

ソーシャルメディアのアイテム群に対する アテンションダイナミクスの学習

松谷 貫司¹ 木村 昌弘^{2,a)}

受付日 2016年11月11日, 再受付日 2016年12月31日,
採録日 2017年1月22日

概要: ソーシャルメディアでは, ユーザは様々な情報を手軽に発信および共有することができ, ユーザが投稿したアイテムは, 他のユーザからの共有や賛意メッセージなどのアテンションの数が増えるにつれてポピュラリティを獲得していく. 最近 Shen ら [12] は, 個々のアイテムへのアテンションの到着過程の確率モデルとして *RPP* モデル (*Reinforced Poisson Process model*) を提案し, それが既存のポピュラリティ予測法よりも精度が高いことを示した. 本論文では, 対象とするソーシャルメディアのアイテム群全体に対するアテンションの到着過程の確率モデルとして, ディリクレ過程と *RPP* モデルを融合した *DPM-RPP* モデルを提案し, アテンションダイナミクスの観点からそれらアイテム群の関係の分析を目指す. 我々は, 観測データに基づく *DPM-RPP* モデルの効率的な学習法および, *DPM-RPP* モデルによる各アイテムの将来ポピュラリティの予測法を与える. 料理レシピ共有サイトの実データを用いた実験により, *DPM-RPP* モデルはポピュラリティ予測において *RPP* モデルを含む従来モデルよりも精度が高いことを示す. また, アテンションダイナミクスの観点における料理レシピ群のコミュニティ構造を明らかにする.

キーワード: アテンションダイナミクス, *RPP* モデル, ディリクレ過程混合モデル

Learning Attention Dynamics for Items in Social Media

KANJI MATSUTANI¹ MASAHIRO KIMURA^{2,a)}

Received: November 11, 2016, Revised: December 31, 2016,
Accepted: January 22, 2017

Abstract: Social media sites allow users easily to post and share online-items. Such items gain their popularity through the amount of attention received. Recently, Shen et al. [12] presented a probabilistic model, called the *RPP* (*Reinforced Poisson Process*) model, for the arrival process of attention to an individual item, and showed that this model outperforms existing popularity prediction methods. In this paper, we aim at analyzing and mining the relation of items in a social media site from a perspective of attention dynamics, and by incorporating a Dirichlet process into the *RPP* model, we propose a probabilistic model, called the *DPM-RPP* model, for the arrival processes of attention to all the items involved. We present both an efficient method of learning the *DPM-RPP* model from the observed data and an effective method of predicting the future popularity of each item by the *DPM-RPP* model. Using real data from a cooking-recipe sharing site, we demonstrate that the *DPM-RPP* model outperforms conventional models including the *RPP* model for popularity prediction, and uncover the community structure of cooking-recipes in terms of attention dynamics.

Keywords: attention dynamics, Reinforced Poisson Process model, Dirichlet process mixture model

¹ 龍谷大学大学院理工学研究科電子情報学専攻
Division of Electronics and Informatics, Graduate School of
Science and Technology, Ryukoku University, Otsu, Shiga
520–2194, Japan

² 龍谷大学理工学部電子情報学科
Department of Electronics and Informatics, Faculty of Sci-
ence and Technology, Ryukoku University, Otsu, Shiga 520–
2194, Japan

1. はじめに

Facebook, Twitter, YouTube, @cosme, クックパッド
など, ソーシャルメディアは, Web 空間におけるコミュニ
ケーションの重要な場として進化し続けており, 人々の日

^{a)} kimura@rins.ryukoku.ac.jp

常生活に大きな影響を与えつつある。ソーシャルメディアでは、ニュース、アイデア、オピニオンなど様々な情報を手軽に発信できる。また、発信された情報は多くの人々に共有されることがあるので、意思決定や社会のトレンド形成にも影響を与えうる。したがって、ソーシャルメディアにおいてユーザが投稿したアイテムに対し、他のユーザから共有されたり賛意メッセージが送られたりするというアイテムへのアテンションが、どのような時間間隔で到着するかというアテンションダイナミクスのモデリングは、アイテムのポピュラリティ予測、トレンド予測、ユーザ行動の理解、マーケティング、意思決定などの観点から重要である [13]。

アイテムのポピュラリティ予測に関しては、一般に、時系列モデルを用いる研究や特徴量を導入して回帰問題や分類問題として扱う研究などがある [1], [4], [10], [13], [15]。アテンションダイナミクスは、Twitter でのリツイートによるツイートへのアテンション到着過程など、ソーシャルネットワーク上の情報拡散としてとらえられる場合がある。ソーシャルネットワーク上の情報拡散に関しては、複数の情報拡散サンプルが与えられたとき、背後にあるネットワーク構造を推定する問題 [5], [7] や、ネットワークデータと情報拡散の確率モデルが与えられたとき、指定された数のノード群で期待影響度が最大となるものを求める影響最大化問題 [3], [8] に対し、多くの研究がなされている。

最近 Shen ら [12] は、個々のアイテムへのアテンションダイナミクスの確率モデルとして、RPP モデル (Reinforced Poisson Process model) を提案した。RPP モデルは、平均ポピュラリティに関する時系列を対象とした一般的な従来手法と違って、個々のアテンションの到着過程に関する確率的生成モデルであり、アイテムの魅力を表すフィットネス、アイテムの新たなアテンション獲得能力に対する継時効果 (すなわち、エイジング効果) を表す時間緩和関数、および “rich-get-richer” 現象を表す強化関数という 3 つの構成要素からなる非一様ポアソン過程である。Shen らは、論文引用ダイナミクスのモデリングにおいて、ポピュラリティ予測性能が従来法よりも良いことを実データを用いた実験で示すことにより、RPP モデルの有用性を実証した。Gao ら [6] は、特に、Twitter でのリツイートによるツイートへのアテンションダイナミクスの場合に対して、RPP モデルを改良した ERPP モデル (Extended Reinforced Poisson Process model) を提案した。また、Zhao ら [16] は、各ユーザのフォロー数データを利用することにより、リツイート到着過程を Galton-Watson ツリー上の Hawkes 過程としてモデル化し、ツイートのポピュラリティを予測する手法 SEISMIC を提案した。しかしながら、これらの研究は、個々のアイテムを完全に独立にモデル化するというアプローチをとっており、対象とするソーシャルメディアのアイテム群におけるコミュニティ構造の抽出など、アイテ

ム群全体を対象とした関係の分析やマイニングに適用するには限界がある。特に Zhao らのモデルは、フォロー数データを必要とし、適用範囲にも限界があることに注意しておく。

本論文では、対象とするソーシャルメディアにおけるアイテム群全体に対して、各アイテムへのアテンション到着過程のより詳細なモデルの獲得と、エイジングタイプ (すなわち、新たなアテンション獲得に対する継時効果のタイプ) の観点からアイテム群におけるコミュニティ構造を抽出することを目指し、ディリクレ過程を融合して RPP モデルを拡張することにより、それらアイテム群全体のアテンションダイナミクスの確率モデル (DPM-RPP モデル) を提案する。さらに、観測データに基づく DPM-RPP モデルの効率的な学習法および、DPM-RPP モデルによる各アイテムの将来ポピュラリティの予測法を与える。料理レシピ共有サイトの実データを用いた実験により、提案する DPM-RPP モデルは、受け取るアテンションの総数による料理レシピのポピュラリティ予測において、RPP モデルと ERPP モデルよりも精度が高いことを示す。また、エイジングタイプの観点における料理レシピ群のコミュニティ構造を明らかにする。

本論文の構成は次のとおりである。2 章では、RPP モデルの拡張として DPM-RPP モデルを提案する。3 章では、観測データに基づく DPM-RPP モデルの学習法、DPM-RPP モデルによるアテンション総数の予測法および、コミュニティ抽出法を詳説する。4 章では、料理レシピ共有サイトの実データと人工データを用いた実験、および分析の結果を報告する。そして、5 章はまとめである。

2. 準備

対象とするソーシャルメディアに投稿されるアイテム全体の集合を D とし、正数 T を任意に固定する。任意のアイテム $d \in D$ に対して、 d が投稿されてから T 時間内に到着する d へのアテンションの総数を $n_d(T)$ とし、特に、 d が投稿されてから i 個目のアテンションが到着するまでの時間を t_i^d とする。ここに、 $i = 1, \dots, n_d(T)$ であり、 $0 \leq t_1^d \leq \dots \leq t_{n_d(T)}^d \leq T$ である。本論文では、任意のアイテム $d \in D$ への T 時間内のアテンションの到着過程 $\mathbf{t}^d(T) = \{t_i^d; i = 1, \dots, n_d(T)\}$ を同時にモデル化する確率的生成モデルを提案する。その準備として、本章においてまず、個々のアイテム d へのアテンション到着過程 $\mathbf{t}^d(T)$ をそれぞれ独立にモデル化する確率的生成モデルである 2 つの既存手法 RPP モデルと ERPP モデルについて詳しく述べる。

RPP モデルと ERPP モデルでは、各アイテム $d \in D$ へのアテンションの到着過程 $\mathbf{t}^d(T)$ はレート関数 $\lambda_d(t)$ の非一様ポアソン過程としてモデル化される。まず、RPP モデルの $\lambda_d(t)$ は、 d の魅力を表すフィットネス η_d , d

のエイジング効果を表す時間緩和関数 $g(t; \mu_d, \sigma_d)$ および, “rich-get-richer” メカニズムを表す強化関数 $r_d(t)$ の積 $\lambda_d(t) = \eta_d g(t; \mu_d, \sigma_d) r_d(t)$ で表される. ここに, $\eta_d > 0$ はパラメータ, $g(t; \mu_d, \sigma_d)$ は対数正規分布の密度関数

$$g(t; \mu_d, \sigma_d) = \frac{1}{\sqrt{2\pi\sigma_d^2}t} \exp\left\{-\frac{(\ln t - \mu_d)^2}{2\sigma_d^2}\right\}, \quad (1)$$

そして $r_d(t)$ は $t_{i-1}^d \leq t < t_i^d$ ならば

$$r_d(t) = i - 1 + \nu \quad (2)$$

である. ただし, $\mu_d \in \mathbb{R}$ と $\sigma_d > 0$ はパラメータであり, ν は事前に設定するハイパーパラメータである. 次に, ERPP モデルにおける $\lambda_d(t)$ は, RPP モデルと同様のフィットネス $\eta_d (> 0)$, 時間緩和関数 $\tilde{g}(t; \rho_d)$ および, 強化関数 $\tilde{r}(t; \zeta_d)$ の積 $\lambda_d(t) = \eta_d \tilde{g}(t; \rho_d) \tilde{r}(t; \zeta_d)$ で表される. ここに, $\tilde{g}(t; \rho_d)$ は冪関数

$$\tilde{g}(t; \rho_d) = t^{-\rho_d},$$

そして $\tilde{r}(t; \zeta_d)$ は $t_{i-1}^d \leq t < t_i^d$ ならば

$$\tilde{r}(t; \zeta_d) = \sum_{j=0}^{j=i} \exp\{-\zeta_d j\} + \nu$$

である. ただし, $\rho_d > 0$ と $\zeta_d > 0$ はパラメータであり, ν は RPP モデルと同様に事前に設定するハイパーパラメータである. これらの既存手法では, すべてのモデルパラメータの値は, 一般にアイテムごとに異なりうる. すなわち, 各アイテム $d \in D$ に対して, アテンションの到着過程 $\mathbf{t}^d(T)$ が完全に独立にモデル化されており, D 全体を対象とした関係の分析やマイニングに適用するには限界があることに注意する.

3. 提案モデル

我々は RPP モデルを拡張して, 任意のアイテム $d \in D$ への T 時間内のアテンションの到着過程 $\mathbf{t}^d(T) = \{t_i^d; i = 1, \dots, n_d(T)\}$ の確率的生成として, まず, RPP 混合モデルを提案し, さらに, それを無限混合モデルである DPM-RPP モデルへと拡張する.

3.1 RPP 混合モデル

対象とするソーシャルメディアに投稿されたアイテム群 D は完全に独立ではなく, なんらかの関係性を持ちうると考えられる. 我々は, アイテム独自の魅力を表すフィットネスと異なり, アイテムのエイジング効果を表す時間緩和関数はアイテムごとに完全に異なるのではなく, いくつかのタイプに分かれると仮定して, エイジングタイプに基づく D のコミュニティ構造を RPP モデルに組み込み, 各アイテム $d \in D$ のアテンションダイナミクスのより詳細なモデル化と, アイテム集合 D 全体の関係分析を目指す. 我々

は, 任意のアイテム $d \in D$ への T 時間内のアテンションの到着過程 $\mathbf{t}^d(T) = \{t_i^d; i = 1, \dots, n_d(T)\}$ を以下のようにモデル化する.

まず, D におけるエイジング効果を表す時間緩和関数を $f(t; \theta)$ とする^{*1}. ここに, θ はパラメータである. D には K 個のエイジングタイプ (コミュニティ) があると, 時間緩和関数 $f(t; \theta)$ における K 個のパラメータ値 $\theta_1, \dots, \theta_K$ をハイパーパラメータ τ を持つ θ の事前分布 $p(\theta | \tau)$ から

$$\theta_k | \tau \sim p(\theta | \tau) \quad (k = 1, \dots, K) \quad (3)$$

と生成する. また, K 次元ディリクレ分布から K 次元離散確率分布 $\boldsymbol{\pi} = (\pi_1, \dots, \pi_K)$ を

$$\boldsymbol{\pi} | \gamma \sim \text{Dirichlet}(\gamma/K, \dots, \gamma/K)$$

と生成する. $\gamma > 0$ はハイパーパラメータである.

次に, d のエイジングタイプ z_d を $\boldsymbol{\pi}$ が定義する多項分布から

$$z_d | \boldsymbol{\pi} \sim \text{Multinomial}(\boldsymbol{\pi})$$

と生成する. d のフィットネス η_d をパラメータが $\alpha > 0$ と $\beta > 0$ のガンマ分布 $\text{Gamma}(\alpha, \beta)$ から,

$$\eta_d | \alpha, \beta \sim \text{Gamma}(\alpha, \beta) \quad (4)$$

と生成する. ただし, α と β はハイパーパラメータである. d に対する強化関数 $r_d(t)$ を式 (2) で定義する. このとき, $\lambda_d(t)$ を

$$\lambda_d(t) = \eta_d f(t; \theta_{z_d}) r_d(t) \quad (5)$$

で定義し, レート関数 $\lambda_d(t)$ の非一様ポアソン過程により, d への T 時間内のアテンション到着過程 $\mathbf{t}^d(T) = \{t_i^d; i = 1, \dots, n_d(T)\}$ を生成する. 我々は, ここで提案したモデルを RPP 混合モデルと呼ぶ.

3.2 DPM-RPP モデル

RPP 混合モデルでは, エイジングタイプ (コミュニティ) の総数 K を事前に与えなければならない. K の値も観測データから学習するために, エイジングタイプの生成過程をディリクレ過程 [9] を用いてモデル化し, RPP 混合モデルをノンパラメトリックモデルに拡張することを考える. 我々は, このディリクレ過程と RPP モデルを融合した提案モデルを DPM-RPP モデルと呼ぶ. DPM-RPP モデルでは, 任意のアイテム $d \in D$ への T 時間内のアテンションの到着過程 $\mathbf{t}^d(T) = \{t_i^d; i = 1, \dots, n_d(T)\}$ を以下のようにモデル化する.

まず, D におけるエイジング効果を表す時間緩和関数

^{*1} $f(t; \theta)$ は, 対象とするソーシャルメディアに応じて異なり, 対数正規分布やべき乗則などの密度関数を用いる.

$f(t; \theta)$ のパラメータ値 θ_k ($k = 1, 2, 3, \dots$) を θ の事前分布 $p(\theta | \tau)$ から,

$$\theta_k | \tau \sim p(\theta | \tau) \quad (k = 1, 2, 3, \dots)$$

と生成する．無限次元離散確率分布 $\pi = (\pi_k)_{k=1}^{\infty}$ を Stick-Breaking 過程から $k = 1, 2, 3, \dots$ に対して,

$$v_k | \gamma \sim \text{Beta}(1, \gamma), \quad \pi_k = v_k \prod_{\ell=1}^{k-1} (1 - v_\ell)$$

と生成する．ここに, $\text{Beta}(1, \gamma)$ はパラメータが 1 と $\gamma > 0$ のベータ分布である．ランダム測度 G を

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\theta_k}$$

と定義する．ここに, δ_{θ_k} は位置 θ_k に集中した確率測度を表す．このとき G は, 密度関数 $p(\theta | \tau)$ が定める確率測度を基底測度とし, γ を集中度パラメータとするディリクレ過程 $\text{DP}(\gamma, p(\theta | \tau))$ に従って分布していることに注意しておく．

次に, d のエイジングタイプ z_d をランダム測度 G から

$$z_d | G \sim G$$

と生成する．RPP 混合モデルと同様に, d のフィットネス η_d を式 (4) で生成し, d に対する強化関数 $r_d(t)$ を式 (2) で定義する．このとき, $\lambda_d(t)$ を式 (5) で定義し, レート関数 $\lambda_d(t)$ の非一様ポアソン過程により, d への T 時間内のアテンション到着過程 $t^d(T) = \{t_i^d; i = 1, \dots, n_d(T)\}$ を生成する．

DPM-RPP モデルでは, t_{i-1}^d 時間後に $i-1$ 個目のアテンションが到着したとき, t_i^d 時間後に i 個目のアテンションが到着する確率は,

$$p_1(t_i^d | t_{i-1}^d) = \lambda_d(t_i^d) \exp \left\{ - \int_{t_{i-1}^d}^{t_i^d} \lambda_d(t) dt \right\}$$

で与えられる．また, $t_{n_d(T)}^d$ 時間後から T 時間後までに 1 度もアテンションが到着しない確率は,

$$p_0(T | t_{n_d(T)}^d) = \exp \left\{ - \int_{t_{n_d(T)}^d}^T \lambda_d(t) dt \right\}$$

で与えられる．したがって, T 時間内の d へのアテンション到着 $t^d(T) = \{t_i^d; i = 1, \dots, n_d(T)\}$ の尤度は,

$$\begin{aligned} p(t^d(T) | z_d, \eta_d, \theta_{z_d}) \\ &= p_0(T | t_{n_d(T)}^d) \prod_{i=1}^{n_d(T)} p_1(t_i^d | t_{i-1}^d) \\ &= \eta_d^{n_d(T)} \exp \{ - \eta_d X_d(\theta_{z_d}) \} \prod_{i=1}^{n_d(T)} f(t_i^d; \theta_{z_d}) (i-1 + \nu) \end{aligned} \quad (6)$$

となることに注意しておく．ここに,

$$\begin{aligned} X_d(\theta_{z_d}) &= (n_d(T) + \nu) \int_0^T f(t; \theta_{z_d}) dt \\ &\quad - \sum_{i=1}^{n_d(T)} \int_0^{t_i^d} f(t; \theta_{z_d}) dt \end{aligned} \quad (7)$$

である．

4. DPM-RPP モデルの学習

本章では, DPM-RPP モデルの推定法を与え, さらに, 同モデルをポピュラリティ予測とコミュニティ抽出に適用する手法を示す．

4.1 推定法

アイテム集合 D に対して, T 時間後までのアテンション到着過程の観測データ $\mathcal{O} = \{t^d(T); d \in D\}$ から, DPM-RPP モデルを効率良く推定する手法を与える． D における潜在変数の集合を $Z = \{z_d; d \in D\}$ とし, 時間緩和関数のパラメータの集合を $\Theta = \{\theta_k; k = 1, \dots, K\}$ とする*2．ディリクレ過程は Chinese restaurant process (CRP) と同値であることが知られている [9]．我々は, 提案した RPP 混合モデルの枠組みで, CRP に基づく近似推論アプローチにより DPM-RPP モデルを学習することを考える．パラメータ π と各アイテム $d \in D$ のフィットネス η_d を事前分布に基づいて周辺化し, 以下に示す 3 つのステップ「1) Z のサンプリング, 2) Θ のサンプリング, 3) α, β, τ の推定」を反復して, $K, \Theta, \alpha, \beta, \tau$ の値を推定し, 潜在変数 $z_d = k$ の事後確率を推定する．

ステップ 1) Z のサンプリング: Gibbs サンプラを用いる．エイジングタイプ k に割り当てられているアイテムの数を m_k とする．また, z_d を除いたアイテムのエイジングタイプの集合を $Z_{-d} = Z \setminus \{z_d\}$ とし, そのときのエイジングタイプの総数を K^- とすると, CRP に基づいた z_d の生成確率は,

$$p(z_d = k | Z_{-d}, \gamma) = \begin{cases} \frac{m_k - 1}{|D| - 1 + \gamma} & (1 \leq k \leq K^-) \\ \frac{\gamma}{|D| - 1 + \gamma} & (k > K^-) \end{cases} \quad (8)$$

となる．ここに, $|\cdot|$ は集合の要素数を表す．一方, 式 (6) で与えられるアテンション到着過程の尤度を, 式 (4) の事前分布に基づき η_d について周辺化することで, $t^d(T)$ の周辺尤度として

$$\begin{aligned} p(t^d(T) | z_d, \theta_{z_d}, \alpha, \beta) \\ &= \int p(t^d(T) | z_d, \eta_d, \theta_{z_d}) p(\eta_d | \alpha, \beta) d\eta_d \\ &\propto \frac{\Gamma(n_d(T) + \alpha)}{(X_d(\theta_{z_d}) + \beta)^{n_d(T) + \alpha}} \prod_{i=1}^{n_d(T)} f(t_i^d; \theta_{z_d}) \end{aligned} \quad (9)$$

*2 K は無限大にもなりうることに注意しておく．

を得る．ここに、式 (7) で与えられる $X_d(\theta_{z_d})$ を用いた．DPM-RPP モデルの時間緩和関数 $f(t; \theta)$ に対して、一般的に用いられる対数正規分布や冪乗則関数が組み込まれたとき、そのパラメータに対して共役な事前分布を見つけるのは一般に困難である．したがって、我々は非共役な事前分布を含むモデルに対しても適用するために、補助パラメータによる Gibbs サンプラ (文献 [9] の Algorithm 8) を用いる．新たなタイプ $k \geq K^- + 1$ に対して c 個の補助成分を考え、それらに対する補助パラメータ $\theta_{K^-+1}, \dots, \theta_h$ を与える．ここに、 $h = K^- + c$ である．もし d が固有のタイプに割り当てられているとき、すなわち、 $d' \neq d$ に対し $z_d \neq z_{d'}$ であるならば、 $\theta_{K^-+1} = \theta_{z_d}$ とし、 $\theta_{K^-+2}, \dots, \theta_h$ を式 (3) から生成する．もし d が既存のタイプ k に割り当てられているとき、すなわち、 $\exists d' \neq d$ に対し $z_d = z_{d'}$ であるならば、 $\theta_{K^-+1}, \dots, \theta_h$ を式 (3) から生成する．補助パラメータが与えられると、式 (8), (9) より、条件付き確率

$$p(z_d = k | Z_{-d}, \mathbf{t}^d(T), \Theta, \alpha, \beta, \gamma) \propto \begin{cases} \frac{m_k - 1}{|D| - 1 + \gamma} p(\mathbf{t}^d(T) | z_d = k, \theta_k, \alpha, \beta) & (1 \leq k \leq K^-) \\ \frac{\gamma/c}{|D| - 1 + \gamma} p(\mathbf{t}^d(T) | z_d = k, \theta_k, \alpha, \beta) & (K^- < k \leq h) \end{cases} \quad (10)$$

に従う Gibbs サンプラから z_d をサンプルする．1 度 z_d がサンプルされるごとに、アイテムが 1 つも割り当てられていないエイジングタイプ k に対応するパラメータ θ_k を破棄する．このようにして、 Z と K のサンプルを得る．

ステップ 2) Θ のサンプリング： K と Z の現在のサンプルに対する Θ の不変分布を得るために、我々は Metropolis-Hastings アルゴリズムを適用する．候補点集合 Θ' の提案分布 $q(\Theta' | \Theta)$ として正規分布を用いる^{*3}．また、式 (6) より Θ の事後分布は、

$$p(\Theta | \mathcal{O}, Z, \alpha, \beta, \tau) \propto \prod_{d \in D} p(\mathbf{t}^d(T) | z_d, \theta_{z_d}, \alpha, \beta) \times p(\Theta | \tau) \quad (11)$$

で与えられる．提案分布の対称性 ($q(\Theta' | \Theta) = q(\Theta | \Theta')$) より、受理確率は式 (11) を用いて、

$$A(\Theta', \Theta) = \min \left[1, \frac{p(\Theta' | \mathcal{O}, Z, \alpha, \beta, \tau)}{p(\Theta | \mathcal{O}, Z, \alpha, \beta, \tau)} \right]$$

となり、これに従って Θ' を受理する．この操作を繰り返して Θ のサンプルを得る．

ステップ 3) α, β, τ の推定： K と Z と Θ の現在のサンプルに基づいて、 α, β, τ の値を推定する．まず、最尤推定により α と β の値を推定する．目的関数は、

$$F(\alpha, \beta) = \sum_{d \in D} \ln p(\mathbf{t}^d(T) | z_d, \theta_{z_d}, \alpha, \beta)$$

^{*3} 平均が Θ 、分散共分散行列が単位行列の正数倍である正規分布を用いる．

となり、

$$\begin{aligned} \frac{\partial F}{\partial \beta} &= - \sum_{d \in D} \frac{n_d(T) + \alpha}{X_d(\theta_{z_d}) + \beta} + \frac{|D| \alpha}{\beta} \\ \frac{\partial F}{\partial \alpha} &= \sum_{d \in D} \Psi(n_d(T) + \alpha) - \sum_{d \in D} \ln(X_d(\theta_{z_d}) + \beta) \\ &\quad + |D| (\ln \beta - \Psi(\alpha)) \end{aligned}$$

を得る．ここに、 $\Psi(\cdot)$ はディガンマ関数である．これらを用いて勾配法により α と β を推定する．同様に、 τ の値についても、与えられた θ の事前分布に基づいて目的関数 $F_1(\tau) = \sum_{k=1}^K \ln p(\theta_k | \tau)$ を定義し、最尤推定の枠組みで推定する．

上記の手法により観測データ $\mathcal{O}$ から推定された、エイジングタイプ数 K の値、パラメータ Θ の値および、ハイパーパラメータ α, β, τ の値をそれぞれ、 $K^*, \Theta^* = \{\theta_1^*, \dots, \theta_{K^*}^*\}$ および、 $\alpha^*, \beta^*, \tau^*$ とする．さらに、式 (10) に基づいて潜在変数 z_d をサンプリングすることにより、事後確率 $P(z_d = k | \mathbf{t}^d(T), \Theta^*, \alpha^*, \beta^*, \gamma)$ の推定値 $\phi_{d,k}$ を求め、アイテム $d \in D$ の各エイジングタイプに対する帰属度を K^* 次元離散確率分布 $\phi_d = (\phi_{d,1}, \dots, \phi_{d,K^*})$ として推定する．

4.2 ポピュラリティ予測とコミュニティ構造抽出

観測データ $\mathcal{O}$ から推定した DPM-RPP モデルを用いて、各アイテム $d \in D$ が将来獲得する総アテンション数の予測、およびアイテム群 D におけるコミュニティ構造の抽出を行う．我々は、アイテム d のフィットネス η_d の事後分布

$$\begin{aligned} p(\eta_d | \mathbf{t}^d(T), z_d, \theta_{z_d}^*, \alpha^*, \beta^*) &= \frac{(X_d(\theta_{z_d}^*) + \beta^*)^{n_d(T) + \alpha^*}}{\Gamma(n_d(T) + \alpha^*)} \eta_d^{n_d(T) + \alpha^* - 1} \\ &\quad \times \exp \left\{ - (X_d(\theta_{z_d}^*) + \beta^*) \eta_d \right\} \end{aligned} \quad (12)$$

を用いる．ここに、 $\Gamma(\cdot)$ はガンマ関数である．

まず、DPM-RPP モデルを用いたポピュラリティ予測法を与える．ポピュラリティ予測問題では、将来の t ($\geq T$) 時間後までに、各アイテム $d \in D$ が獲得する総アテンション数の平均値 $x^d(t)$ を予測することを考える．与えられた DPM-RPP モデルでは、式 (5) により、境界条件 $x^d(T) = n_d(T)$ の下での微分方程式

$$\frac{dx^d(t)}{dt} = \eta_d f(t; \theta_{z_d}^*) (x^d(t) + \nu) \quad (t \geq T)$$

の解としてそれらは予測される．よって、

$$x^d(t) = (n_d(T) + \nu) \exp \{ \eta_d Y_d(\theta_{z_d}^*) \} - \nu \quad (13)$$

である．ここに、

$$Y_d(\theta_{z_d}^*) = \int_T^t f(s; \theta_{z_d}^*) ds$$

である．さらに，帰属度 ϕ_d および式 (12) で与えられる η_d の事後分布を用いることで，我々は d に対する $t (\geq T)$ 時間後までのアテンション総数を

$$\begin{aligned} \langle x^d(t) \rangle &= \sum_{k=1}^{K^*} \phi_{d,k} \int x^d(t) p(\eta_d | t^d(T), z_d = k, \theta_k^*, \alpha^*, \beta^*) d\eta_d \\ &= (n_d(T) + \nu) \\ &\quad \times \sum_{k=1}^{K^*} \phi_{d,k} \left(\frac{X_d(\theta_k^*) + \beta^*}{X_d(\theta_k^*) - Y_d(\theta_k^*) + \beta^*} \right)^{n_d(T) + \alpha^*} - \nu \end{aligned}$$

と予測する．

次に，DPM-RPP モデルを用いたコミュニティ構造の抽出法を与える．エイジングタイプに係るエイジング効果パターンの多様性は，各タイプのアイテム群が形成するコミュニティ構造の違いによってもたらされる．したがって，DPM-RPP モデルを用いてアイテムをクラスタリングすることで，それらコミュニティの構造を抽出し，さらにエイジング効果の情報に基づいてコミュニティの分析を行うことができる．アイテム集合 D のクラスタリング（コミュニティ構造の抽出）は各アイテム $d \in D$ に対して推定値 z_d^* を

$$z_d^* = \underset{1 \leq k \leq K^*}{\operatorname{argmax}} \phi_{d,k}$$

と決定することで行う．クラスタリング結果を $Z^* = \{z_d^*; d \in D\}$ とする．

5. 評価実験

本論文ではクックパッドの実データと人工データを用いて提案法を評価する．まず，データの統計解析に基づいて，DPM-RPP モデルに対しクックパッドデータにおける適切な時間緩和関数を与える．次に，クックパッドデータのポピュラリティ予測における DPM-RPP モデルの性能を評価する．また，人工データを用いた実験により，データに存在するコミュニティの数と予測性能の関係を調べる．さらに，アテンションの到着パターンに基づいた料理レシピのコミュニティ構造の抽出を行う．

5.1 クックパッドデータにおける実験設定

ユーザが投稿したアイテムに対して，他のユーザから獲得するアテンションが観測可能なソーシャルメディアとして，料理レシピ共有サイト「クックパッド^{*4}」を対象とした．クックパッドにおいてユーザは創作した料理レシピを投稿でき，他のユーザはその料理レシピを気に入ったならば「つくれぽ」と呼ばれる賛意メッセージを送ることができる．したがって，アイテムはユーザが投稿した料理レシ

^{*4} <http://cookpad.com/>

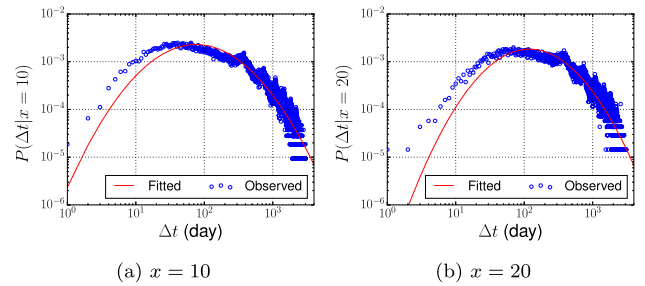


図 1 クックパッドデータにおけるエイジング効果

Fig. 1 Aging effects for the cookpad data.

ピであり，それへのアテンションは他のユーザからのつくれぽである．我々は，アテンションの到着過程が観測可能な期間「2006 年 2 月 27 日から 2014 年 9 月 30 日」における，アイテム数 1,618,419，アテンション数 9,449,479 のデータを利用した．

まず，クックパッドにおける新アテンション獲得能力に対するエイジング効果を分析した．時間緩和関数の形状を決定するためには，エイジング効果以外の要因を制御する必要があるため，我々は Wang ら [14] の方法に従って，アイテムがある数のアテンション数を獲得するまでの時間を分析した．具体的には，アテンション数 x のアイテムに対して，それが生成されてから $x+1$ となるまでの時間の分布 $P(\Delta t | x)$ を調べた．図 1(a) と図 1(b) にそれぞれ $x=10$ と $x=20$ に対する結果を示す．実線は対数正規分布によるフィッティングであり，クックパッドにおけるアテンションダイナミクスのエイジング効果をよくとらえている．したがって，我々は，DPM-RPP モデルの時間緩和関数として，式 (1) と同様に対数正規分布の密度関数

$$f(t; \theta_k) = g(t; \mu_k, \sigma_k)$$

を用いた．ここに， $\mu_k \in \mathbb{R}$ と $\sigma_k > 0$ はエイジングタイプ k に対するパラメータであり，その事前分布として Normal-Inverse-Gamma 分布

$$\begin{aligned} p(\mu_k, \sigma_k^2 | \tau_1, \tau_2, \tau_3, \tau_4) \\ = \frac{\sqrt{\tau_2 \tau_4^3}}{\sqrt{2\pi} \sigma_k^{2\tau_3+3} \Gamma(\tau_3)} \exp \left\{ -\frac{\tau_2 (\mu_k - \tau_1)^2 + 2\tau_4}{2\sigma_k^2} \right\} \end{aligned}$$

を用いた．ここに， $\tau_1 \in \mathbb{R}$ ， $\tau_2, \tau_3, \tau_4 > 0$ はハイパーパラメータ τ である．

次に，クックパッドにおける“rich-get-richer”現象を分析した．我々は，Redner [11] と同様の手法で，獲得アテンション数と，attachment rate と呼ばれる新アテンション獲得能力の関係を調べた．ある期間 I における attachment rate A_j は，その直前の期間に j 個のアテンションを獲得したアイテムが， I 内に獲得した新アテンション数の平均値として与えられる．図 2(a) および図 2(b) にそれぞれ，直前の 2 カ月間の j と $I = [2013-1-1, 2013-2-1)$ (すなわち，2013 年 1 月 1 日から 2013 年 2 月 1 日の直前まで) およ

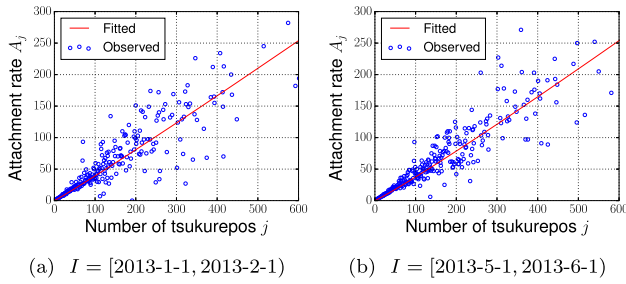


図 2 クックパッドデータにおける“rich-get-richer”現象
Fig. 2 Rich-get-richer phenomena for the cookpad data.

び $I = [2013-5-1, 2013-6-1]$ (すなわち, 2013 年 5 月 1 日から 2013 年 6 月 1 日の直前まで) における A_j の結果を示す. 実線は線形関数によるフィッティングである. これらの結果は, クックパッドのアテンションダイナミクスに“rich-get-richer”現象が存在することを示しており, また, j の増加にともなって A_j がほぼ線形に増加することが観察されることから, $r_d(t)$ として式 (2) が適切であることが示唆される.

実験では, 2006 年 2 月 27 日から 2010 年 1 月 31 日のデータからデータセット $\mathcal{D}_1$, 2010 年 2 月 1 日から 2013 年 1 月 31 日のデータからデータセット $\mathcal{D}_2$ をそれぞれ構築した. ここに, $\mathcal{D}_1$ ではアイテム数 4,777, アテンション数 647,958, $\mathcal{D}_2$ ではアイテム数 5,766, アテンション数 705,992 であった.

5.2 ポピュラリティ予測実験

すべてのアイテム $d \in D$ に対し, 等しく T 時間内のアテンション到着過程 $t^d(T)$ を用いて将来の $t (\geq T)$ 時間後までに獲得するアテンション総数 $x^d(t)$ の予測を行い, その性能を評価した. 本論文では, 特に T 時間内に 50 以上のアテンションを獲得したアイテムのみを対象とした.

DPM-RPP モデルを 2 つの手法 RPP モデル [12] および ERPP モデル [6] と総アテンション数予測に対する性能で比較する. 我々は, ポピュラリティ予測において標準的に用いられる Mean Absolute Percentage Error (MAPE) を指標として予測性能を測定した. ここに, アイテム $d \in D$ が $t (\geq T)$ 時間後までに獲得する真のアテンション総数を $y^d(t)$ とすると, MAPE は

$$MAPE = \frac{1}{|D|} \sum_{d \in D} \left| \frac{x^d(t) - y^d(t)}{y^d(t)} \right|$$

で定義される. 実験では, 観測時間 T を 50 日および 70 日として, その 1 週間後から 24 週間後までのアテンション総数を予測した. ハイパーパラメータに関しては, ν の値は既存研究 [6] に従いすべてのモデルで 10 と設定した. また, 提案モデルでは $\gamma = 1$ と設定した. 各モデルのパラメータをランダムに初期化することで, 異なる初期値における学習結果の予測性能を測定した. 10 回の試行に対する

結果を図 3 に示す. いずれの条件においても, DPM-RPP モデルが最も高性能であり, RPP モデルがそれに続いている. また, 予測する時間が長くなるにつれて, RPP モデルとの性能差が大きくなることが観察される. さらに我々は, DPM-RPP モデルと RPP モデルの性能差を評価するために, 両側 5% 有意水準におけるウェルチの t 検定を行った. データセット $\mathcal{D}_1$ の $T = 50$ では 7 週間後以降, $T = 70$ では 10 週間後以降においてそれぞれ有意差が認められた. また, データセット $\mathcal{D}_2$ の $T = 50$ と $T = 70$ ではともに 9 週間後以降において有意差が認められた. 性能差はデータ依存であるが, 観測時間 T が短い場合により大きな有意差が現れる傾向が見られた. ここに, 有意差が認められるまでの時間においても, DPM-RPP モデルの予測性能は少なくとも RPP モデルとほぼ同等であったことに注意しておく. これらの結果は, エイジングタイプに基づくアイテム群のコミュニティ構造をとらえた DPM-RPP モデルの有効性を示している.

5.3 人工データによる実験

5.2 節では, クックパッドの実データを用いたポピュラリティ予測実験により, 長期的な予測をする場合において特に DPM-RPP モデルが有効であることを示した. 次に我々は, エイジングタイプの総数 K と予測性能の関係を調べるために, 人工データを用いた実験を行った. ここでは, クックパッドデータにおいて DPM-RPP モデルに続く予測性能を示した RPP モデルとの比較を行った.

5.1 節におけるクックパッドデータに対する設定と同様に, DPM-RPP モデルの時間緩和関数として対数正規分布の密度関数を式 (14) のように与え, 人工データの生成においても同様の時間緩和関数を用いた. まず, エイジングタイプの真の総数 K とアイテムの総数 $|D|$ を指定して, RPP 混合モデルに従い, $k = 1, \dots, K$ に対する時間緩和関数のパラメータ μ_k, σ_k と, アイテム $d \in D$ のエイジングタイプ z_d およびフィットネス η_d を生成し, 非一様ポアソン過程のレート関数

$$\lambda_d(t) = \eta_d g(t | \mu_{z_d}, \sigma_{z_d}) r_d(t) \quad (14)$$

を定義した. 次に, 観測時間 T を指定し, 時間伸縮理論に基づくシミュレーションアルゴリズム [2] を用いて, 式 (14) のレート関数に従う $d \in D$ のアテンション到着過程 $t^d(T) = \{t_i^d; i = 1, \dots, n_d(T)\}$ を次の手順で生成した.

- i $t_0^d = 0, i = 1$ として初期化する.
- ii 平均が 1 の指数分布から乱数 $\xi_i^d (\geq 0)$ を

$$\xi_i^d | 1 \sim \text{Exponential}(1)$$

と生成する.

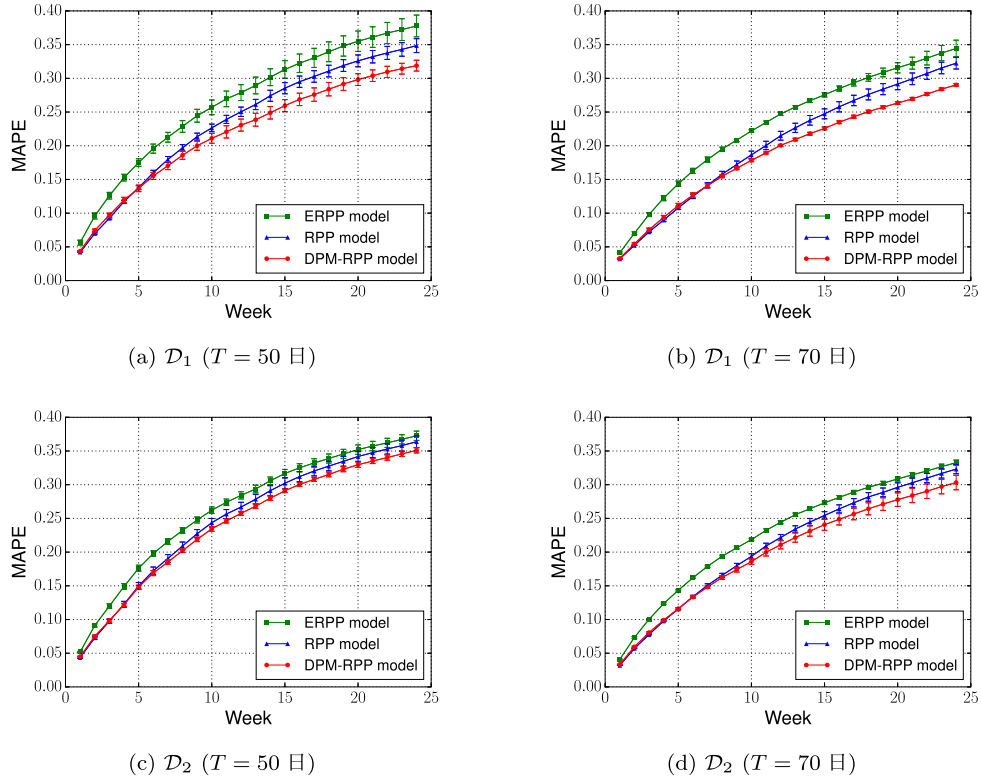


図 3 クックパッドデータにおけるポピュラリティ予測性能の比較。横軸は観測終了からの時間 $(t - T)$ を週で表している

Fig. 3 Performance comparison in popularity prediction for the cookpad data. The horizontal axis indicates $t - T$ (how many weeks since the final observation).

- iii 乱数 ξ_i^d と式 (14) のレート関数を用いて定義される方程式

$$\xi_i^d = \int_{t_{i-1}^d}^{t_i^d} \lambda_d(t) dt$$

の解として t_i^d を生成する。

- iv $t_i^d > T$ であるならば終了し、さもなければ、 $i = i + 1$ として ii へ戻る。

実験では、 $|D| = 500$, $T = 50$ と固定し、 $K = 1$, $K = 5$ および $K = 20$ の設定においてそれぞれ 5 つのデータセットを構築した。紙数の関係上、ここでは、各設定におけるデータセットの 1 つに対する結果のみを示す*5。それらのアテンション数は、 $K = 1$ では 7,734, $K = 5$ では 15,163, さらに $K = 20$ では 19,075 であった。

10 回の試行における結果を図 4 に示す。クックパッドデータに対する結果と同様に、DPM-RPP モデルが RPP モデルよりも高性能であり、予測する時間が長くなるにつれて、RPP モデルとの性能差が大きくなっている。性能差は $K = 1$ の場合に最も大きく、 K が増加するにつれて小さくなるのが観察される。また、両側 5% 有意水準にお

けるウェルチの t 検定の結果、 $K = 1$ では $t - T = 20$ 以降において、 $K = 5$ では $t - T = 10$ 以降において、さらに $K = 20$ では $t - T = 30$ 以降においてそれぞれ有意差が認められた。したがって、 K が $|D|$ に近づくにつれて、各アイテムを独立にモデル化した RPP モデルとの性能差が小さくなるものの、少数のコミュニティ構造により特徴づけられるようなアイテム群に対しては、特に DPM-RPP モデルが有効であると考えられる。明確なコミュニティ構造を持つこれらの人工データにおいては、顕著にそれが表れていると思われる。

5.4 コミュニティ構造の抽出

DPM-RPP モデルを用いて、エイジングタイプの観点からクックパッドのアイテム群におけるコミュニティ構造を抽出した。各データセットにおいて興味深いコミュニティ構造が観察されたが、特に、本論文ではデータセット D_1 に対する結果を報告する。

コミュニティ構造分析のためにデータセット D_1 の全データを用いて DPM-RPP モデルの学習を行い、クラスタリング結果 Z^* を得た。推定されたエイジングタイプの総数は $K^* = 12$ であった。また、各タイプ $k = 1, \dots, 12$ に対し、推定されたパラメータ μ_k^* および σ_k^* を用いて時間緩和関数 $f(t; \theta_k^*) = g(t; \mu_k^*, \sigma_k^*)$ の最頻値を求め、その値に

*5 いずれの設定においても、すべてのデータセットでほぼ同様な結果であった。

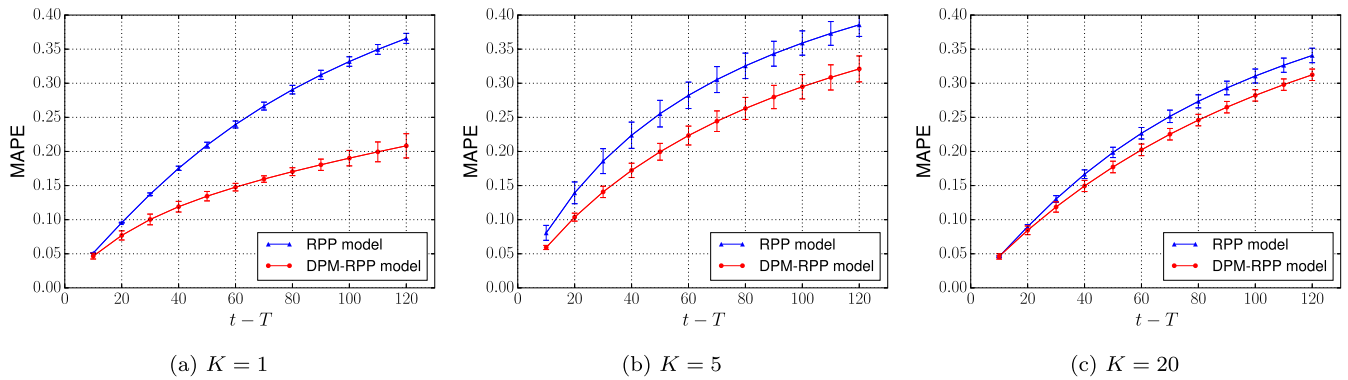


図 4 人工データにおけるポピュラリティ予測性能の比較

Fig. 4 Performance comparison in popularity prediction for the synthetic data.

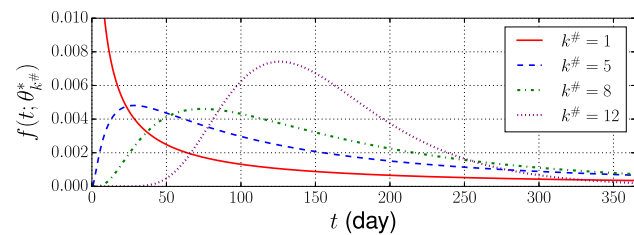


図 5 データセット D_1 の各エイジングタイプにおける時間緩和関数
Fig. 5 Temporal-relaxation functions of aging types extracted for dataset D_1 .

基づいてエイジングタイプを昇順に並べ替えることで新たにインデックス $k\# = 1, \dots, 12$ を与えた. 特徴的であった 4 つのコミュニティ $k\# = 1, 5, 8, 12$ について, それらの時間緩和関数を図 5 に示す. さらに, 各コミュニティ $k\#$ を代表する料理レシピ群として, それへの帰属度 $\phi_{d,k\#}$ に関する料理レシピランキングの上位 3 位までを表 1 に示す. まず, エイジングタイプ $k\# = 1$ の時間緩和関数は指数関数的なエイジング効果を持つことが観察される. $k\# = 1$ においては, バレンタインやひな祭り, クリスマスなどに関する料理レシピが見られたことから, イベント料理に関連するコミュニティが形成されていると考えられ, それらの料理レシピに対しては, 投稿された直後からつくればが送られるが, その後急速にユーザの関心が薄れてゆくことが分かる. また, $k\# = 5$ ではケーキ, タルト, マフィンといった料理レシピが見られ, 洋菓子に関連するコミュニティが形成されていると考えられる. $k\# = 5$ の時間緩和関数を見ると, これらの料理レシピは投稿後およそ 1 カ月間で急速にユーザの関心を集めてゆき, その後は緩やかに風化しながらも継続的につくればを獲得し続ける特徴があることが分かる. 一方, $k\# = 8$ では煮物やカレーなど一般家庭の食卓に並ぶ料理レシピが多く, またレシピ名にも「家族」や「夫」などの家庭を連想させる単語が見受けられることから, 家庭料理に関連するコミュニティが形成されていると考えられる. 時間緩和関数の特徴より, $k\# = 8$ の料理レシピに対する関心は, 投稿後およそ 2 カ月かけて

表 1 データセット D_1 において DPM-RPP モデルにより抽出されたコミュニティ

Table 1 Communities extracted by the DPM-RPP model for dataset D_1 .

(a) $k\# = 1$	
レシピ名	カテゴリ名
バレンタインに**チョコ♡マドレーヌ**	マドレーヌ
ひな祭りのお寿司♪ひしもち風♪	お祝い・パーティ寿司
クリスマスに Kiss を♡Mistletoe Salad	クリスマスのサラダ
(b) $k\# = 5$	
レシピ名	カテゴリ名
ノンオイル☆ノン小麦粉☆おから林檎ケーキ	おからを使ったケーキ
さつまいもとナッツのキャラメルタルト	パイ・ドナルド
パン屋さんのマフィン♪	色々なマフィン
(c) $k\# = 8$	
レシピ名	カテゴリ名
家族喜ぶ♪豚バラこんにゃく	豚バラ肉
♪美味すぎて・旦那が帰ってくるカレー♪	ビーフカレー
夫に好評☆なすのひき肉はさみ煮	なす
(d) $k\# = 12$	
レシピ名	カテゴリ名
本当のチキンティッカマサラカレー	インド風カレー
本当のインド家庭のジンジャーチャイ	チャイ
簡単なのに本格的♪チャプチェ	

徐々に高まってゆき, その後は $k\# = 5$ と比較してより緩やかに減衰してゆくことが分かる. さらに, $k\# = 12$ では他のコミュニティに対して比較的多様な種類の料理レシピが見られたが, インドカレーやチャイと呼ばれるインドの茶, チャプチェと呼ばれる韓国料理などアジア圏の料理レシピが特徴的であり, ある種のエスニック料理に関連するコミュニティが形成されていると考えられる. 図 5 より, これらの料理レシピは投稿後およそ 1 カ月間はほとんどつくればを獲得せず, 4 カ月後にかけて徐々に関心を集めた後, $k\# = 5$ と $k\# = 8$ に比べて急速に風化してゆくエイジング効果を持つことが分かる.

ところで, コミュニティ構造の抽出に関しては, RPP モデルにより各アイテムについて独立に学習し, 後処理として, アイテム群を学習した時間緩和関数のパラメータ値に

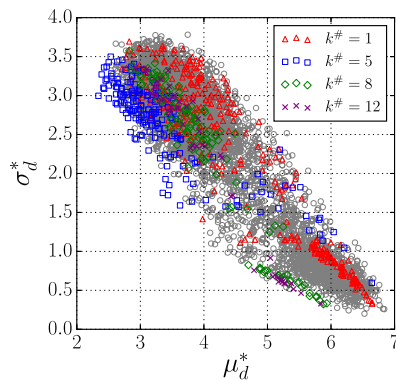


図 6 RPP モデルにおける時間緩和関数のパラメータ推定値と DPM-RPP モデルにより抽出されたコミュニティの関係 (データセット $\mathcal{D}_1$)

Fig. 6 Relation of the estimated parameter-values in temporal-relaxation functions and the extracted communities for the DPM-RPP model for dataset $\mathcal{D}_1$.

基づいてクラスタリングする手法も考えられる。そこで、DPM-RPP モデルによるコミュニティ構造が、RPP モデルによるそのような手法によって容易に抽出しうるかどうかを調べた。まず、データセット $\mathcal{D}_1$ の全データを用いて、各アイテム d の RPP モデルのパラメータ μ_d^* , σ_d^* を推定した。図 6 にその散布図とともに、DPM-RPP モデルによるコミュニティ $k^{\#} = 1, 5, 8, 12$ を示す。コミュニティ $k^{\#} = 1, 5, 8, 12$ が、その散布図において広い範囲に複雑に分布していることが観察される。したがって、RPP モデルで推定された時間緩和関数のパラメータ値に基づいて単純にアイテム群をクラスタリングすることで DPM-RPP モデルによるコミュニティ構造を抽出するのは、一般に困難と考えられる。すなわち、DPM-RPP モデルは、対象とするアイテム群全体に対するアテンションの到着過程のデータから、既存手法では抽出が困難なエイジングタイプに基づくコミュニティ構造を抽出できるといえる。

6. まとめ

本論文では、対象とするソーシャルメディアのアイテム群全体に対するアテンション到着過程の確率的生成モデルとして、ディリクレ過程と RPP モデルを融合した DPM-RPP モデルを提案した。また、観測データに基づく DPM-RPP モデルの効率的な学習法を与え、それに基づいて、各アイテムが将来に獲得するポピュラリティの予測法および、エイジングタイプの観点からのアイテム群のコミュニティ抽出法を与えた。そして、料理レシピ共有サイト「クックパッド」における料理レシピへのアテンション到着過程の実データを用いて提案モデルを評価した。提案した DPM-RPP モデルは、料理レシピの将来ポピュラリティ予測において、従来の RPP モデルと ERPP モデルよりも予測性能が高いことを示した。また、人工データを用いた実験により、少数のコミュニティ構造によって特徴

づけられるようなアイテム群のポピュラリティ予測において、DPM-RPP モデルが特に有効であることを示した。さらに、DPM-RPP モデルを用いて、エイジングタイプの観点における料理レシピ群のコミュニティ構造を抽出し、その特徴を明らかにした。

提案手法を様々なソーシャルメディアに適用して、そのポピュラリティ予測性能を評価し、また、それが抽出するエイジングタイプに基づくコミュニティ構造を分析することは、今後の重要な課題である。しかしながら、本研究において、従来のようなアイテムごとに完全に独立なアテンションダイナミクスのモデリングと比較して、アイテム群の関係性を組み込んだ提案モデルの有望性が示されたことは、ソーシャルメディアにおけるアテンションダイナミクスのモデル構築に対する重要な第 1 歩と考えている。

謝辞 本研究はクックパッド株式会社と国立情報学研究所が提供するクックパッドデータを利用し、JSPS 科研費 17K00433 の助成を受けたものである。

参考文献

- [1] Bandari, R., Asur, S. and Huberman, B.: The pulse of news in social media: Forecasting popularity, *Proc. ICWSM'12*, pp.26–33 (2012).
- [2] Brown, E.N., Barbieri, R., Ventura, V., Kass, R.E. and Frank, L.M.: The time-rescaling theorem and its application to neural spike train data analysis, *Neural Computation*, Vol.14, No.2, pp.325–346 (2002).
- [3] Chen, W., Lakshmanan, L. and Castillo, C.: Information and influence propagation in social networks, *Synthesis Lectures on Data Management*, Vol.5, pp.1–177 (2013).
- [4] Cheng, J., Adamic, L., Dow, P., Kleinberg, J. and Leskovec, J.: Can cascades be predicted?, *Proc. WWW'14*, pp.925–936 (2014).
- [5] Daneshmand, H., Gomez-Rodriguez, M., Song, L. and Schölkopf, B.: Estimating diffusion network structures: Recovery conditions, sample complexity & soft-thresholding algorithm, *Proc. ICML'14*, pp.793–801 (2014).
- [6] Gao, S., Ma, J. and Chen, Z.: Modeling and predicting retweeting dynamics on microblogging platforms, *Proc. WSDM'15*, pp.107–116 (2015).
- [7] Gomez-Rodriguez, M., Leskovec, J. and Krause, A.: Inferring networks of diffusion and influence, *Proc. KDD'10*, pp.1019–1028 (2010).
- [8] Kempe, D., Kleinberg, J. and Tardos, E.: Maximizing the spread of influence through a social network, *Proc. KDD'03*, pp.137–146 (2003).
- [9] Neal, R.M.: Markov chain sampling methods for Dirichlet process mixture models, *Journal of computational and graphical statistics*, Vol.9, No.2, pp.249–265 (2000).
- [10] Pinto, H., Almedia, J. and Goncalves, M.: Using early view patterns to predict the popularity of youtube videos, *Proc. WSDM'13*, pp.365–374 (2013).
- [11] Redner, S.: Citation Statistics from 110 Years of Physical Review, *Physics Today*, Vol.58, pp.49–54 (2005).
- [12] Shen, H., Wang, D., Song, C. and Barabási, A.-L.: Modeling and predicting popularity dynamics via reinforced Poisson processes, *Proc. AAAI'14*, pp.291–297 (2014).
- [13] Szabo, G. and Huberman, B.: Predicting the popularity

of online content, *Comm. ACM*, Vol.53, No.8, pp.80–88 (2010).

- [14] Wang, D., Song, C. and Barabási, A.-L.: Quantifying long-term scientific impact, *Science*, Vol.342, No.6154, pp.127–132 (2013).
- [15] Yang, J. and Leskovec, J.: Patterns of temporal variation in online media, *Proc. WSDM'11*, pp.177–186 (2011).
- [16] Zhao, Q., Erdogdu, M., He, H., Rajaraman, A. and Leskovec, J.: SEISMIC: A self-exciting point process model for predicting tweet popularity, *Proc. KDD'15*, pp.1513–1522 (2015).



松谷 貫司

2016 年龍谷大学工学部電子情報学科卒業。現在、同大学大学院理工学研究科電子情報学専攻修士課程在学中。



木村 昌弘

1987 年大阪大学理学部数学科卒業。
1989 年同大学大学院理学研究科数学
専攻修士課程修了。同年日本電信電
話株式会社入社。NTT コミュニケー
ション科学基礎研究所を経て、現在、
龍谷大学工学部電子情報学科教授。

複雑ネットワーク科学，データマイニングおよび機械学習
の研究と教育に従事。博士（理学）。日本数学会，日本応用
数理学会，人工知能学会，電子情報通信学会各会員。