

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

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 [10] 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 KIMURA²

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 [10] presented a probabilistic model, called the *RPP 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. 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

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, Ryukoku University

² 龍谷大学理工学部電子情報学科
Department of Electronics and Informatics, Ryukoku University

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 [11]。

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], [3], [9], [11], [12]。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捉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 [4], [6] 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 [2], [7] に対し、多くの研究がなされている。

最近 Shen ら [10] 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 [5] 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 [13] は、各ユーザのフォロワ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、フォロワ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. 提案モデル

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。我々は RPP モデルを拡張して、任意の $d \in D$ への T 時間内のアテンションの到着過程 $\mathbf{t}^d(T) = \{t_i^d; i = 1, \dots, n_d(T)\}$ の確率的生成モデルを提案する。

2.1 RPP モデル

RPP 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、 $\lambda_d(t)$ 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。RPP モデルでは, すべてのモデルパラメータの値は, 一般にアイテム毎に異なり得る。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。

2.2 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得ると考えられる。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毎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 をもつ 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

^{*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。

$\beta > 0$ 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 混合モデルとよぶ。

2.3 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 [8] 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 $f(t; \theta)$ 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 $\boldsymbol{\pi} = (\pi_k)_{k=1}^{\infty}$ 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 $\text{DP}(\gamma, p(\theta | \tau))$ にしたが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 $\mathbf{t}^d(T) = \{t_i^d; i = 1, \dots, n_d(T)\}$ を生成する.

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 $\mathcal{O} = \{\mathbf{t}^d(T); d \in D\}$ から, DPM-RPP モデルを推定する. 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, 事後確率 $P(z_d = k | \mathbf{t}^d(T), \Theta^*, \alpha^*, \beta^*, \tau^*)$ 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. これらを用いて, 各アイテム $d \in D$ に対する $t (\geq T)$ 時間後までのアテンション総数を

$$\langle x^d(t) \rangle = (n_d(T) + \nu) \times \sum_{k=1}^{K^*} \phi_{d,k} \left(\frac{X_d(\theta_k^*) + \beta^*}{X_d(\theta_k^*) - Y_d(\theta_k^*) + \beta^*} \right)^{n_d(T) + \alpha^*} - \nu$$

と予測する. ここに,

$$X_d(\theta_k^*) = (n_d(T) + \nu) \int_0^T f(t; \theta_k^*) dt - \sum_{i=1}^{n_d(T)} \int_0^{t_i^d} f(t; \theta_k^*) dt,$$

$$Y_d(\theta_k^*) = \int_T^t f(s; \theta_k^*) ds$$

である. さらに, 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

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

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.

3. 評価実験

本論文ではクックパッドの実データを用いて提案法を評価する. まず, DPM-RPP モデルに時間緩和関数を与える. 次に, ポピュラリティ予測における DPM-RPP モデルの性能を評価し, 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.

3.1 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*2」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

*2 <http://cookpad.com/>

ば「つくれば」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ばである. 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.

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

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

を用いた. ここに, $\Gamma(\cdot)$ はガンマ関数であり, $\tau_1 \in \mathbb{R}$, $\tau_2, \tau_3, \tau_4 > 0$ がハイパーパラメータ τ である.

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.

3.2 ポピュラリティ予測実験

全てのアイテム $d \in D$ に対し, 等しく T 時間内のアテンション到着過程 $\mathbf{t}^d(T)$ 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 [10] およびモデル [5] と総アテンション数予測に対する性能で比較する. ERPP モデルは, 冪乗関数による時間緩和関数と, 指数強化関数と呼ばれる強化関数を用いている. 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

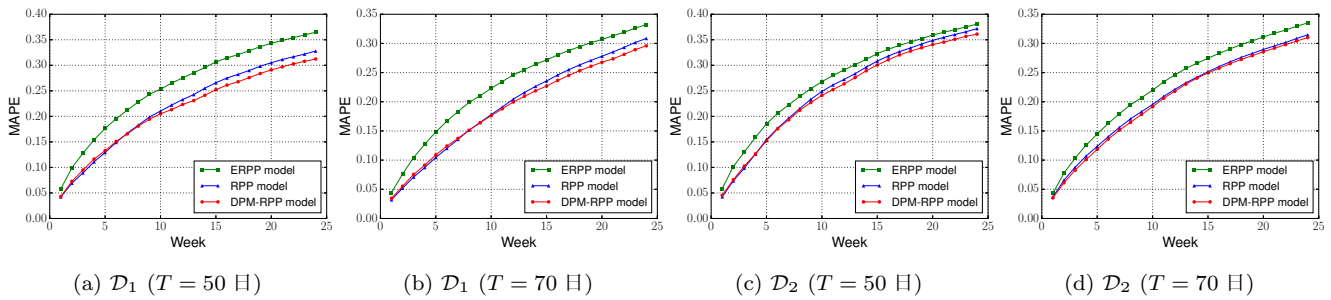


図 1: クックパッドデータにおけるポピュラリティ予測性能の比較

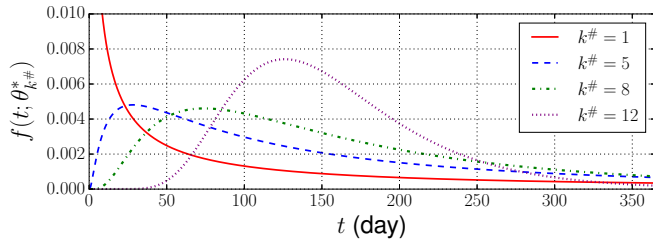


図 2: データセット \mathcal{D}_1 の各エイジングタイプにおける時間緩和関数

既存研究 [5] に従い全てのモデルで 10 と設定した。また、提案モデルでは $\gamma = 1$ と設定した。MAPE による比較結果を図 1 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、DPM-RPP モデルは、他のモデルと比較して、観測期間が短い場合において特に良い性能を示している。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捉えた DPM-RPP モデルの有効性を示している。

3.3 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 \mathcal{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 \mathcal{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 2 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。

表 1: データセット \mathcal{D}_1 において抽出されたコミュニティ

(a) $k^\# = 1$

レシピ名	カテゴリー名
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	カテゴリー名
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	ベイクドタルト
パン屋さんのマフィン♪	色々なマフィン

(c) $k^\# = 8$

レシピ名	カテゴリー名
家族喜ぶ♪豚バラこんにやく	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	カテゴリー名
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	

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 か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か月かけて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2より、これらの料理レシピは投稿後およそ1か月間はほとんどつくればを獲得せず、4か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。

4. 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大規模な実データを用いて、提案モデルを評価した。まず、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。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毎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一歩と考えている。

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た。

参考文献

- [1] Bandari, R., Asur, S. and Huberman, B.: The pulse of news in social media: Forecasting popularity, *Proceedings of ICWSM'12*, pp. 26–33 (2012).
- [2] Chen, W., Lakshmanan, L. and Castillo, C.: Infor-

- mation and influence propagation in social networks, *Synthesis Lectures on Data Management*, Vol. 5, pp. 1–177 (2013).
- [3] Cheng, J., Adamic, L., Dow, P., Kleinberg, J. and Leskovec, J.: Can cascades be predicted?, *Proceedings of WWW'14*, pp. 925–936 (2014).
- [4] Daneshmand, H., Gomez-Rodriguez, M., Song, L. and Schölkopf, B.: Estimating diffusion network structures: Recovery conditions, sample complexity & soft-thresholding algorithm, *Proceedings of ICML'14* (2014).
- [5] Gao, S., Ma, J. and Chen, Z.: Modeling and predicting retweeting dynamics on microblogging platforms, *Proceedings of WSDM'15*, pp. 107–116 (2015).
- [6] Gomez-Rodriguez, M., Leskovec, J. and Krause, A.: Inferring networks of diffusion and influence, *Proceedings of KDD'10*, pp. 1019–1028 (2010).
- [7] Kempe, D., Kleinberg, J. and Tardos, E.: Maximizing the spread of influence through a social network, *Proceedings of KDD'03*, pp. 137–146 (2003).
- [8] 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).
- [9] Pinto, H., Almedia, J. and Goncalves, M.: Using early view patterns to predict the popularity of youtube videos, *Proceedings of WSDM'13*, pp. 365–374 (2013).
- [10] Shen, H., Wang, D., Song, C. and Barabási, A.-L.: Modeling and predicting popularity dynamics via reinforced Poisson processes, *Proceedings of AAAI'14*, pp. 291–297 (2014).
- [11] Szabo, G. and Huberman, B.: Predicting the popularity of online content, *Communications of the ACM*, Vol. 53, No. 8, pp. 80–88 (2010).
- [12] Yang, J. and Leskovec, J.: Patterns of temporal variation in online media, *Proceedings of WSDM'11*, pp. 177–186 (2011).
- [13] Zhao, Q., Erdogdu, M., He, H., Rajaraman, A. and Leskovec, J.: SEISMIC: A self-exciting point process model for predicting tweet popularity, *Proceedings of KDD'15*, pp. 1513–1522 (2015).