

社会環境から受ける影響を考慮した選択行動モデルの検討

A Study on Probabilistic Behavior Models Considering Influences from Social Environments

田中 佑典*¹
Yusuke Tanaka

倉島 健*¹
Takeshi Kurashima

藤原 靖宏*²
Yasuhiro Fujiwara

岩田 具治*³
Tomoharu Iwata

澤田 宏*¹
Hiroshi Sawada

*¹NTT サービスエボリューション研究所
NTT Service Evolution Laboratories

*²NTT ソフトウェアイノベーションセンタ
NTT Software Innovation Center

*³NTT コミュニケーション科学基礎研究所
NTT Communication Science Laboratories

Our decision-making in purchasing is shaped by social influences. For example, a user may be attracted towards a popular item, an item recommended by others, or an item promoted by a television commercial. In this paper, we propose a probabilistic behavior model which takes into account three factors: (a) preferences of each user, (b) social influences from others, and (c) commercial stimuli from media. For modeling user's behaviors, we employ marked Poisson processes in which each factor is represented by their parameters. The parameters are estimated based on Bayesian inference. The proposed model can be used for finding influential users or companies, and evaluating which factor influences users. In experiments, the effectiveness of the proposed model is demonstrated by using real-world data sets.

1. はじめに

ユーザを取り巻く社会環境には、ユーザの選択行動に影響を与える様々な要因が存在する。ユーザは他者や企業などと影響関係にあり、それらから影響を受け、その結果として、商品購入などの選択を行うと考えられる。例えば、家族や友人に勧められて商品を購入する場合もあれば、TV コマーシャルを通じた企業からの商業的刺激によって商品を購入する場合もある。また、社会環境だけでなく、個人の嗜好が商品の選択に影響する場合もある。

マーケティング分野において、ユーザがどのような影響を受けて商品の選択を行ったかについて分析することは重要な課題である。例えば、TV コマーシャルが購買行動に与える影響の大きさを推定することができればマーケターが広告マネジメントを行う際に有用である。いつ、どのようなターゲットユーザに対して、どのくらいの広告を打てば効果を最大化できるかといった分析が可能となるためである。このような課題を解決するために、マーケティングリサーチ会社はユーザ ID に紐付いた複数メディアの接触および購買のログデータ（シングルソースデータ）を大規模に収集している。シングルソースデータを用いれば、各ユーザに対してメディア接触と購買行動とを関連付けて分析することが可能である。しかし、シングルソースデータを用いても TV コマーシャルの影響をどの程度受けて商品を購入したか、どの人から影響を受けて商品を購入したかといった、なぜ購買に至ったかという要因を観測できない。

本稿では、ユーザの購買ログおよび TV コマーシャル視聴ログからなるシングルソースデータを用いてユーザの選択行動をモデル化する。ここで、与えられるデータは、いつ、誰が、何を買ったか（視聴したか）という時系列情報のみであり、なぜ購買に至ったかという要因は未観測である。提案モデルでは、ユーザの選択行動が、(a) 個人の嗜好によるレート、(b) 他者の影響によるレート、(c) TV コマーシャルを通じた企業の影響によるレートの三つのレートを足し合わせたレートで決まると仮定する。ここで、レートとは単位時間当たりの商品の選択確

率を表す。提案モデルでは、ユーザの選択行動を、ユーザ間の影響の伝わりやすさ、企業からユーザへの影響の伝わりやすさおよび個人の嗜好をパラメータとして持つマーク付きポアソン過程に基づいてモデル化する。そして、これらのパラメータをベイズ推定に基づき推定する。これにより、TV コマーシャルを通して企業がユーザの購買行動に与えた影響の大きさの分析や、影響力の大きい企業およびユーザの抽出などが可能となる。

評価実験では、実シングルソースデータを用いて提案モデルの妥当性を検証する。また、実用性の観点から、TV コマーシャルを通して企業がユーザの購買行動に与えた影響の大きさについて分析した結果について述べる。

2. 関連研究

ユーザの選択行動が、他者の選択の影響によって決まると仮定し、情報や影響がユーザからユーザへと伝播する現象を表す代表的な数理モデルとしてカスケードモデル [4] がある。カスケードモデルでは、ユーザをノード、ユーザ間の影響関係の強さをエッジの重みとしたグラフを考え、観測現象をそのグラフ上での確率的な伝播プロセスとして捉える。未知パラメータとして推定するのは、ネットワーク構造、つまり、ユーザ間の影響関係の強さを示すエッジの重みである。これを用いて影響力の大きいユーザの抽出を行い、バイラルマーケティング [7] やインフルエンサー抽出 [10] などに利用できる。近年、カスケードモデルに基づく様々な手法 [2, 3] が提案されている。Netrate [2] は各ユーザ間において異なるレートに基づき影響の伝播が起こると仮定し、ユーザ間の影響関係を推定する。また、ユーザ間の影響の伝わりやすさをパラメータとして持つポアソン過程を用いて、ベイズ推定に基づき他者の選択の影響を考慮したユーザの選択行動モデルを推定する手法も提案されている [3]。しかし、従来のモデルは、ユーザの選択行動が他者の選択の影響のみによって決まると仮定しており、個人の嗜好やメディアの影響などによる選択行動を説明することができない。本稿で提案するモデルは、カスケードモデルを基礎として、他者の選択の影響に加えて、個人の嗜好および TV コ

表 1: 主な記号とその定義

| 記号 | 定義 |
|----------------------|---|
| I | アイテム集合 |
| U | ユーザ集合 |
| M | 企業集合 |
| u_{ij} | アイテム i の j 番目の購買行動を起こしたユーザ |
| t_{ij} | アイテム i の j 番目の購買行動が起こった時刻 |
| m_i | アイテム i を取り扱う企業 |
| N_i | アイテム i の購買回数 |
| T | 観測期間 |
| 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} | アイテム i とユーザ u との相対的な関係の強さ $\phi_{iu} \geq 0, \sum_{u \in U} \phi_{iu} = 1$ |
| $\theta_{u'u}$ | ユーザ u' からユーザ u への相対的な関係の強さ $\theta_{u'u} \geq 0, \sum_{u \in U \setminus u'} \theta_{u'u} = 1$ |
| θ_{mu} | 企業 m からユーザ u への相対的な関係の強さ $\theta_{mu} \geq 0, \sum_{u \in U} \theta_{mu} = 1$ |
| γ_1, γ_2 | 減衰パラメータ |

マーシャルを通じた企業の影響も併せて推定することができる。提案モデルにより、ユーザおよび企業をノードとし、ノード間をつなぐエッジの重みが影響の伝わりやすさを表すネットワーク構造を得ることができる。あるノード A から別のノード B へのエッジの重みが大きいことは、ノード A でイベントが起きた後にノード B で同じイベントが起きやすいことを意味する。これを用いることにより、TV コマーシャルを通して企業がユーザの購買に与えた影響の大きさの分析や、影響力の大きい企業およびユーザの抽出などが可能となる。

一方、カスケードモデルとは異なる文脈において、文献 [8] では、他者の選択の影響に加え、個人の嗜好がユーザの選択にもたらす影響をモデル化している。しかし、SNS における友人関係を既存知識としてモデルに組み込むことを前提としているため、購買データ解析のように友人関係が観測できない場合には用いることができない。また、文献 [9] では、TV コマーシャルや Web におけるアイテム閲覧などの履歴情報を入力とした回帰モデルを学習することにより、ユーザの購買行動を予測するための手法が提案されている。しかし、この手法はユーザ間の影響の伝播現象を扱うカスケードモデルとは異なるため、ユーザ間の影響関係を推定することはできない。提案モデルでは、カスケードモデルに基づき、各ユーザが購買および TV コマーシャル視聴を行った時刻情報を用いて、その時間的な近さや前後関係を考慮することにより、ユーザ間および企業とユーザ間の影響関係を推定することができる。

3. モデル

アイテムの集合を I としたとき、 $|I|$ 種類のアイテムの選択イベント集合を $\{D_i\}_{i \in I}$ と表す。またアイテム i の選択イベント数を N_i 個としたとき、 $D_i = \{(t_{in}, u_{in})\}_{n=1}^{N_i}$ と表す。ここで、 n 番目のイベントはユーザ u_{in} が時刻 t_{in} にアイテム i を選択したことを表す。主な記号とその定義について表 1 にまとめる。

提案モデルでは、 D_i は $\lambda_i(t, u)$ をレートとするマーク付き

ポアソン過程に従って以下のように生成されると仮定する。

$$D_i \sim \text{PoissonProcess}(\lambda_i(t, u)) \quad (1)$$

ここで、マーク付きポアソン過程とは、イベントの起こった時刻を生成する通常のポアソン過程に対して、イベントの起こった時刻と付加的な情報（今回の場合ユーザ情報）とを同時に生成する確率過程のことである。提案モデルのレートを以下の式で表す。

$$\lambda_i(t, u) = \xi_{(i)}^{\text{pref}}(u) + \sum_{(t', u') \in D_i} \xi_{(i, t', u')}^{\text{social}}(t, u) + \sum_{t' \in C_{iu}} \xi_{(i, t', m_i)}^{\text{media}}(t, u) \quad (2)$$

ここで、 C_{iu} は、ユーザ u がアイテム i の TV コマーシャルを視聴した時刻の集合を表す。また、 m_i はアイテム i を取り扱う企業を表す。提案モデルでは、複数の独立なポアソン過程の重ねあわせが、それぞれのレートの和をレートとするポアソン過程となる性質を用いた。詳細は文献 [5] に述べられている。以下では、式 (2) の各項にあたる各要因に基づくレートの詳細について述べる。

(a) 個人の嗜好によるレート

ユーザ u がアイテム i を個人の嗜好に基づいて選択するレートを以下の式で表す。

$$\xi_{(i)}^{\text{pref}}(u) = \alpha_i \phi_{iu} \quad (3)$$

ここで、個人の嗜好は時間によって変動しないものとする。 $\alpha_i \geq 0$ はアイテム i の人気度を表す。 $\phi_{iu} \geq 0$ はアイテム i とユーザ u の相対的な関係の強さを表しており、 U をユーザ集合とすると $\sum_{u \in U} \phi_{iu} = 1$ を満たすものとする。これら 2 つの因子の積 $\alpha_i \phi_{iu}$ は、ユーザ u のアイテム i に対する嗜好の強さを表す。

(b) 他者の選択の影響によるレート

ユーザ u' が時刻 t' にアイテム i を選択したことによる影響で、ユーザ u が時刻 t にアイテム i を選択するレートを以下の式で表す。

$$\xi_{(i, t', u')}^{\text{social}}(t, u) = \begin{cases} \alpha_{u'} \theta_{u'u} e^{-\gamma_1(t-t')} & \text{if } t' < t \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

ここで、式 (4) は $t' < t$ のときに値を持ち、時刻 t 以前の選択イベントから影響を受けるものとする。したがって、 $t' > t$ のときレートは 0 となる。また、 $\theta_{uu} = 0$ とし、自身の選択イベントからは影響を受けないものとする。式 (2) の第二項は、式 (4) をアイテム i の選択イベント集合について和をとったものである。これは、ユーザ毎の影響力を考慮したアイテム i のトレンドを表すと考えることもできる。 $\alpha_{u'} \geq 0$ 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 への相対的な関係の強さを表しており、 $\sum_{u \in U \setminus u'} \theta_{u'u} = 1$ を満たすものとする。ここで、 $\setminus u'$ は、ユーザ集合 U からユーザ u' を除くことを表す。これら二つの因子の積 $\alpha_{u'} \theta_{u'u}$ は、ユーザ u' からユーザ u への影響の伝わりやすさを表す。また、 $e^{-\gamma_1(t-t')}$ は、減衰パラメータを γ_1 として、他者の選択の影響が時間とともに減衰することを表している。アイテムのトレンドは時間とともに沈静化していくため、自然な仮定であると言える。

(c)TV コマーシャルを通した企業の影響によるレート

ユーザ u が時刻 t' にアイテム i の TV コマーシャルを視聴したことによる影響で、ユーザ u が時刻 t にアイテム i を選択するレートを以下の式で表す。

$$\xi_{(i,t',m_i)}^{\text{media}}(t, u) = \begin{cases} \alpha_{m_i} \theta_{m_i u} e^{-\gamma_2(t-t')} & \text{if } t' < t \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

ここで、TV コマーシャルを通した企業の影響は、アイテム i を取り扱う企業 m_i からユーザ u に与えられるものとする。 $\alpha_{m_i} \geq 0$ は企業 m_i の影響力を表す。 $\theta_{m_i u} \geq 0$ は企業 m_i からユーザ u への相対的な関係の強さを表しており、 $\sum_{u \in U} \theta_{m_i u} = 1$ を満たすものとする。ここで、TV コマーシャルを通した企業の影響は、企業からユーザへの一方である点に注意する。企業の影響を考慮するための最も素朴な方法として、企業をユーザの一種として扱い、企業の影響によるレートを式 (4) で計算することもできる。しかし、この方法では、企業とユーザが双方向に影響しあうことになってしまうため適切ではない。ユーザが TV コマーシャルを視聴した後、時間とともにその効果は薄れるものと考えられる。そのため、TV コマーシャルの影響は時間とともに減衰すると仮定する。ここで、TV コマーシャルの影響の減衰の仕方は、他者の影響の減衰の仕方とは異なると考えられる。例えば、口コミなどの他者の影響はある期間継続的に効果があるが、TV コマーシャルは短期間で効果がなくなってしまうといった、影響の継続期間に違いがあると考えられる。したがって、他者の影響の減衰パラメータとは別に、TV コマーシャルの影響の減衰パラメータを γ_2 とする。

観測データ $D = \{D_i\}_{i \in I}$ および $C = \{C_{iu}\}_{i \in I, u \in U}$ が与えられたとしたとき、以下の尤度関数を用いて各パラメータ $\alpha = \{\alpha_i\}_{i \in U \cup M \cup I}$ 、 $\Theta = \{\theta_i\}_{i \in U \cup M}$ 、 $\Phi = \{\phi_i\}_{i \in I}$ 、 $\gamma = \{\gamma_1, \gamma_2\}$ を推定する。

$$P(D|C, \alpha, \Theta, \Phi, \gamma) = \prod_{i \in I} \exp\left(-\int_0^T \sum_{u \in U} \lambda_i(t, u) dt\right) \prod_{n=1}^{N_i} \lambda_i(t_{in}, u_{in}) \quad (6)$$

ここで、 $\theta_{u'}$ = $\{\theta_{u'u}\}_{u \in U \setminus u'}$ 、 θ_m = $\{\theta_{mu}\}_{u \in U}$ 、 ϕ_i = $\{\phi_{iu}\}_{u \in U}$ である。また、企業集合を M とし、データの観測期間を T とした。ベイズ推定に基づき、各パラメータに対して事前分布を仮定し Stochastic EM アルゴリズム [1] を用いることにより、パラメータおよび事前分布のハイパーパラメータを推定することができる。紙面の都合上、詳細は割愛する。

4. 評価実験

4.1 実験データ

評価実験では、実店舗における購買ログおよび自宅での TV コマーシャル視聴ログからなるシングルソースデータを用いてモデルの妥当性を検証する。本データは、株式会社インテージのスキャンパネルデータ (i-SSP*1 および SCI*2) である。ここで、購買および TV コマーシャル視聴ログは共通するユーザ ID により紐付けられている。観測期間は 2013 年 1 月 1 日から 2013 年 12 月 31 日である。実験では、2013 年 1 月 1 日から 2013 年 11 月 30 日までの期間に行われた購買および視聴ログをモデルの学習のために用いる学習データとし、2013 年 12 月 1 日から 2013 年 12 月 31 日までの期間に行われた購買

表 2: データセットの基本的な統計量

| 商品カテゴリ | 購買数 | 視聴数 | ユーザ数 | 企業数 | 商品数 |
|--------|--------|--------|------|-----|-----|
| 缶コーヒー | 46,818 | 98,865 | 743 | 11 | 40 |

表 3: テストデータの対数尤度

| 商品カテゴリ | SCPP | Proposed 1 | Proposed 2 |
|--------|----------------------|----------------------|----------------------|
| 缶コーヒー | -1.738×10^4 | -1.736×10^4 | -1.681×10^4 |

および視聴ログをモデルの妥当性を検証するためのテストデータとした。また、今回は商品カテゴリとして缶コーヒーのデータを用いて実験を行った。ここで、購買回数が 5 回以上のユーザのみを用いた。データセットについての基本的な統計量を表 2 に示す。

4.2 実験結果

テストデータを用いて、時刻とユーザの対数尤度を以下の式に従って計算し、比較することによって提案モデルの妥当性を検証した。

$$L = \sum_{i \in I} \left(-\int_T^{T'} \sum_{u \in U} \lambda_i(t, u) dt \right) \sum_{(t,u) \in D_i^{\text{test}}} \lambda_i(t, u) \quad (7)$$

ここで、テスト期間は $T' - T$ であり、この期間のアイテム i の購買データを D_i^{test} とした。それぞれのモデルにおけるテストデータの対数尤度を表 3 に示す。比較手法として Shared Cascade Poisson Process (SCPP) [3] を用いた。SCPP は、他者の選択の影響のみを考慮してユーザの選択行動をモデル化したものである。提案手法は、個人の嗜好および他者の影響を考慮したモデルを Proposed 1 と表し、個人の嗜好、他者の影響および TV コマーシャルを通した企業の影響を考慮したモデルを Proposed 2 と表す。表 3 より、Proposed 2 におけるテストデータの対数尤度が SCPP に比べて高いことが分かる。これにより、個人の嗜好および TV コマーシャルを通した企業の影響を考慮できる提案モデルが、ユーザの選択行動を説明するモデルとして、より妥当なものであると言える。

次に、提案モデルにおいて推定された商品の人気度、ユーザの影響力および企業の影響力を表すパラメータを図 1 に示す。ここで、商品、ユーザおよび企業をパラメータの値が大きい順にそれぞれ並び替えた。図 1(a) より、人気度が高い商品が少数存在することが分かる。提案モデルでは、他者の影響および TV コマーシャルを通した企業の影響を考慮しながら個人の嗜好を推定する。そのため、社会的な影響を分離した上で、より純度の高い商品の人気度を推定可能であると考えられる。また、図 1(b) および図 1(c) より、影響力の大きいユーザおよび企業が少数存在しており、大部分は影響が小さいことが分かる。このような結果を用いれば、TV コマーシャルの効果が大きい企業を抽出したり、トレンドの先駆けとなるような影響力の大きいユーザを抽出することも可能である。これらはマーケティングが広告の最適化を行う際に有用な知見であると言える。

図 1(c) に赤色で示した企業について、学習期間における各購買が三つの要因のうちどの要因で生じたかを提案モデルにより推定した結果を用いて、各週毎の購買を三色で塗り分けたグラフを図 2 に示す。青は個人の嗜好、赤は他者の影響、緑は TV コマーシャルを通した企業の影響による購買をそれぞれ表す。図 2 より、全学習期間において、個人の嗜好による購買がトレンドに依らず一定量存在すると推定された。また、学習期

*1 インテージシングルソースパネル

*2 全国消費者パネル調査

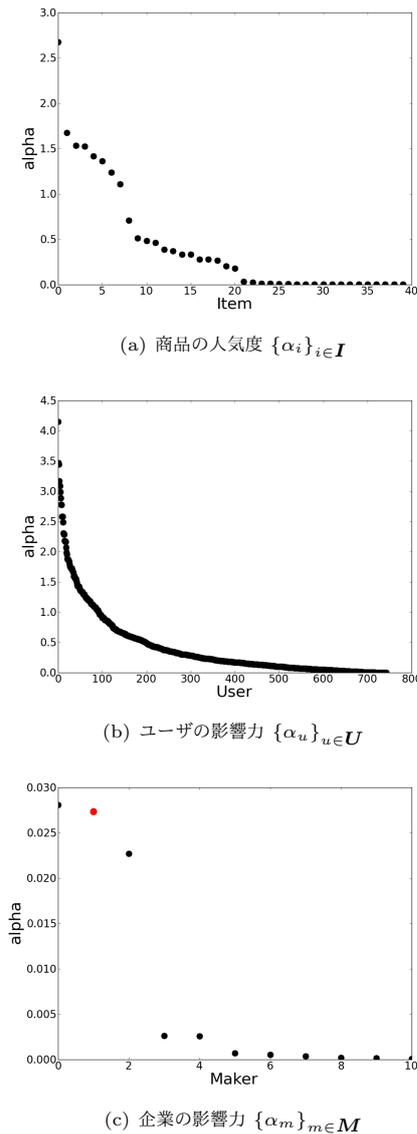


図 1: 提案モデルにより推定されたパラメータ。商品、ユーザおよび企業をパラメータの値が大きい順にそれぞれ並び替えた。

間の後半では、購買量の盛り上がりが存在し、そのときに TV コマーシャルの影響があったと推定された。実際、この期間には、該当企業の新商品発売および商品リニューアルによる TV コマーシャルが打たれており、それによる効果が抽出されたと考えられる。さらに、提案モデルでは、どのようなユーザが TV コマーシャルの影響によって購買を行いやすいか、ということも分析可能である。紙面の都合上、分析結果は割愛する。

5. おわりに

本稿では、ユーザの選択行動が、個人の嗜好、他者の影響、TV コマーシャルを通じた企業の影響の三つの要因によって起こると仮定したモデルを提案した。評価実験では、購買ログおよび TV コマーシャル視聴ログからなるシングルソースデータを用いてモデルの妥当性について検証を行った。テストデータの対数尤度を指標として、提案モデルが従来モデルに比べて、ユーザの選択行動を説明するモデルとして、より妥当であることを確認した。また、提案モデルにおいて推定された商品の人気度、ユーザの影響力および企業の影響力を表すパラメー

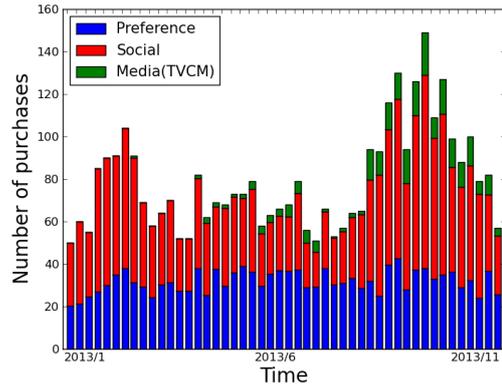


図 2: 提案モデルによって推定された購買要因の分析結果。ある企業における各週毎の購買を三つの要因についてグラフを塗り分けた。青は個人の嗜好、赤は他者の影響、緑は TV コマーシャルを通じた企業の影響による購買をそれぞれ表す。

タの分析を行った。これにより、人気度が高い商品および影響力の大きいユーザや企業が少数存在しているという推定結果を得た。さらに、新商品発売や商品リニューアルに伴った TV コマーシャルの効果を抽出することができた。このような知見は、マーケターが広告の最適化を行う際に有用なものであると考えられる。今後は、商品の季節性やユーザの購買行動の周期性などについても考慮できるようにさらなるモデルの検討を行う予定である。

参考文献

- [1] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*, Chapman and Hall, 1996,
- [2] M. Gomez-Rodriguez, D. Balduzzi, and B. Scholkopf, “Uncovering the temporal dynamics of diffusion networks,” In *ICML’11*, pp. 561–568, 2011.
- [3] T. Iwata, A. Shah, and Z. Ghahramani, “Discovering latent influence in online social activities via shared cascade Poisson processes,” In *KDD’13*, pp. 266–274, 2013.
- [4] D. Kempe, J. Kleinberg, and E. Tardos, “Maximizing the spread of influence through a social network,” In *KDD’03*, pp. 137–146, 2003.
- [5] J. F. C. Kingman, *Poisson processes*, vol. 3, Oxford University Press, USA, 1993. p
- [6] T. Kurashima, T. Iwata, N. Takaya, and H. Sawada, “Probabilistic latent network visualization: Inferring and embedding diffusion networks,” In *KDD’14*, pp. 1236–1245, 2014.
- [7] J. Leskovec, L. A. Adamic, and B. A. Huberman, “The dynamics of viral marketing,” In *Proceedings of the 7th ACM Conference on Electronic Commerce*, pp. 228–237, 2006.
- [8] W. Lu, S. Ioannidis, S. Bhaget, and L. Lakshmanan, “Optimal recommendations under attraction, aversion, and social influence,” In *KDD’14*, pp. 811–820, 2014.
- [9] 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*, pp. 727–736, 2013.
- [10] X. Song, Y. Chi, K. Hino, and B. L. Tseng, “Information flow modeling based on diffusion rate for prediction and ranking,” In *WWW’07*, pp. 191–200, 2007.