

## A Study on Time Window Topic Model for Analyzing Customer Browsing Behavior

ITO Fumiyo

### 1. はじめに

現在、EC サイトを初めとした様々なサービスがインターネット上で展開されており、各企業は購買履歴に加え、Web ページの閲覧履歴などユーザの行動に関するより詳細なデータを取得することが可能になっている。そのため、閲覧履歴を活用することにより、ユーザの嗜好をより詳細に分析し、効果的な施策を実施することが実務上望まれている。EC サイト上の購買においては、ユーザは Web ページを閲覧する中で、徐々に興味の対象を絞っていき、最終的にある商品を購入するという行動が想定される。しかしながら、EC サイトにおける購買率は高々数%であることが一般的であり、興味の対象を絞り込むことができず購買に至らないケースが大多数を占める。このように、EC サイト上にアクセスしたユーザの様々なセッションでは、徐々に興味の対象の絞り込みが行われていく少数のユーザ（以下、収束ユーザ）と、興味を絞り込むことができず様々な商品を閲覧している状態に留まる多数のユーザ（以下、未収束ユーザ）が混在していることが考えられる。そのため、各ユーザの興味の変化を分析し、興味の変動具合に基づき適切なタイミングで施策を実施することが、施策の効果を高める上で重要であると考えられる。以上により、本研究では Web サイトの閲覧履歴データを用いて、時間と共に変化するユーザの興味をモデル化し、分析可能とする手法の提案を目的とする。

従来より、顧客の嗜好は多様であり、観測可能な購買や閲覧の背後には複数の潜在的な状態が存在していることが考えられることから、観測可能な状態の背後に潜在的な状態（以下、トピック）を仮定したトピックモデルや隠れマルコフモデルが広く活用されてきた。このうち、時間と共に変化するユーザの嗜好をモデル化することが可能なトピックモデルとして、購買履歴データを対象とした Topic Tracking Model（以下、TTM）[1] が提案されている。ここで、トピックモデルでは、ユーザの嗜好を表現するトピック分布と、各トピックにおける商品やサイトの購買または閲覧確率であるアイテム分布を同時に推定し、これらの分布を用いてユーザの興味を分析することが可能である。TTM では、トピック分布とアイテム分布の両方に時系列性を仮定しており、各時刻において、これらの分布のパラメータを同時に推定する。このモデルは、比較的長い購買履歴データを想定し提案された手法であり 1 つのセッション内の閲覧系列のように、数十ページ程度のサイトの遷移の中でユーザの興味が変わっていくという状況を考慮できていない。そのため、TTM を閲覧履歴データに適用した際には、各時刻で観測されるデータ長が

十分に存在している状態を除きトピック分布とアイテム分布を同時に推定することは困難であり、ユーザの興味の収束状況を検出できない恐れがある。

そこで、本研究では、ユーザの興味を表現するトピック分布と各トピックの特徴を表現する分布であるアイテム分布を段階的に推定するトピックモデル、時間窓トピックモデル（Time Window Topic Model: TWTM）を提案する。提案モデルでは、各時刻におけるデータが比較的少数な場合でもパラメータの安定的な推定を可能とするために、始めに全時刻のデータを用いてアイテム分布の推定を行う。この際、Web サイトの閲覧においては、ユーザは連続する一定期間内の閲覧系列（以下、ウィンドウ）の中では類似した興味に従ってページを閲覧するという行動に着目し、ウィンドウごとに潜在トピックを割り当てを行う。この後に、推定したアイテム分布のパラメータを固定したもとの、時刻ごとにユーザのトピック分布の推定を行う。提案手法により、収束ユーザと未収束ユーザが混在し、ユーザの興味が徐々に移り変わるという複雑な状況が仮定される閲覧履歴データにおいても、安定したパラメータ推定を行うことができ、収束ユーザの興味の変化を明らかにすることが可能となる。これによって、各ユーザの興味の変動具合に基づいた施策の実施による施策効果の向上が期待される。本研究では、実データに対して提案手法を適用し、その有効性を検証する。

### 2. 準備

#### 2.1. 関連研究

購買履歴や閲覧履歴データから顧客の興味を分析するために、観測可能な購買や閲覧の背後に複数の潜在的な状態（トピック）の存在を仮定したトピックモデルや隠れマルコフモデルを起点にして様々な手法が提案されている。

代表的なトピックモデルである、Latent Dirichlet Allocation（以下、LDA）[2] はユーザの嗜好や興味をトピック混合割合であるトピック分布として解釈することが可能であり、購買履歴データを初めとした様々なデータに対して適用と拡張がなされている。さらにこの LDA から発展し、時系列的な特徴を捉えるために文書データを対象として、Dynamic Topic Model[3] に代表される時系列トピックモデルが提案されている。これらの手法は、元来、文書データを前提としており各時刻において 1 文書ずつ生成されることを想定している。そのため、これらの時系列トピックモデルを購買履歴や閲覧履歴データに適用し、ユーザの嗜好の変化を分析することは不可能である。これらの手法に対し、1 人のユーザの嗜好の変化をトピック分布の変化としてモデル化するために

提案された手法として TTM がある。TTM を購買履歴に適用することにより、ユーザの嗜好の変化について解析することが可能である。

一方、購買履歴や閲覧履歴データを対象として潜在マルコフモデル (以下、HMM) を起点とした研究も広く行われている。HMM では、トピック間の遷移のしやすさについて解析することが可能であるが、ユーザの興味の変化を分析するためには、全ての購買や閲覧に対するトピックの変化を分析することが必要となる。これに対し、HMM にトピックモデルの考え方を援用し、トピック間の遷移にマルコフ性を仮定しながら各ユーザのトピック分布を推定可能な手法として、Hidden Topic Markov Model[4] (以下、HTMM) が知られている。保戸田ら [5] は、この HTMM を閲覧履歴データに適用するために拡張したモデルを提案し、実データに適用して有効性を示している。このモデルでは、トピック間の遷移割合とユーザのトピック分布を推定可能であるが、トピック分布が時刻によって変化することは想定されていない。

## 2.2. 変数の定義

本研究は、Web サイト内の閲覧履歴データを対象としているため、これを前提として変数の定義を行う。Web サイトのページ集合を  $\mathcal{X} = \{b_d\}_{d=1}^D$  とし、ユーザ  $a_u$  ( $u \in \mathcal{U} := \{1, 2, \dots, U\}$ ) の閲覧ページ数を  $M_u$ 、ユーザ  $a_u$  の  $i$  番目の閲覧ページを  $x_i^u \in \mathcal{X}$  ( $i \in \mathcal{M}_u := \{1, \dots, M_u\}$ ) と表現する。このとき、ユーザ  $a_u$  の閲覧系列を  $\mathbf{X}_u = x_1^u, x_2^u, \dots, x_{M_u}^u$  と表すこととする。さらに、ある時刻  $t$  ( $t \in \mathcal{T} := \{1, \dots, T\}$ ) におけるユーザ  $a_u$  の閲覧ページ数を  $M_{t,u}$ 、時刻  $t$  における  $i$  番目の閲覧ページを  $x_i^{t,u} \in \mathcal{X}$  ( $i \in \mathcal{M}_{t,u} := \{1, \dots, M_{t,u}\}$ ) とし、このときの閲覧系列を  $\mathbf{X}_{t,u} = x_1^{t,u}, \dots, x_{M_{t,u}}^{t,u}$  とする。ただし、 $\mathbf{X}_u = \mathbf{X}_{1,u}, \dots, \mathbf{X}_{T,u}$  である。また、潜在トピック数を  $K$ 、潜在トピック集合を  $\mathcal{Z} := \{1, \dots, K\}$  とする。

## 2.3. 従来手法 (Topic Tracking Model)

Topic Tracking Model (TTM) は、購買履歴データを想定しユーザの嗜好の変化とアイテムの流行を解析するために提案された手法である。TTM においてユーザ  $a_u$  の時刻  $t$  における閲覧の出現確率は、以下の式 (1) で与えられる。

$$P(\mathbf{X}_{t,u} | \boldsymbol{\theta}_{t,u}, \boldsymbol{\Phi}_t) = \prod_{i=1}^{M_{t,u}} \sum_{k=1}^K \{P(z_{t,u,m} = k | \boldsymbol{\theta}_{t,u}) \times P(x_i^{t,u} | z_{t,u,m} = k, \boldsymbol{\phi}_k)\}. \quad (1)$$

ここで、 $\boldsymbol{\theta}_{t,u} = (\theta_{t,u,1}, \dots, \theta_{t,u,K}) \in \mathbb{R}^K$ 、 $\boldsymbol{\phi}_{t,k} = (\phi_{t,k,1}, \dots, \phi_{t,k,D})$  であり、 $\boldsymbol{\theta}_{t,u}$  は時刻  $t$  におけるユーザ  $a_u$  のトピック分布のパラメータ、 $\boldsymbol{\phi}_{t,k}$  は時刻  $t$  におけるトピック  $k \in \mathcal{Z}$  のアイテム分布のパラメータである。ただし、 $\theta_{t,u,k} = P(z_{t,u,m} = k | \boldsymbol{\theta}_{t,u})$ 、 $\phi_{t,k,x_m^{t,u}} = P(x_m^{t,u} = k, \boldsymbol{\phi}_k)$  である。また、 $\boldsymbol{\Phi}_t = (\boldsymbol{\phi}_{t,1}, \dots, \boldsymbol{\phi}_{t,K})^\top \in \mathbb{R}^{K \times D}$  であり、 $z_{t,u,m}$  は時刻  $t$  におけるユーザ  $a_u$  の  $m$  番目の閲覧に対して割り当てられたトピックである。

さらに、TTM では各ユーザのトピック分布とアイテム分布に対して 1 次のマルコフ性を仮定しており、 $\boldsymbol{\theta}_{t,u}$ 、 $\boldsymbol{\phi}_{t,k}$  の

事前分布を以下の式 (2),(3) によってモデル化する。

$$P(\boldsymbol{\theta}_{t,u} | \hat{\boldsymbol{\theta}}_{t-1,u}, \alpha_{t,u}) \propto \prod_{k=1}^K \theta_{t,u,k}^{\alpha_{t,u} \hat{\theta}_{t-1,u,k}^{-1}}. \quad (2)$$

$$P(\boldsymbol{\phi}_{t,k} | \hat{\boldsymbol{\phi}}_{t-1,k}, \beta_{t,k}) \propto \prod_{d=1}^D \phi_{t,k,d}^{\beta_{t,k} \hat{\phi}_{t-1,k,d}^{-1}}. \quad (3)$$

ただし、 $\alpha_{t,u}$  と  $\beta_{t,k}$  は、分布の形状を制御するハイパーパラメータであり、 $\hat{\boldsymbol{\theta}}_{t-1,u}$  と  $\hat{\boldsymbol{\phi}}_{t-1,k}$  は、それぞれ  $t-1$  時刻において推定されたトピック分布とアイテム分布のパラメータである。

## 3. 提案手法

### 3.1. 概要

閲覧履歴データから、ユーザの興味の変化をトピック分布としてモデル化するためには、TTM のような時系列トピックモデルは有効であると考えられる。しかしながら、TTM は比較的長い閲覧系列、もしくは各時点で観測されるデータ長が十分に大きいことを前提とし、ユーザの興味も緩やかに変化することを想定している。一方、閲覧履歴データにおいてはユーザの興味は数〜数十単位の閲覧の中で切り替わることが想定され、従来のモデルではこのような状況においてはトピック分布とアイテム分布のパラメータの推定性能が悪化する恐れがある。

そこで、本研究ではユーザの興味が短期間で切り替わる閲覧履歴データにおいても高精度なパラメータ推定を可能とするために、アイテム分布とトピック分布を 2 段階で推定する手法を提案する。提案手法では、各時刻におけるデータ量に依らず安定したパラメータ推定を可能にするために、各ユーザのトピック分布のみに時系列性を仮定し、STEP.1 において全時刻の閲覧履歴データを用いてアイテム分布の推定を行う。ここで、アイテム分布の際に LDA などのトピックモデルを適用することも考えられるが、LDA ではユーザの各閲覧ごとに潜在トピックの割り当てを行うため、閲覧の中の局所的なユーザの興味の偏りなどを考慮することができない。そのため、STEP.1 においてはウィンドウごとに潜在トピックの割り当てを行うことによりこの問題を解決する。そして、次の STEP.2 において、STEP.1 で推定したアイテム分布のパラメータを固定し、ユーザのトピック分布の推定のみを行う。このように段階的にパラメータを学習することによって、数〜数十単位の閲覧の中でユーザの興味が様々切り替わるといった状況においてもユーザの興味の変化をトピック分布として適切に捉えることが可能となる。

### 3.2. STEP.1 アイテム分布の推定

アイテム分布の推定の際には、同一ウィンドウ内のページは同じトピックのもとで共起していると仮定し、パラメータの推定を行う。ここで、ユーザ  $a_u$  の  $m$  番目のウィンドウ内部分系列  $\mathbf{w}_{u,m}$  ( $m \in \mathcal{W}_u := \{1, 2, \dots, M_u - c + 1\}$ ) は、ウィンドウサイズを  $c$  として、以下の式 (4) によって定義される。

$$\mathbf{w}_{u,m} = x_m^u, x_{m+1}^u, \dots, x_{m+c-1}^u. \quad (4)$$

そして、ユーザ  $a_u$  の閲覧系列は、式 (5) のように表現される。

$$\mathbf{X}_u = \mathbf{w}_{u,1}, \mathbf{w}_{u,2}, \dots, \mathbf{w}_{u,M_u-c+1}. \quad (5)$$

このとき、ユーザ  $a_u$  の閲覧系列  $\mathbf{X}_u$  の出現確率は、以下の式 (6) で与えられる。

$$P(\mathbf{X}_u | \boldsymbol{\theta}, \boldsymbol{\Phi}) = \prod_{m=1}^{M_u-c+1} \sum_{k=1}^K \{P(z_{u,m} = k | \boldsymbol{\theta}) \times \prod_{i=m}^{m+c-1} P(x_i^u | z_{u,m} = k, \boldsymbol{\phi}_k)\}. \quad (6)$$

ただし、 $z_{u,m}$  はユーザ  $a_u$  の  $m$  番目の閲覧の潜在トピックである。また、 $\boldsymbol{\theta} = (\theta_1, \dots, \theta_K)$  はトピック分布のパラメータ、 $\theta_k = P(z_{u,m} = k | \boldsymbol{\theta})$ 、 $\boldsymbol{\phi}_k = (\phi_{k,1}, \dots, \phi_{k,D})$  は、トピック  $k$  のアイテム分布のパラメータであり、 $\phi_{k,x_i^u} = P(x_i^u | z_{u,m} = k, \boldsymbol{\phi}_k)$  とする ( $i = m, \dots, m+c-1$ )。ただし、 $\boldsymbol{\Phi} = (\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_K)^\top \in \mathbb{R}^{K \times D}$  とする。また、トピック分布  $\boldsymbol{\theta}$  とアイテム分布  $\boldsymbol{\phi}_k$  は、式 (7), (8) のディリクレ分布から生成されるものとする。

$$p(\boldsymbol{\theta} | \boldsymbol{\alpha}) \propto \prod_{k=1}^K \theta_k^{\alpha_k - 1} = \text{Dirichlet}(\boldsymbol{\alpha}) \quad (7)$$

$$p(\boldsymbol{\phi}_k | \boldsymbol{\beta}_k) \propto \prod_{d=1}^D \phi_{k,d}^{\beta_k - 1} = \text{Dirichlet}(\boldsymbol{\beta}_k) \quad (8)$$

ただし、 $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K) \in \mathbb{R}^K$ 、 $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_K)$ 、 $\boldsymbol{\beta}_k = (\beta_{k,1}, \beta_{k,2}, \dots, \beta_{k,D}) \in \mathbb{R}^D$  である。提案モデルの学習には、ギブスサンプリングによる潜在トピックの割り当てと、不動点反復法による事前分布のパラメータ  $\boldsymbol{\alpha}$ 、 $\boldsymbol{\beta}$  の更新を繰り返す確率的 EM アルゴリズム [6] を用いる。十分な反復を行った後、アイテム分布は以下の式 (9) によって求められる。

$$\hat{\phi}_{k,d} = \frac{N_{k,d} + \beta_k}{N_k + D\beta_k} \quad (9)$$

ただし、 $N_{k,d}$  は、 $d$  番目のサイトにトピック  $k$  が割り当てられた回数であり、 $N_k = \sum_d N_{k,d}$  である。

### 3.3. STEP.2 トピック分布の推定

STEP.2 では、式 (9) で求めたアイテム分布のパラメータを固定したもとの、ユーザのトピック分布の推定を行う。ここでは、ユーザの興味は時刻  $t$  ごとに変化していることを仮定し、時刻ごとにトピック分布を推定する。また、TTM と同様に、式 (3) によって、ユーザの興味の連続的な変化をモデル化する。このとき、ある時刻  $t$  におけるユーザ  $a_u$  の閲覧系列  $\mathbf{X}_{t,u}$  の出現確率は以下の式 (10) で与えられる。ただし、 $\hat{\boldsymbol{\Phi}} = (\hat{\boldsymbol{\phi}}_1, \dots, \hat{\boldsymbol{\phi}}_K)$  であり、 $\hat{\boldsymbol{\phi}}_k = (\hat{\phi}_{k,1}, \dots, \hat{\phi}_{k,D})$  は式 (9) によって推定されたアイテム分布のパラメータである。

$$P(\mathbf{X}_{t,u} | \boldsymbol{\theta}_{t,u}, \hat{\boldsymbol{\Phi}}) = \prod_{i=1}^{M_{t,u}} \sum_{k=1}^K \{P(z_{t,u,m} = k | \boldsymbol{\theta}_{t,u}) \times P(x_i^{t,u} | z_{t,u,m} = k, \hat{\boldsymbol{\phi}}_k)\}. \quad (10)$$

STEP.1 と同様に、STEP.2 においても確率的 EM アルゴリズムを用いた学習を行う。学習の反復後、時刻  $t$  におけるユーザ  $a_u$  のトピック分布は以下の式 (11) によって推定される。

$$\hat{\theta}_{t,u,k} = \frac{N_{t,u,k} + \alpha_{t,u} \hat{\theta}_{t-1,u,k}}{N_{t,u} + \alpha_{t,u}} \quad (11)$$

ただし、 $N_{t,u,k}$  は、時刻  $t$  においてユーザ  $a_u$  にトピック  $k$  が割り当てられた回数であり、 $N_{t,u} = \sum_k N_{t,u,k}$  である。

## 4. 実データ分析

本稿では、提案モデルの有効性を検証するために実際の閲覧履歴データに提案モデルを適用し、従来手法 (TTM) との性能の比較とユーザの興味の変化に関する分析を行う。

### 4.1. 分析条件

ここでは、株式会社ヴァリユーズ提供の楽天市場の閲覧履歴データを用いる。対象データのデータ取得期間は 2019 年 2 月 1 日–4 月 31 日であり、4.2 節においては、1 期間を 1 週間\*として設定し、全期間において閲覧が存在したユーザを抽出した。その結果抽出されたユーザ数は  $U = 47$  であり、全 Web ページ数  $D = 13,527$ 、総閲覧回数は 29,888 であった。このとき、ウィンドウサイズは  $C = 5$  として設定した。また、4.3 節においては、1 期間を 1 日として設定し、1 期間あたりの閲覧回数が 10 回以上存在する期間が 30 以上あるユーザを抽出した。その結果、全ユーザ数  $U = 178$ 、全 Web ページ数  $D = 5,670$ 、総閲覧回数は 301,916 であり、ウィンドウサイズは  $C = 10$  とした<sup>†</sup>。また、これらの閲覧のうち各ユーザについて最後の 1 閲覧をテストデータとして抽出した。ハイパーパラメータの初期値は、STEP.1 では  $\alpha = 0.1$ 、 $\beta_k = 0.1$  ( $k = 1, \dots, K$ )、STEP.2 では  $\alpha_{1,u} = 10.0$  ( $u = 1, \dots, U$ ) とした。

### 4.2. 提案モデルの性能評価

提案モデルの性能を評価するために、TTM とテストデータに対する Perplexity (以下、PPL) の比較を行なった。ここで、PPL とはトピックモデルの性能を評価する指標であり、小さいほどモデルの性能が高いことを意味する。PPL は式 (12) で求められる。

$$\text{PPL} = \prod_t \prod_u P(\mathbf{X}_{t,u} | \boldsymbol{\theta}_{t,u}, \boldsymbol{\Phi}_t)^{-N_{t,u}} \quad (12)$$

ここで、トピック数  $K$  を 10 から 50 まで変化させたときの提案手法と TTM のテストデータ PPL の値を表 1 に示す。

表 1 の結果より、いずれのトピック数においても提案手法は TTM よりも低い PPL を示しており、提案手法による段階的なパラメータの推定が有効であったといえる。

### 4.3. 提案モデルを用いた分析

実データセットに対して提案手法を適用し、ユーザの興味に関する分析と実応用について検討を行う。ここで、モデルの学習によって推定したトピック分布の時刻ごとの変化を図 1 に示す。分析のために、(a) 興味が収束して購買に至った

\*TTM では全ての期間において閲覧履歴が存在しないと学習を行うことができないため、この条件を満たすように 1 週間を 1 期間とした。

<sup>†</sup>ウィンドウサイズ  $C$  は、 $C = 2, 3, 4, 5, 7, 10$  と変化させたときに最もテストデータ PPL が低かったものを採用した。

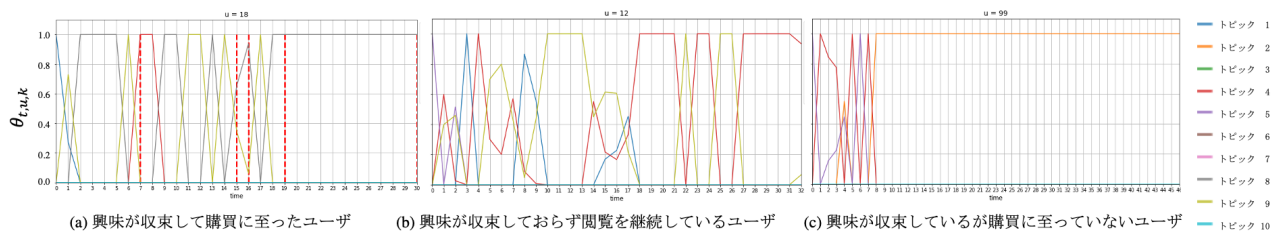


図 1: 各ユーザのトピック分布の変化

表 1: テストデータ PPL の比較

$K$	TWTM	TTM
10	<b>2702.13</b>	63211.64
20	<b>5135.67</b>	13627.13
30	<b>3192.74</b>	19154.29
40	<b>3300.28</b>	6592.41
50	<b>3676.83</b>	8596.18

ユーザ, (b) 興味が収束しておらず閲覧を継続しているユーザ, (c) 興味が収束しているが購買に至っていないユーザ, を抽出した。ただし, 図の縦軸は, 各時刻における所属確率  $\theta_{t,u,k}$  を表し, 赤色の波線は購買の発生を意味する。

(a) に示すユーザは, 時刻  $t = 18$  以降トピック 8 への所属確率が大きくなり, 時刻  $t = 19$  において購買を行なっている。さらに, 購買後も同一のトピックに対する所属確率が高い状態が継続していることから, このユーザはトピック 8 へ興味が収束しており, このトピックへの所属確率が高い店舗のクーポンの発行やメルマガの送付により追加購買を促すことが可能であると考えられる。一方で, (b) に示すユーザは時刻ごとのトピック分布の変動が大きく, ユーザの興味の対象が十分に絞られていないことが考えられる。このようなユーザは, 購買に至る可能性が低く, クーポンの発行などの施策の効果は十分に見込めないことが考えられるため, 施策実施の優先度が低いユーザ群である。また, (c) は時刻  $t = 8$  以降にトピック 2 への興味の高まりと収束が確認されるが, 購買には至っていないユーザである。このようなユーザについては, 十分に興味の対象が絞れていることから, 直近の閲覧店舗のクーポンの発行や商品の追加情報の提示などによって, 購買を誘発できる可能性がある。

以上のように, 提案手法を用いることによってユーザの興味の変化を明らかにし, 興味の収束度合いに基づくマーケティング施策の立案が可能であるといえる。

## 5. 考察

提案手法で推定されたトピック分布を用いることにより, ユーザの興味の収束度合いを把握することが可能となり, これに基づいて施策実施の優先度が高いユーザを抽出したり, 興味の高まったタイミングで施策を実施するなど, 施策の実施タイミングの最適化が可能になると考えられる。

提案手法を用いて施策に関する意思決定を行う際には, 意思決定の透明性を保つため, 解釈性の高いトピックを形成可

能なトピック数, ウィンドウサイズの決定が重要である。これらの決定のためには, PPL や Coherence などのトピックモデルの性能を評価する指標を用いた定量的評価に加え, トピック数やウィンドウサイズを変化させて得られたトピックを知識を持った分析者が解釈するなど, 定性的な評価も必要と考えられる。

## 6. 結論と今後の課題

本研究では, Web サイトの閲覧履歴データを対象として, ユーザの興味の収束度合いを明らかにするために, ユーザの興味が短期間において切り替わるという状況においても安定してモデルのパラメータを推定可能な手法を提案した。提案手法では, アイテム分布とトピック分布を個別に推定し, さらに, アイテム分布の推定の際にウィンドウごとに潜在トピックを割り当てることにより, 従来手法と比較して高精度なパラメータ推定が可能となった。最後に, 実データセットを用いた分析により提案手法の実応用性を示した。今後の課題として, トピック分布を用いた収束ユーザの抽出方法の確立や, 出現頻度の低いサイトを考慮したモデル化が挙げられる。

## 謝辞

本研究は, 株式会社ヴァリユーズと早稲田大学の共同研究であり, データ提供元である株式会社ヴァリユーズの皆様へ深く感謝致します。

## 参考文献

- [1] Tomoharu Iwata, Shinji Watanabe, Takeshi Yamada, and Naonori Ueda. Topic tracking model for analyzing consumer purchase behavior. In *Twenty-First International Joint Conference on Artificial Intelligence*, 2009.
- [2] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *The Journal of Machine Learning Research*, Vol. 3, pp. 993–1022, 2003.
- [3] David M Blei and John D Lafferty. Dynamic topic models. In *Proceedings of the 23rd International Conference on Machine Learning*, pp. 113–120, 2006.
- [4] Amit Gruber, Yair Weiss, and Michal Rosen-Zvi. Hidden topic markov models. In *Artificial intelligence and statistics*, pp. 163–170. PMLR, 2007.
- [5] Mio Hotoda, Gendo Kumoi, and Masayuki Goto. A study on customer purchase behavior analysis based on hidden topic markov models. *Industrial Engineering & Management Systems*, Vol. 20, No. 1, pp. 48–60, 2021.
- [6] Hanna M Wallach. Topic modeling: beyond bag-of-words. In *Proceedings of the 23rd International Conference on Machine Learning*, pp. 977–984, 2006.