と表される。ここで、 $d_{iz} \in U \cup M$ はダミー変数であり、 z_{ij} が購買イベントを指し示す場合は $d_{iz} = u_{iz}$ を、 z_{ij} が広告閲覧イベントを指し示す場合は $d_{iz} = m_i$ を表す。

式 (5)~式 (7) を用いることにより、購買イベント集合 D および潜在変数集合 $Z = \{\{z_{ij}\}_{j=1}^{J_i}\}_{i \in I}$ の同時分布は、

$$P(D, Z | C, \alpha, \Theta, \Omega, \Phi, \gamma) = \prod_{i \in I} \exp \left[\alpha_i T + \frac{1}{\gamma^p} \sum_{j=1}^{J_i} \alpha_{u_{ij}} \left(1 - e^{-\gamma^p (T - t_{ij}^p)}\right) + \frac{\alpha_{m_i}}{\gamma^v} \sum_{k=1}^{K_i} \left(1 - e^{-\gamma^v (T - t_{ik}^v)}\right) \right] \prod_{j=1}^{J_i} \lambda_i(t_{ij}^p, u_{ij}^p, z_{ij}) \quad (8)$$

と表される。さらに、パラメータ θ_u , ω_m , ϕ_i に対する事前分布としてディリクレ分布を用い、それぞれ $P(\Theta) \propto \prod_{u \in U} \prod_{u' \in U \setminus u} \theta_{uu'}^{\beta_1 - 1}$, $P(\Omega) \propto \prod_{m \in M} \prod_{u \in U} \omega_{mu}^{\beta_2 - 1}$, $P(\Phi) \propto \prod_{i \in I} \prod_{u \in U} \phi_{iu}^{\beta_3 - 1}$ とする。ここで、各ディリクレ分布のハイパーパラメータをまとめて $\beta = \{\beta_1, \beta_2, \beta_3\}$ と表す。ハイパーパラメータ β は fixed-point iteration 法 [7] を用いて推定した。また、パラメータ $\alpha = \{\alpha_l\}_{l \in U \cup M \cup I}$ に対する事前分布としてガンマ分布を用い、 $P(\alpha_l) \propto \alpha_l^{a-1} \exp(-b\alpha_l)$ とする。ここで、ガンマ分布のハイパーパラメータを a, b とし、実験では $a = b = 1$ に固定した。これらの事前分布は、各パラメータに対する共役事前分布であるため、パラメータに関して周辺化を行うことができる。周辺化された同時分布は、

$$P(D, Z | C, \gamma, \beta, a, b) = \int P(D, Z | C, \alpha, \Theta, \Omega, \Phi, \gamma) P(\alpha | a, b) P(\Theta | \beta_1) \times P(\Omega | \beta_2) P(\Phi | \beta_3) d\alpha d\Theta d\Omega d\Phi \quad (9)$$

により解析的に計算できる。

確率的 EM アルゴリズムの手順について述べる。E ステップにおいて以下の式に従い Z をサンプリングする。

$$P(z_{ij} = y | D, Z_{\setminus ij}, C, \gamma, \beta, a, b) \propto \frac{P(D, Z_{\setminus ij}, z_{ij} = y | C, \gamma, \beta, a, b)}{P(D_{\setminus ij}, Z_{\setminus ij} | C, \gamma, \beta, a, b)} \quad (10)$$

式 (10) は、アイテム i の j 番目の購買イベントが、過去の y 番目のイベントによって引き起こされた事後確率を表す。さらに、本稿では M ステップにおいて γ をニュートン法を用いて最適化する。E ステップと M ステップをパラメータの値が収束するまで交互に繰り返す。これにより得られた Z を用いて、 α_l , θ_u , ω_m , ϕ_i に関する事後分布を計算し、事後分布最大化に基づいてパラメータの点推定の結果を得ることが可能である。紙面の都合上、詳細は割愛する。

5. 評価実験

本章では、実シングルソースデータを用いて提案モデルの有効性を検証する。5.1 節において、従来手法との購買予測精度の比較を行い提案モデルの妥当性を示す。さらに、5.2 節において、購買行動を引き起こした要因を定量的に評価した結果について述べる。

5.1 定量的評価

本実験の目的は、提案モデルがユーザの購買行動をより正確に予測可能なことを示し、5.2 節で述べるユーザの購買行動に関する知見の妥当性を担保することである。学習されたモデルの正しさを評価するために予測精度を用いることは一般的な方法である [3], [10]。

実験では、ユーザの購買および TV 広告閲覧履歴からな

表 3 データセットの基本的な統計量

カテゴリ	缶コーヒー	ビール
購買数	51,120	54,830
広告閲覧数	102,480	339,331
ユーザ数	743	1,135
企業数	11	7
アイテム数	40	36

る実シングルソースデータ^{*3}を用いた。データセットは、缶コーヒーとビールの2つのカテゴリからなり、観測期間は2013年1月1日から2013年12月31日までの52週間である。ここで、記録されている時刻の単位は日である。データセットの詳細を表3に示す。また、各データセットに対してテスト期間を8月、10月、12月と変えることにより3つの異なる訓練データとテストデータを用意し、それぞれを用いて評価を行った。例えば、8月をテスト期間とする際は、1月から7月までの期間を訓練期間とした。また、購買数があまりに少ないユーザは評価の対象外として、訓練期間において購買数が5回以下のユーザを除外した。

従来手法である shared cascade Poisson process (SCPP) [3] との比較を行う。SCPPは、他ユーザからの社会的影響のみを考慮して購買レートを計算するモデルである。社会的影響の大きさは、データから推定されたユーザ間の関係の強さにしたがって決まる。また、ベースライン手法として定常ポアソン過程 (homogeneous Poisson process, 以下 PP) との比較を行う。PPは、一定の購買レートにしたがって起こるイベントの発生時刻をモデル化するための確率過程である。本実験では、PPの購買レートを提案モデルにおける式(2)とする。これは、各ユーザが個人の嗜好のみに基づいて購買商品を選択するという仮定をおくことを意味する。各モデルを用いたときの予測精度をテストデータに対する負の対数尤度で評価する。テスト期間を T から T' としたとき、購買イベント (t, u) の負の対数尤度 L_1 は、

$$L_1 = \sum_{i \in I} \left(- \int_T^{T'} \sum_{u \in U} \lambda_i(t, u) dt \right) \sum_{(t, u) \in D_i^{\text{test}}} \log \lambda_i(t, u) \quad (11)$$

と表される。ここで、 D_i^{test} はアイテム i のテストデータである。 L_1 の値が小さいほど予測精度が高いことを意味する。図3に各データセットを用いたときの結果を示す。 x 軸はテストデータの観測月を、 y 軸は L_1 の値を表し、最も精度が高い手法に(*)を付けた。図3より、テスト期間に依らず、提案モデルは従来手法に比べて同等かそれ以上の予測精度を示すことが分かる。これにより、提案モデルは3つの要因を考慮することで、ユーザの購買行動をより適切に予測可能であることが示された。さらに、提案モデルは従来手法と異なり、社会的影響だけでなく個人の嗜好およびTV広告の影響も推定可能であるという特徴を持つ。

^{*3} 株式会社インテージによって収集された全国個人消費者パネル調査 (SCI) およびインテージシングルソースパネル (i-SSP)。

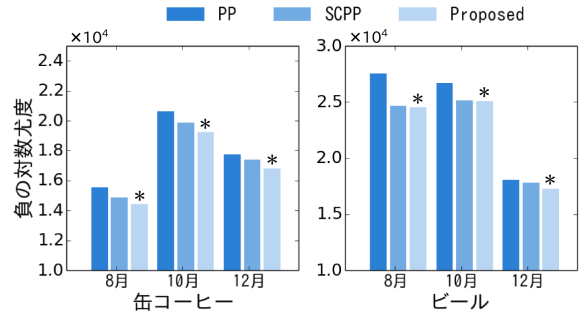


図 3 テスト期間における購買イベント (t, u) の負の対数尤度。

5.2 定性的評価

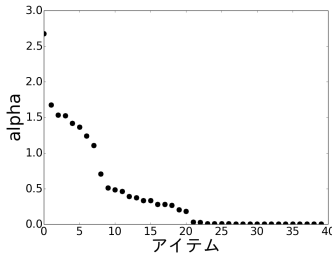
本節では、データから推定されたモデルパラメータを用いて、ユーザの購買行動に関する重要な知見が得られたことを示す。また、マーケターが活用することを想定し、抽出された知見がマーケティング戦略を立てる上でどのように有用であるかについて述べる。

購買行動に影響を与える要素の抽出

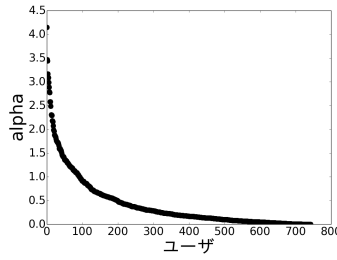
推定された提案モデルのパラメータを用いて、人気商品や影響力の大きいユーザおよび企業を抽出することができる。図4に、缶コーヒーのデータセットを用いたときのアイテムの人気度 $\{\alpha_i\}_{i \in I}$ 、ユーザの影響力 $\{\alpha_u\}_{u \in U}$ 、企業の影響力 $\{\alpha_m\}_{m \in M}$ の値をプロットした。ここで、アイテム、ユーザ、企業を推定されたパラメータの値にしたがって降順に並び替えた。この結果から、少数のアイテム、ユーザおよび企業がそれぞれ大きい値のパラメータを持つことが分かる。図4(a)から、缶コーヒー市場において、特定のアイテムだけが人気を持つということが示唆され、図4(b)から、消費者の中に他者の購買を促進するような影響力の大きい特定のユーザが存在することが示唆される。また、図4(c)から、TV広告を通してユーザに大きな影響を与える特定の企業が存在することが示唆される。これらの知見は、マーケティング戦略の最適化を行う際に有用である。例えば、人気商品を発見することにより不要な広告費用を削減できる可能性がある。提案モデルにより推定された人気商品は、外部からの影響とは関係なく購入されやすい商品の意味するため、広告量を少なくしても商品の購買量を維持できる可能性がある。また、バイラルマーケティングを行う際に影響力の大きいユーザをターゲットとすることで、より多くのユーザに対して二次的な波及効果を期待できる。さらに、広告を出す媒体として複数のメディアがあり、各メディアに対する広告費の分配を最適化したときにも活用できる。例えば、提案モデルによって影響が大きいと推定された企業は、TV広告に費用を分配することが購買促進に有効であると期待できる。

購買行動を引き起こした要因の分析

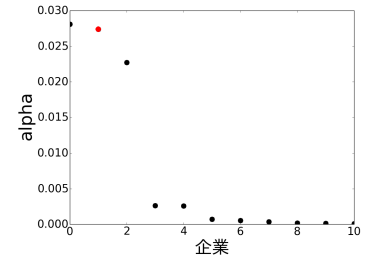
式(10)を用いて、購買を引き起こした要因を分析した結果について述べる。Algorithm 1に各購買イベントに対する3つの要因の影響の大きさを計算する手順を示す。Algorithm 1における入力、推定された式(10)の事後分



(a) アイテムの人気度 $\{\alpha_i\}_{i \in I}$



(b) ユーザの影響度 $\{\alpha_u\}_{u \in U}$



(c) 企業の影響度 $\{\alpha_m\}_{m \in M}$

図 4 缶コーヒーのデータセットを用いたときに推定された提案モデルのパラメータ. (a) アイテムの人気度, (b) ユーザの影響度, (c) 企業の影響度を表す. アイテム, ユーザおよび企業は, パラメータの値に基づいて降順に並び替えた. 図 (c) の赤点で表す企業は購買要因の分析対象とした企業を表す.

Algorithm 1 アイテム i の j 番目の購買イベントに対する 3 つの要因の影響の大きさの計算手順.

Require: $P(z_{ij}|D, Z_{\setminus ij}, C, \gamma, \beta, a, b)$, $\{d_{iy}\}_{y=1}^{N_{itu}}$

$E_f \leftarrow P(z_{ij} = 0|D, Z_{\setminus ij}, C, \gamma, \beta, a, b)$

$E_g \leftarrow 0, E_h \leftarrow 0$

$y \leftarrow 1$

while $y \leq N_{itu}$ **do**

if $d_{iy} \in U$ **then**

$E_g \leftarrow E_g + P(z_{ij} = y|D, Z_{\setminus ij}, C, \gamma, \beta, a, b)$

else

$E_h \leftarrow E_h + P(z_{ij} = y|D, Z_{\setminus ij}, C, \gamma, \beta, a, b)$

end if

$y \leftarrow y + 1$

end while

return E_f, E_g, E_h

布および 4.2 節で述べたダミー変数 $\{d_{iy}\}_{y=1}^{N_{itu}}$ である. また, E_f, E_g, E_h は, 個人の嗜好, 社会的影響, TV 広告の影響の大きさをそれぞれ表し, $E_f + E_g + E_h = 1$ を満たす.

表 4 に, 訓練期間を 2013 年 1 月 1 日から 2013 年 11 月 30 日としたときの各要因の影響の大きさの比率を示す. 両方のデータセットにおいて社会的影響の比率が大きいと推定された. この結果は, 様々なカテゴリにおいて, 他ユーザの影響が購買行動に対して大きな影響を与えることを示唆している. 一方で, TV 広告の影響の比率は, 全データセットにおいて小さいと推定された. この理由として以下のことが考えられる. 文献 [16] には, 広告の影響は, 直接的な影響と間接的な影響に大別されると述べられている. 直接的な影響とは, 例えば, 新商品の TV 広告をみてその商品を購入するというような影響を表す. 間接的な影響は, 広告がユーザの企業への印象や態度を形成し, 間接的に商品の購買意欲に影響を与える. 例えば, TV 広告によってユーザが企業に対して好印象を持ち, 将来商品を購入する際に, その企業の商品を選ぶというような場合である. 提案モデルはこのような二つの影響のうち, 直接的な広告の影響のみを推定する. 提案モデルでは, 間接的な広告の影響を推定することはできず, その影響は個人の嗜好か社会的影響かのどちらかの要因に吸収される. そのため, 表 4 における広告の影響の比率は直接的な広告の影響のみを表

表 4 推定された要因の比率 (%)

データセット	缶コーヒー	ビール
個人の嗜好	8.3	9.6
社会的影響	89.3	89.0
TV 広告の影響	2.4	1.4

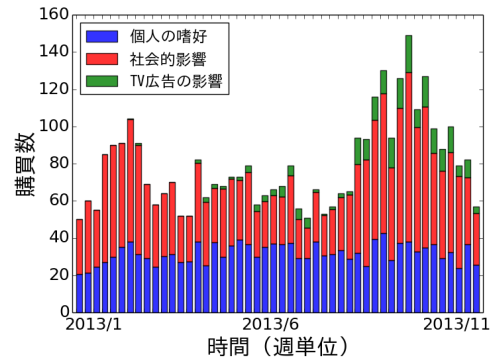
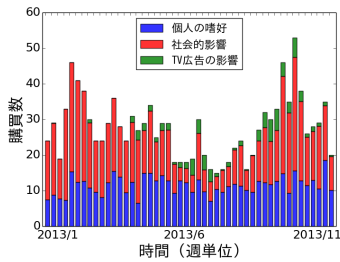


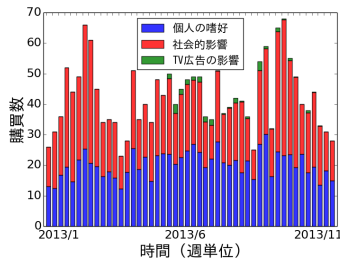
図 5 ある企業に着目したときの各要因によって引き起こされた推定された購買数. 棒グラフは各週の総購買数を表し, 3 つの色は各要因に基づく推定購買数を表す.

しており, 結果としてその影響は小さいと推定された.

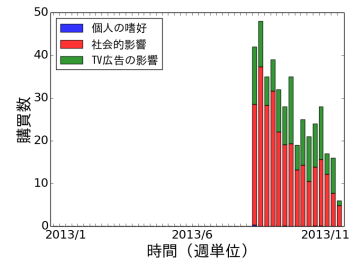
次に, 影響力が大きいと推定されたある企業に注目し, 購買を引き起こした要因について詳細に分析を行う. 図 4(c) における赤点で示す企業は, 缶コーヒーのデータセットにおいて影響力が大きいと推定された企業を表す. この企業に注目した理由は, 訓練期間において新商品やリニューアル商品の発売があり分析対象として適切であると考えたためである. 図 5 に, 缶コーヒーのデータセットにおいて推定された購買要因を示す. 各購買イベントに対して推定された E_f, E_g, E_h の値を週毎に集計し棒グラフで表した. ここで, 棒グラフの青, 赤, 緑は, 個人の嗜好, 社会的影響および TV 広告の影響によって引き起こされた推定された購買数をそれぞれ表す. 図 5 より, 個人の嗜好に基づく購買数は時間によってあまり変化せず, 社会的影響や TV 広告の影響に基づく購買数は時間によって変動すると推定された. このことは, 流行や TV 広告などに関係なく個人の嗜好によって買われやすい商品が存在することを示唆している. また, TV 広告の影響は訓練期間の最後の四半期において比較的大きいと推定された. 缶コーヒーは, 秋冬には温かい缶コーヒーが, 春夏には冷たい缶コーヒーが多



(a) リニューアル商品 1



(b) リニューアル商品 2



(c) 新商品

図 6 新商品およびリニューアル商品に対して推定された各要因に基づく購買数.

く販売される。缶コーヒーの冷温が切り替わる時期に伴って新商品やリニューアル商品が発売されている。図 6(a) および図 6(b) にリニューアル商品に対する購買要因, 図 6(c) に新商品に対する購買要因をそれぞれ示す。リニューアル商品は新商品に比べて個人の嗜好に基づく購買が多く, 一方で新商品は社会的影響や TV 広告の影響による購買が多いと推定された。この理由として, 既存の商品 (リニューアル商品) はすでに嗜好や購買態度がある程度形成されており, 外部からの影響とは関係なく購入するユーザが存在することが考えられる。また, 新商品に対する TV 広告は直接購買に結びつきやすく, 既存の商品に対しては直接の効果が得られにくいということを示している。以上の結果より, 提案モデルは, 購買行動を引き起こす要因について定量的に評価が可能である。このような評価結果はマーケティングにとって非常に有用である。例えば, マーケターがある広告戦略を実施したとき, 提案モデルを用いれば, その広告戦略がどの程度購買につながったかという観点で戦略の評価を行うことができる。

6. 結論

本稿では, シングルソースデータからユーザの購買行動を引き起こした要因を推定するための新たな購買行動モデルを提案した。提案モデルは, 1) 個人の嗜好, 2) 他者から受ける影響, 3) メディア広告から受ける影響の 3 つの要因を考慮し, 各購買行動が起こった要因を推定可能である。実シングルソースデータを用いた評価実験では, 従来手法に比べて高い精度で購買予測を行うことができることを示した。また, 推定された提案モデルのパラメータを用いて, 人気商品や影響力の大きいユーザおよび企業の存在を確認した。さらに, 新商品やリニューアル商品の発売に伴った TV 広告がどの程度購買につながったかという観点で広告効果の定量的な評価が可能であることを示した。今後は, 推定された購買要因をユーザ毎に集計し, 各ユーザが影響を受けやすい要因について分析することを検討している。

参考文献

- [1] M. Gomez-Rodriguez, D. Balduzzi, and B. Scholkopf, "Uncovering the Temporal Dynamics of Diffusion Networks," In *ICML'11*, pages 561–568, 2011.
- [2] A. Ishii, H. Arakaki, N. Matsuda, S. Umemura, T. Urushidani, N. Yamagata, and N. Yoshida, "The 'Hit' Phenomenon: a Mathematical Model of Human Dynamics Interactions as a Stochastic Process," *New Journal of Physics*, 14, 2012.
- [3] T. Iwata, A. Shah, and Z. Ghahramani, "Discovering Latent Influence in Online Social Activities via Shared Cascade Poisson Processes," In *KDD'13*, pages 266–274, 2013.
- [4] G. Last and A. Brandt, *Marked Point Processes on the Real Line: the Dynamical Approach (Probability and Its Applications)*, Springer, 1995.
- [5] J. Leskovec, L. A. Adamic, and B. A. Huberman, "The Dynamics of Viral Marketing," In *Proceedings of the 7th ACM Conference on Electronic Commerce*, pages 228–237, 2006.
- [6] N. Mehta, X. Chen, O. Narasimhan, "Informing, Transforming, and Persuading: Disentangling the Multiple Effects of Advertising on Brand Choice Decisions," *Marketing Science*, 27(3):334–355, 2008.
- [7] T. Minka, "Estimating a Dirichlet distribution," 2000.
- [8] S. A. Myers and J. Leskovec, "On the Convexity of Latent Social Network Inference," In *NIPS'10*, pages 1741–1749, 2010.
- [9] P. A. Naik, K. Raman, R. S. Winer, "Planning Marketing-Mix Strategies in the Presence of Interaction Effects," *Marketing Science*, 24(1):25–34, 2005.
- [10] P. Smyth, "Model Selection for Probabilistic Clustering Using Cross-Validated Likelihood," *Statistics and Computing*, 10(1):63–72, 2000.
- [11] D. L. Snyder and M. I. Miller, *Random Point Processes in Time and Space*, Springer, 1991.
- [12] X. Song, Y. Chi, K. Hino, and B. L. Tseng, "Information Flow Modeling Based on Diffusion Rate for Prediction and Ranking," In *WWW'07*, pages 191–200, 2007.
- [13] R. Takahashi, H. Mizuta, N. Abe, R. L. Kennedy, V. J. Jeffs, R. Shah, and R. H. Crites, "Collective Response Spike Prediction for Mutually Interacting Consumers," In *ICDM'13*, pages 727–736, 2013.
- [14] G. J. Tellis, R. K. Chandy, D. Macinnis, and P. Thaiwanich, "Modeling the Microeffects of Television Advertising: Which Ad Works, When, Where, for How Long, and Why?," *Marketing Science*, 24(3):351–366, 2005.
- [15] N. Terui, M. Ban, G. M. Allenby, "The Effect of Media Advertising on Brand Consideration and Choice," *Marketing Science*, 30(1):74–91, 2011.
- [16] D. Vakratsas and T. Ambler, "How Advertising Works: What Do We Really Know?," *Journal of Marketing*, 63(1):26–43, 1999.
- [17] D. J. Watts and P. S. Dodds, "Influence, Networks, and Public Opinion Formation," *Journal of Consumer Research*, 34:441–458, 2007.