原著論文

Original Paper

オンラインアイテム群における共有イベント系列に基づいた協調構造の抽出

Extracting Cooperative Structure among Online Items from Share-Event Sequence

松谷 貫司
Kanji Matsutani
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
t16m024@mail.ryukoku.ac.jp

熊野 雅仁
Masahito Kumano
龍谷大学理工学部電子情報学科
Department of Electronics and Informatics, Ryukoku University
kumano@rins.ryukoku.ac.jp

木村 昌弘
Masahiro Kimura
（同　　上）
kimura@rins.ryukoku.ac.jp

斉藤 和巳
Kazumi Saito
静岡県立大学経営情報学部
School of Administration and Informatics, University of Shizuoka
k-saito@u-shizuoka-ken.ac.jp

大原 剛三
Kouzou Ohara
青山学院大学理工学部情報テクノロジー学科
Department of Integrated Information Technology, Aoyama Gakuin University
ohara@it.aoyama.ac.jp

元田 浩
Hiroshi Motoda
大阪大学産業科学研究所
Institute of Scientific and Industrial Research, Osaka University
motoda@ar.sanken.osaka-u.ac.jp

keywords: cooperative structure, social media mining, stochastic process model

Summary

Social media allows people to post widely and evaluate diverse information including ideas, news and opinions. Once such an online item is posted on a social media site, it can be appreciated and shared by many people and become popular. This kind of phenomenon can have a large influence on people's daily life and social trends. Thus, studies on modeling the arrival process of shares to an individual item have recently attracted a great deal of interest in the field of social media mining. In this paper, we propose, by combining a Dirichlet process with a Hawkes process in a novel way, a probabilistic model, called cooperative Hawkes process (CHP) model, to discover the cooperative structure among all the items involved. The proposed model takes into account all the arrival processes of shares for those items. We develop an efficient method of inferring the CHP model from the observed sequences of share-events, and present an effective framework for predicting the future popularity of each of these items. Using synthetic and real data, we demonstrate that the CHP model outperforms the Hawkes process model without interaction among items (HP model) and the multivariate Hawkes process model (MHP model) in terms of popularity prediction. Moreover, for real data from a cooking-recipe sharing site, we discover the cooperative structure among cooking-recipes in view of popularity dynamics by applying the CHP model.

1. は　じ　め　に

Facebook，Twitter，YouTube，@cosme，Cookpad など，ソーシャルメディアサイトが Web 空間における人々の重要なコミュニケーションの場として発展し続けている．人々はソーシャルメディアを利用して，アイデアやニュース，オピニオンなどの多種多様な情報を容易に，世界に向けて広く発信したり評価したりすることができるようになってきた．ソーシャルメディアサイトに投稿されたそのようなオンラインアイテムは，多くの人々に高く評価され共有されていくことによって，そのポピュラリティを獲得していく．このような現象は，人々の日常生活や社会のトレンドにも大きな影響を及ぼす場合があるので，オンラインアイテムが共有されポピュラリティを獲得していく過程のモデル化が，近年ソーシャルメディアマイ

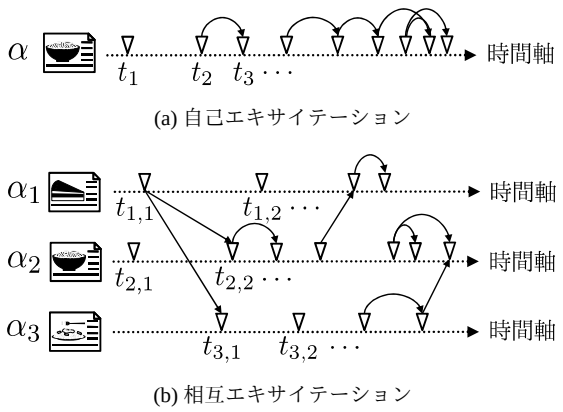


図 1 共有イベント時系列における自己エキサイテーションと相互エキサイテーションの説明図.
逆三角印は，アイテムがその時刻に共有されたことを表し，矢印は共有イベントの誘発を表している．(a) はアイテム α の共有イベント時系列とその自己エキサイテーションの説明図である．例えば，時刻 t_2 に α が共有されたというイベントが，時刻 t_3 にそれが共有されたというイベントを誘発したことを表している．(b) は 3 つのアイテム $\alpha_1,\alpha_2,\alpha_3$ の共有イベント時系列と，それらの相互エキサイテーションの説明図である．例えば，時刻 $t_{1,1}$ に α_1 が共有されたというイベントが，時刻 $t_{2,2}$ に α_2 が共有されたというイベントと，時刻 $t_{3,1}$ に α_3 が共有されたというイベントを誘発したことを表している．

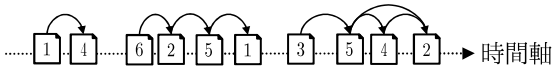
ニングの分野で注目されている [Gao 15, Shen 14, Szabo 10, Wang 13, Zhao 15].

ソーシャルメディアサイトに投稿されたオンラインアイテムに対して，それがいつ共有されたかについては一般に観測可能であるが，それを誰が共有したかについてはプライバシーやセキュリティの関係上，必ずしも公開されているとは限らない．したがって少なくとも，オンラインアイテムの共有イベント発生に関する時系列データについては，一般にそれを容易に獲得できると言える．ここに，アイテム α の共有イベント (t,α) とは，あるユーザが賛意を表明して時刻 t に α を共有したことを表す．例えば，アイテムの共有イベントは，Facebook では記事の「シェア」，Twitter ではツイートに対する「リツイート」，Cookpad では料理レシピへの賛意メッセージである「つくれぽ」に相当する．本論文で我々は，ソーシャルメディアサイトに投稿されたオンラインアイテム群に対して，それらの共有イベント時系列の発生過程を同時にモデル化するという問題に取り組む．まず，個々のアイテムはそれ独自の魅力を有していると考えられ，またアイテムの共有イベントについては，その過去の共有イベントがそれ自身の将来の共有イベントの発生を誘発するという，自己エキサイテーション性を有していると考えられる（図 1 (a) 参照）．例えば，料理レシピ共有サイトでは，レシピに対する賛意メッセージがそれ自身への将来の賛意メッセージを誘発することがありうる．さらに，こうしたエキサイテーションの性質は異なるアイテム間にも存在し，アイテムの共有イベントは相互エキサイテーション性をもつと仮定するのが自然である．すなわち，あるアイテムの共有イベントは，別のアイテムの将来の共有イベントの発生をも誘発しうると考えられる（図 1 (b) および図 2 (a) 参照）．例えば，料理レシピ共有サイトでは，パスタのあるカルボナーラレシピに対する賛意メッセージが別のカルボナーラレシピの将来の賛意メッセージをも誘発することがありうる．このようなアイテム間の相互エキサイテーション性を特徴づけるために，アイテム α' からアイテム α への影響度 $\tilde{w}_{\alpha,\alpha'}$ を考える．このとき，すべてのアイテムペア (α,α') に対して影響度 $\tilde{w}_{\alpha,\alpha'}$ が異なりうると仮定すること（図 2 (b) 参照）は現実的でない．実際，アイテムは多数あるが，それに比べて相対的に少量の共有イベント観測データしか得られないことが多く，すべてのアイテム間の影響度を推定することは一般に困難と考えられるからである．

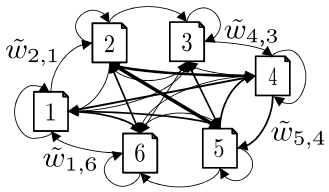
本論文では，共有イベント発生の時系列に基づいて，オンラインアイテム群における関係性を抽出するための確率モデルを提案し，各アイテムの将来ポピュラリティを精度よく予測することを目指す．そのためにまず，我々はアイテム群における**協調構造** $\mathcal{A}_1,\ldots,\mathcal{A}_K$ を導入する．ここに，任意の協調グループ $\mathcal{A}_k$ に対して，それに属する各アイテム α' は，共有イベントの相互エキサイテーション性に関し，任意のアイテム α に等しく影響 $w_{\alpha,k}$ を及ぼし（図 2 (c) 参照），アイテム α' からアイテム α への影響度 $\tilde{w}_{\alpha,\alpha'}$ は $\tilde{w}_{\alpha,\alpha'}=w_{\alpha,k}$ で与えられる．例えばレシピ共有サイトでは，そうした協調グループは，電子レンジで簡単かつ短時間で調理できるレシピ群のような，レシピのジャンルに対応している可能性がある．

提案モデルは，計数過程 [Aalen 08] の一種であり，“rich-get-richer” 現象を捉えるためによく用いられる Hawkes 過程 [Hawkes 71] を土台としている．我々は，ディリクレ過程 [Neal 00] を新たなやり方で融合することで Hawkes 過程に協調構造を組み込み，指定されたアイテム群すべてに対して，それらの共有イベント時系列の発生過程を同時にモデル化する．我々の提案モデルを ***CHP* モデル** ***(cooperative Hawkes process model)*** と呼ぶ．我々は，共有イベント時系列の観測データから CHP モデルを効率よく推定する手法を開発し，アイテム群の協調構造を同定する手法および，CHP モデルの下で近い将来の各アイテムのポピュラリティを予測する有効な枠組みを与える．人工データおよび料理レシピ共有サイトの実データを用いた実験により，CHP モデルが既存の Hawkes 過程モデルよりもポピュラリティ予測において有効であることを示す．さらに，CHP モデルに基づいて，共有イベント時系列の発生過程の観点から，料理レシピ共有サイトにおける料理レシピ群の協調構造を明らかにする．

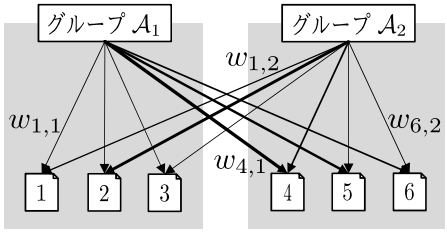
本論文の構成は以下のとおりである．まず，2 章では関連研究について簡潔に述べる．3 章では，Hawkes 過程について述べ，CHP モデルを提案する．4 章において，CHP モデルの確率的推定法を開発し，ポピュラリティ予測のための枠組みを与える．5 章では，人工データおよ



(a) 共有イベント時系列と相互エキサイテーション



(b) 完全結合



(c) 協調構造

図 2 共有イベントの相互エキサイテーションにおけるアイテム間の影響関係の例.
(a) は 6 つのアイテムに対する共有イベントの時系列とそれらの相互エキサイテーションの例である．1 から 6 までの数字が書かれたものがアイテムで，それを時間軸上に配置することで，各アイテムがある時刻に共有されたことを表現している．矢印は共有イベントの誘発を表している．(b) と (c) は，(a) の相互エキサイテーションに関する 6 つのアイテム間の影響関係を例示したものであり，矢印の太さは影響度の強さに対応している．(b) はすべてのアイテムペア (α,α') に対し影響度 $\tilde{w}_{\alpha,\alpha'}$ が異なると仮定した場合の例であり，完全結合グラフとなる．(c) は，6 つのアイテム間の協調構造 $\{\mathcal{A}_k\}$ と影響度 $\{w_{\alpha,k}\}$ を例示したものである．

び料理レシピ共有サイトの実データに対する実験結果を報告する．最後に，6 章はまとめである．

2. 関 連 研 究

個々のアイテムのポピュラリティ予測に関する先行研究のほとんどは，固定された時間窓内に獲得する共有イベント数の平均値を既存の時系列解析法を用いて予測する研究 [Szabo 10, Yang 11] と，網羅的に抽出した特徴量を用いて既存の分類法や回帰分析法を適用する研究 [Bandari 12, Cheng 14, Pinto 13] であった．近年，Wang ら [Shen 14, Wang 13] は，RPP モデルと呼ばれる個々のアイテムへのアテンション到着過程モデルを提案し，論文引用ダイナミクスに対してそれが既存の予測手法よりも高性能であることを示した．Gao ら [Gao 15] は，Twitter におけるリツイートダイナミクスを対象として RPP モデルを改良した．Zhao ら [Zhao 15] もまたリツイートダイナミクスを対象とし，Hawkes 過程に各リツイートユーザのフォロワー数情報を組み込むことにより，与えられたツイートの総リツイート数を効率的に予測する手法 SEISMIC を与えた．しかしこれらの研究では，本論文における我々のアプローチとは異なり，アイテムごとに共有イベントの発生過程を独立にモデル化しているため，対象とするアイテム群全体の関係分析やマイニングには限界がある．本論文では，アイテム群の共有イベント時系列の発生過程を同時にモデル化することにより，それらアイテム群の協調構造の発見を目指す．

ソーシャルメディアサイトに投稿されたアイテムの共有イベントの発生過程に関する研究は，ソーシャルネットワーク上での情報拡散の研究とも関連している．影響最大化問題，すなわち，指定された情報拡散モデルの下で影響力が強い少数ノード群を見つける問題に対して，多くの研究がなされてきた [Chen 13, Kempe 03]．また，情報伝搬の観測系列からその背後にあるユーザ間の影響関係を表すソーシャルネットワークを推定する研究も数多くなされてきた [Daneshmand 14, Gomez-Rodriguez 10]．そこでは，ソーシャルネットワーク上の情報伝搬を形成するイベント系列をモデル化するために，多変量 Hawkes 過程がよく用いられている [Farajtabar 14, Iwata 13, Zhou 13a, Zhou 13b]．多変量 Hawkes 過程は，また，オピニオンダイナミクス [De 16] や情報拡散とソーシャルネットワークの共進化ダイナミクス [Farajtabar 17] のモデル化へも拡張されている．本論文では，上記の先行研究と同様に Hawkes 過程に焦点を置くが，それらのアプローチとは異なり，プライバシー保護の観点から，誰がアイテムを共有したかという情報を用いることなしにアイテム群の関係を明らかにすることを目指す．

3. モ デ ル

対象とするオンラインアイテムの集合 $\mathcal{A}$ を固定し，期間 $[0,T)$ でのそれらに対する共有イベント系列の発生過程を同時にモデル化することを考える．ここに，T はそれほど大きくない正数（例えば，2,3 カ月）である．任意のアイテム $\alpha\in\mathcal{A}$ に対して，時刻 t までのその共有イベントを計数過程 $N_\alpha(t)$ [Aalen 08] としてモデル化することを考える．ここに，$N_\alpha(t)$ は期間 $[0,t)$ 内での α の共有イベント数を表す．$N(t)$ を期間 $[0,t)$ 内での $\mathcal{A}$ に対する共有イベントの総数，すなわち，

$$N(t)=\sum_{\alpha\in\mathcal{A}}N_\alpha(t)$$

とする．各 $n=1,\ldots,N(t)$ に対して，第 n 共有イベントを組 (t_n,α_n) で表す．これは，アイテム α_n が時刻 t_n に

共有されたことを意味する．時刻 t までの $\mathcal{A}$ に対する共有イベント系列を，

$$\mathcal{T}(t) = \{(t_n, \alpha_n); n = 1, \ldots, N(t)\}$$

とする．また，任意のアイテム $\alpha \in \mathcal{A}$ に対して，時刻 t までのその共有イベント系列を，

$$\mathcal{T}_\alpha(t) = \{(t_n, \alpha); n = 1, \ldots, N_\alpha(t)\}$$

とする．各 $\alpha \in \mathcal{A}$ に対して，$\lambda_\alpha(t)$ を α に対する強度関数とする．ここに，$\lambda_\alpha(t)$ は，時刻 t までの共有イベントの観測系列 $\mathcal{T}(t)$ が与えられたとき，微小な時間窓 $[t, t+dt)$ 内で α の共有イベントが発生する条件付き確率，

$$\lambda_\alpha(t)\, dt \;=\; \mathbb{E}[dN_\alpha(t) \mid \mathcal{T}(t)]$$

を表している（[Farajtabar 17] 参照）．本章では，強度関数 $\lambda_\alpha(t)$ のモデル化について考える．

3·1 Hawkes 過 程

まず，基本的な Hawkes 過程について述べる．

§1 一様ポアソン過程

各アイテム $\alpha \in \mathcal{A}$ は，それ固有の魅力 $\mu_\alpha > 0$ をもつと考えられる．最も単純な設定では，$\lambda_\alpha(t)$ が $\mathcal{T}(t)$ と独立でありかつ定数であると仮定することである．すなわち，

$$\lambda_\alpha(t) = \mu_\alpha$$

である．これは一様ポアソン過程と呼ばれる．

§2 アイテム間に相互作用のない Hawkes 過程 (HP)

アイテムの共有イベントは，自己エキサイテーション性をもち，“rich-get-richer” 現象を示すと考えられる．これは，アイテム間に相互作用のない Hawkes 過程，

$$\lambda_\alpha(t) = \mu_\alpha + \tilde{w}_{\alpha,\alpha} \sum_{(t_n,\alpha)\in\mathcal{T}_\alpha(t)} \exp\{-\tilde{\gamma}_\alpha(t - t_n)\} \quad (1)$$

として捉えられる．ここに，$\tilde{w}_{\alpha,\alpha} > 0$ は共有イベントの自己エキサイテーションにおけるアイテム α からそれ自身への影響度を表し，$\tilde{\gamma}_\alpha$ は α からの影響度の時間減衰率を表している．この場合，$\lambda_\alpha(t)\, dt \;=\; \mathbb{E}[dN_\alpha(t) \mid \mathcal{T}_\alpha(t)]$ となることに注意しておく．

3·2 多変量 Hawkes 過程 (MHP)

アイテムの共有イベントは相互エキサイテーション性をも有すると考えられるので，一般に，多変量 Hawkes 過程，

$$\lambda_\alpha(t) = \mu_\alpha + \sum_{(t_n,\alpha_n)\in\mathcal{T}(t)} \tilde{w}_{\alpha,\alpha_n} \exp\{-\tilde{\gamma}_{\alpha_n}(t - t_n)\} \quad (2)$$

がモデル化において必要となる．ここに，$\tilde{w}_{\alpha,\alpha_n} > 0$ は，共有イベントの相互エキサイテーションにおけるアイテム α_n からアイテム α への影響度を表す．しかしながら 1 章でも述べたように，$|\mathcal{A}|$ と比べて $|\mathcal{T}(T)|$ が十分に大きくないような現実問題においては，すべてのアイテムペア $(\alpha, \alpha') \in \mathcal{A} \times \mathcal{A}$ に対して，$\tilde{w}_{\alpha,\alpha'}$ が異なる値をもつと仮定することは現実的でない．したがって我々は，$\mathcal{A}$ 内の関係構造を考慮に入れ，各アイテム $\alpha \in \mathcal{A}$ に対する将来の共有イベントを正確に予測することを目指す．

3·3 協 調 構 造

式 (2) で定義される多変量 Hawkes 過程に対して，我々は $\mathcal{A}$ 内に**協調構造** $Z = \{z(\alpha); \alpha \in \mathcal{A}\}$ を次のように導入する．$\mathcal{A} = \bigcup_{k=1}^{K} \mathcal{A}_k$ (disjoint union) であり，$z(\alpha') = k$ であるのは，$\alpha' \in \mathcal{A}_k$ であるときに限る．そして，強度関数 $\lambda_\alpha(t)$ は，

$$\lambda_\alpha(t \mid Z) = \mu_\alpha + \sum_{(t_n,\alpha_n)\in\mathcal{T}(t)} w_{\alpha,z(\alpha_n)} \exp\{-\gamma_{z(\alpha_n)}(t - t_n)\} \quad (3)$$

となる．ここに，$w_{\alpha,1}, \ldots, w_{\alpha,K} > 0$ が存在して，$\forall \alpha' \in \mathcal{A}$ に対し，$\tilde{w}_{\alpha,\alpha'} = w_{\alpha,z(\alpha')}$ であり，また，$\gamma_1, \ldots, \gamma_K > 0$ が存在して，$\forall \alpha' \in \mathcal{A}$ に対し，$\tilde{\gamma}_{\alpha'} = \gamma_{z(\alpha')}$ である．協調グループ $\mathcal{A}_k$ に属する各アイテムは，任意のアイテム α に $w_{\alpha,k}$ という形で等しく影響を及ぼし，さらに同一の時間減衰率 γ_k をもつということに注意しておく．

3·4 提 案 モ デ ル

一般に，協調グループの数 K を事前に決定することは困難であるため，観測データから K の値を決定できることが望ましい．したがって我々は，ディリクレ過程 [Neal 00] を用いて，$\mathcal{A}$ 内の協調構造を多変量 Hawkes 過程にノンパラメトリックベイズ形式で組み込むことにより，$\mathcal{A}$ に対する共有イベント系列を生成する確率モデルを定義する．我々の提案モデルを ***CHP (cooperative Hawkes process)* モデル**と呼ぶ．ここに，その強度関数 $\lambda_\alpha(t)$ は式 (3) で与えられる．

CHP モデルのパラメータは，$\boldsymbol{\mu} = (\mu_\alpha)_{\alpha\in\mathcal{A}}$，$Z = \{z(\alpha); \alpha \in \mathcal{A}\}$，$\boldsymbol{\gamma} = (\gamma_k)_{k=1}^{K}$ および $W = (\boldsymbol{w}_k)_{k=1}^{K}$ である．ただし，$\boldsymbol{w}_k = (w_{\alpha,k})_{\alpha\in\mathcal{A}}$ である．これらのパラメータは以下の手順で生成される．まず，無限次元離散確率分布 $\boldsymbol{\pi} = (\pi_k)_{k=1}^{\infty}$ が，Stick-Breaking 過程から $k = 1, 2, 3, \ldots$ に対して，

$$v_k \mid \beta \sim \mathrm{Beta}(1, \beta), \quad \pi_k = v_k \prod_{\ell=1}^{k-1} (1 - v_\ell)$$

と生成される．ここに，$\mathrm{Beta}(1, \beta)$ はパラメータが 1 と $\beta > 0$ のベータ分布である．次に，$k = 1, 2, 3, \ldots$ に対して，$\phi_k = (\gamma_k, \boldsymbol{w}_k)$ は事前確率分布 H から，

$$\phi_k \mid H \sim H$$

と生成される．ランダム測度 G を

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

と定義する．ここに，δ_ϕ は位置 ϕ におけるアトム，すなわち，ϕ に集中した確率測度である．G は基底分布 H と集中度パラメータ β のディリクレ過程 $\mathrm{DP}(\beta, H)$ に従って分布していることに注意しておく．Z は，G から各 $\alpha \in \mathcal{A}$ に対して，

$$z(\alpha) \,|\, G \sim G$$

と生成される．さらに $\boldsymbol{\mu}$ は，パラメータ $\boldsymbol{\eta} = (\eta_0, \eta_1)$ のガンマ分布から各 $\alpha \in \mathcal{A}$ に対して，

$$\mu_\alpha \,|\, \boldsymbol{\eta} \sim \mathrm{Gamma}(\boldsymbol{\eta}) \tag{4}$$

と生成される．ここに，$\eta_0, \eta_1 > 0$ である．

CHP モデルにおいて，観測系列 $\mathcal{T}(T)$ の確率密度は，

$$p(\mathcal{T}(T) \,|\, Z, \boldsymbol{\mu}, \boldsymbol{\gamma}, W) = \exp\left\{ -\int_0^T \sum_{\alpha \in \mathcal{A}} \lambda_\alpha(t \mid Z)\, dt \right\} \prod_{(t_n, \alpha_n) \in \mathcal{T}(T)} \lambda_{\alpha_n}(t_n \,|\, Z) \tag{5}$$

で与えられる．

4. 学　習　法

アイテム集合 $\mathcal{A}$ における共有イベントの観測系列を $\mathcal{T}(T) = \{(t_n, \alpha_n); n = 1, \ldots, N(T)\}$ とする．本章では，$\mathcal{T}(T)$ から CHP モデルを推定する手法を開発し，ポピュラリティダイナミクスの観点から $\mathcal{A}$ 内の協調構造 $\mathcal{A}_1, \cdots, \mathcal{A}_K$ を同定する手法を与える．さらに，$\mathcal{A}$ に対する将来の共有イベントを予測する枠組みを与える．

4·1 パラメータ推定

観測系列 $\mathcal{T}(T)$ からパラメータ $Z, K, \boldsymbol{\mu}, \boldsymbol{\gamma}, W$ を推定することを考える．ディリクレ過程は Chinese restaurant 過程 (CRP) [Neal 00] と等価であることから，我々は CRP に基づく近似推論アプローチをとる．事前分布 H は，パラメータ $\boldsymbol{\sigma} = (\sigma_0, \sigma_1)$ のガンマ分布とパラメータ $\boldsymbol{\nu} = (\nu_0, \nu_1)$ のガンマ分布の積とし，$\boldsymbol{\gamma}$ と W は，$k = 1, \ldots, K$ と $\alpha \in \mathcal{A}$ に対して独立に，

$$\gamma_k \,|\, \boldsymbol{\sigma} \sim \mathrm{Gamma}(\boldsymbol{\sigma}) \tag{6}$$

$$w_{\alpha,k} \,|\, \boldsymbol{\nu} \sim \mathrm{Gamma}(\boldsymbol{\nu}) \tag{7}$$

と生成されるとする．ここに，$\sigma_0, \sigma_1, \nu_0, \nu_1 > 0$ である．

まず，計数過程に関する重ね合わせの原理を用いて学習アルゴリズムを単純化することを考える．そのために，第 n 共有イベント (t_n, α_n) が第 x_n 共有イベント (t_{x_n}, α_{x_n}) によって引き起こされたことを表す潜在変数の集合 $X = \{x_n; n = 1, \ldots, N(T)\}$ を導入する．ここに，$x_n = 0, 1, \ldots, n-1$ であり，$x_n = 0$ は第 n 共有イベント (t_n, α_n) がアイテム α_n の固有の魅力によって引き起こされたことを意味する．具体的には，第 n 共有イベント (t_n, α_n) に対する強度関数 $\lambda_{\alpha_n}(t_n, x_n \,|\, Z)$ を，

$$\lambda_{\alpha_n}(t_n, x_n \,|\, Z) = \begin{cases} \mu_{\alpha_n} & \text{if } x_n = 0 \\ w_{\alpha_n, z(\alpha_{x_n})} \exp\left\{ -\gamma_{z(\alpha_{x_n})} (t_n - t_{x_n}) \right\} & \text{if } 1 \le x_n < n \end{cases}$$

と定義する．ここに，$x_n \in \{0, 1, \ldots, n-1\}$ である．このとき，独立なポアソン過程に対する重ね合わせの原理より，強度関数 $\lambda_{\alpha_n}(t_n \,|\, Z)$ (式 (3) 参照) は，

$$\lambda_{\alpha_n}(t_n \,|\, Z) = \sum_{x_n=0}^{n-1} \lambda_{\alpha_n}(t_n, x_n \,|\, Z) \tag{8}$$

となる．これはよく知られた性質であり ([Farajtabar 14, Hawkes 71] 参照)，計数過程のベイズ推定に対してもよく用いられている ([Iwata 13, Linderman 14] 参照)．実際，式 (8) の分解は我々の推定法においても重要な役割を果たし，これにより，$Z, \boldsymbol{\mu}, \boldsymbol{\gamma}, W$ が与えられたときの $\mathcal{T}(T)$ と X の結合尤度は，

$$\begin{aligned} p(\mathcal{T}(T), X \,|\, Z, \boldsymbol{\mu}, \boldsymbol{\gamma}, W) &= \exp\left\{ -T \sum_{\alpha \in \mathcal{A}} \mu_\alpha - \sum_{k=1}^{K} F(Z, \gamma_k) \sum_{\alpha \in \mathcal{A}} w_{\alpha,k} \right\} \\ &\quad \times \prod_{(t_n, \alpha_n) \in \mathcal{T}(T)} \lambda_{\alpha_n}(t_n, x_n \,|\, Z) \end{aligned}$$

と扱いやすい積の形になる (式 (5) 参照)．ここに，

$$F(Z, \gamma_k) = \frac{1}{\gamma_k} \sum_{n=1}^{N(T)} \left(1 - \exp\left\{ -\gamma_k (T - t_n) \right\}\right) \mathbb{I}[z(\alpha_n) = k]$$

であり，$\mathbb{I}[y]$ は，y が真ならば $\mathbb{I}[y] = 1$，y が偽ならば $\mathbb{I}[y] = 0$ である指示関数を表す．よって，尤度 $p(\mathcal{T}(T), X \mid Z, \boldsymbol{\mu}, \boldsymbol{\gamma}, W)$ は，$\boldsymbol{\mu}$ と W について事前確率 (式 (4), (7) 参照) に関し容易に周辺化することができ，

$$\begin{aligned} &p(\mathcal{T}(T), X \,|\, Z, \boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\nu}) \\ &= \iint p(\mathcal{T}(T), X \,|\, Z, \boldsymbol{\mu}, \boldsymbol{\gamma}, W) p(\boldsymbol{\mu} \,|\, \boldsymbol{\eta}) p(W \,|\, \boldsymbol{\nu}) d\boldsymbol{\mu}\, dW \\ &= \prod_{\alpha \in \mathcal{A}} \frac{\Gamma(C_\alpha + \eta_0)}{(T + \eta_1)^{C_\alpha + \eta_0}} \frac{\eta_1^{\eta_0}}{\Gamma(\eta_0)} \\ &\quad \times \prod_{\alpha \in \mathcal{A}} \prod_{k=1}^{K} \frac{\Gamma(D_{\alpha,k} + \nu_0)}{(F(Z, \gamma_k) + \nu_1)^{D_{\alpha,k} + \nu_1}} \frac{\nu_1^{\nu_0}}{\Gamma(\nu_0)} \\ &\quad \times \prod_{k=1}^{K} f(Z, X, \gamma_k) \end{aligned} \tag{9}$$

を得る．ここに，

$$f(Z,X,\gamma_k)=\exp\left\{-\gamma_k\sum_{n=1}^{N(T)}(t_n-t_{x_n})\mathbb{I}[z(x_n)=k]\right\}$$

であり，$\Gamma(s)$ はガンマ関数である．また，

$$C_\alpha=\sum_{(t_n,\alpha_n)\in\mathcal{T}(T)}\mathbb{I}[\alpha_n=\alpha]\mathbb{I}[x_n=0]$$

はアイテム α の共有イベントがその固有の魅力によって引き起こされた回数を表し，

$$D_{\alpha,k}=\sum_{(t_n,\alpha_n)\in\mathcal{T}(T)}\mathbb{I}[\alpha_n=\alpha]\mathbb{I}[z(x_n)=k]$$

はアイテム α の共有イベントが協調グループ $\mathcal{A}_k$ に属するアイテムの共有イベントから引き起こされた回数を表している．

提案推定法では，パラメータ $Z,K,\boldsymbol{\mu},\boldsymbol{\gamma},W$ の推定値は次の 4 ステップを反復することにより得られる．

§1 Z の Gibbs サンプリング
§2 X の Gibbs サンプリング
§3 $\boldsymbol{\gamma}$ の Metropolis-Hastings サンプリング
§4 $\boldsymbol{\mu}$ と W のサンプリングおよび $\boldsymbol{\eta}$，$\boldsymbol{\nu}$，$\boldsymbol{\sigma}$ の更新

以下では，これらのステップについて詳細に記述する．また，提案推定法では上記のパラメータだけでなく，X の推定値も得られること注意する．

§1 Z の Gibbs サンプリング

グループ $\mathcal{A}_k$ に割り当てられたアイテムの数を m_k とする．すなわち，

$$m_k=|\{\alpha\in\mathcal{A};\, z(\alpha)=k\}|$$

とする．また，

$$Z^{-\alpha}=Z\setminus\{z(\alpha)\},\ K^-=\left|Z^{-\alpha}\right|$$

とおく．CRP において $z(\alpha)$ が生成される確率は，

$$P(z(\alpha)=k\,|\,Z^{-\alpha},\beta)=\begin{cases}\dfrac{m_k^{-\alpha}}{|\mathcal{A}|-1+\beta} & \text{for } 1\le k\le K^-\\ \dfrac{\beta}{|\mathcal{A}|-1+\beta} & \text{for } k>K^-\end{cases}\tag{10}$$

で与えられる．ここに，$m_k^{-\alpha}$ はアイテム α を除いて数えたときの m_k の値を表す．CHP モデルに含まれている指数減衰関数に対し，そのパラメータ γ_k の共役事前分布を与えるのは一般に困難である．そこで，このようなパラメータを含む CHP モデルに対して CRP に基づく近似推論を適用するために，補助パラメータによる Gibbs サンプリングを用いる ([Neal 00] のアルゴリズム 8 を参照)．新たな協調グループとして b 個の補助コンポーネントを導入し，それらに対応するパラメータ $\gamma_{K^-+1},\ldots,\gamma_h$ を生成する．ここに，b は正の整数であり，$h=K^-+b$ である．すなわち，$\forall\alpha'\in\mathcal{A}\setminus\{\alpha\}$ に対して $z(\alpha)\neq z(\alpha')$ ならば，$\gamma_{K^-+1}=\gamma_{z(\alpha)}$ とおき，式 (6) から $\gamma_{K^-+2},\ldots,\gamma_h$ を生成する．また，$z(\alpha)=z(\alpha')$ であるような $\exists\alpha'\in\mathcal{A}\setminus\{\alpha\}$ が存在するならば，式 (6) から $\gamma_{K^-+1},\ldots,\gamma_h$ を生成する．補助パラメータが与えられると，$z(\alpha)$ に対する新しい値は $\{1,\ldots,h\}$ の中から条件付き確率，

$$\begin{aligned}&P(z(\alpha)=k\,|\,\mathcal{T}(T),Z^{-\alpha},X,\boldsymbol{\gamma},\boldsymbol{\eta},\boldsymbol{\nu},\beta)\\&\propto P(z(\alpha)=k\,|\,Z^{-\alpha},\beta)p(\mathcal{T}(T),X\,|\,z(\alpha)=k,Z^{-\alpha},\mathcal{Y})\\&=\begin{cases}m_k^{-\alpha}\ p(\mathcal{T}(T),X\,|\,z(\alpha)=k,\mathcal{Y}) & \text{for } 1\le k\le K^-\\ \beta/b\ \ p(\mathcal{T}(T),X\,|\,z(\alpha)=k,\mathcal{Y}) & \text{for } K^-<k\le h\end{cases}\end{aligned}$$

に従う Gibbs サンプラーを用いてサンプルされる (式 (9), (10) 参照)．ここに，$\mathcal{Y}=\{Z^{-\alpha},X,\boldsymbol{\gamma},\boldsymbol{\eta},\boldsymbol{\nu}\}$ である．$z(\alpha)$ の新しい値がサンプルされる毎に，アイテムが 1 つも割り当てられていない協調グループ $\mathcal{A}_k$ のパラメータ γ_k を破棄する．このようにして，Z と K のサンプルを得る．

§2 X の Gibbs サンプリング

Z と K の現在のサンプルが得られると，x_n に対する新しい値は $\{0,1,\ldots,n-1\}$ の中から条件付き確率，

$$\begin{aligned}&P(x_n=i\,|\,\mathcal{T}(T),Z,X^{-n},\boldsymbol{\gamma},\boldsymbol{\eta},\boldsymbol{\nu})\\&\propto p(x_n=i,\mathcal{T}(T)\,|\,Z,X^{-n},\boldsymbol{\gamma},\boldsymbol{\eta},\boldsymbol{\nu})\\&=\begin{cases}\dfrac{C_{\alpha_n}^{-n}+\eta_0}{T+\eta_1} & \text{for } i=0\\ \dfrac{D_{\alpha_n,z(\alpha_i)}^{-n}+\nu_0}{F(Z,\gamma_{z(\alpha_i)})+\nu_1}\exp\{-\gamma_{z(\alpha_i)}(t_n-t_i)\} & \text{for } 1\le i<n\end{cases}\end{aligned}$$

に従う Gibbs サンプラーを用いてサンプルされる (式 (9) 参照)．ここに，上付き文字 $-n$ は第 n 共有イベントを除いて得られる集合もしくは値を意味する．

§3 $\boldsymbol{\gamma}$ の Metropolis-Hastings サンプリング

上述のように $\boldsymbol{\gamma}$ の共役事前分布を与えるのは一般に困難なため，現在のサンプル Z,K,X に対する $\boldsymbol{\gamma}$ の不変分布を得るために，我々は Metropolis-Hastings アルゴリズムを採用する．候補ベクトル $\boldsymbol{\gamma}'$ の提案分布には正規分布を用いる．このとき $\boldsymbol{\gamma}'$ の事後確率分布は，

$$\begin{aligned}&p(\boldsymbol{\gamma}\,|\,\mathcal{T}(T),Z,X,\boldsymbol{\eta},\boldsymbol{\nu},\boldsymbol{\sigma})\\&\qquad\propto p(\mathcal{T}(T),X\,|\,Z,\boldsymbol{\gamma},\boldsymbol{\eta},\boldsymbol{\nu})\,p(\boldsymbol{\gamma}\,|\,\boldsymbol{\sigma})\end{aligned}\tag{11}$$

となる (式 (6), (9) を参照)．提案分布の対称性，$q(\boldsymbol{\gamma}'\,|\,\boldsymbol{\gamma})=q(\boldsymbol{\gamma}\,|\,\boldsymbol{\gamma}')$ より，$\boldsymbol{\gamma}'$ の受理確率は式 (11) から，

$$A(\boldsymbol{\gamma}'\,|\,\boldsymbol{\gamma})=\min\left[1,\frac{p(\boldsymbol{\gamma}'\,|\,\mathcal{T}(T),Z,X,\boldsymbol{\eta},\boldsymbol{\nu},\boldsymbol{\sigma})}{p(\boldsymbol{\gamma}\,|\,\mathcal{T}(T),Z,X,\boldsymbol{\eta},\boldsymbol{\nu},\boldsymbol{\sigma})}\right]$$

と得られる．$\boldsymbol{\gamma}'$ は，$A(\boldsymbol{\gamma}'\,|\,\boldsymbol{\gamma})$ に従って受理される．これらの操作を繰り返すことにより $\boldsymbol{\gamma}$ のサンプルが得られる．

§4 μ と W のサンプリングおよび η,ν,σ の更新

Z,K,X,γ の現在のサンプルが与えられたとき，μ と W をそれらの事後分布，

$$p(\mu_\alpha \,|\, \mathcal{T}(T), X, \boldsymbol{\eta}) = \mathrm{Gamma}(C_\alpha + \eta_0, T + \eta_1)$$

および

$$p(w_{\alpha,k} \,|\, \mathcal{T}(T), Z, X, \boldsymbol{\gamma}, \boldsymbol{\nu}) = \mathrm{Gamma}(D_{\alpha,k} + \nu_0, F(Z,\gamma_k) + \nu_1)$$

の期待値として，すなわち

$$\mu_\alpha = \frac{C_\alpha + \eta_0}{T + \eta_1}$$
$$w_{\alpha,k} = \frac{D_{\alpha,k} + \nu_0}{F(Z,\gamma_k) + \nu_1}$$

としてサンプルする．次に，ハイパーパラメータ $\boldsymbol{\eta},\boldsymbol{\nu},\sigma$ を最尤推定により更新する．まず，$\boldsymbol{\eta}$ の目的関数は，

$$\mathcal{L}_1(\boldsymbol{\eta}) = \sum_{\alpha\in\mathcal{A}} \ln \frac{\Gamma(C_\alpha+\eta_0)}{(T+\eta_1)^{C_\alpha+\eta_0}} + |\mathcal{A}| \ln \frac{\eta_1^{\eta_0}}{\Gamma(\eta_0)}$$

で与えられる．また，$\boldsymbol{\nu}$ と $\boldsymbol{\sigma}$ の目的関数はそれぞれ，

$$\mathcal{L}_2(\boldsymbol{\nu}) = \sum_{\alpha\in\mathcal{A}}\sum_{k=1}^{K} \ln \frac{\Gamma(D_{\alpha,k}+\nu_0)}{(F(Z,\gamma_k)+\nu_1)^{D_{\alpha,k}+\nu_1}} + |\mathcal{A}| K \ln \frac{\nu_1^{\nu_0}}{\Gamma(\nu_0)}$$

および

$$\mathcal{L}_3(\boldsymbol{\sigma}) = \sum_{k=1}^{K} ((\sigma_0-1)\ln\gamma_k - \sigma_1\gamma_k) + K \ln \frac{\sigma_1^{\sigma_0}}{\Gamma(\sigma_1)}$$

で与えられる．これらの目的関数をニュートン法に基づいて最大化することで，$\boldsymbol{\eta},\boldsymbol{\nu},\boldsymbol{\sigma}$ の更新式が導出される．

§5 潜在変数の事後確率

上記の手法により，$K,\boldsymbol{\mu},W,\boldsymbol{\gamma}$ の推定値 $K^*,\boldsymbol{\mu}^*,W^*,\boldsymbol{\gamma}^*$ をそれぞれ得る．このとき，事後確率，

$$\theta_{\alpha,k} = P(z(\alpha)=k \,|\, \mathcal{T}(T), \boldsymbol{\mu}^*, \boldsymbol{\gamma}^*, W^*, \beta)$$

が推定できる ($\alpha\in\mathcal{A}, k=1,\ldots,K^*$)．よって，各 $\alpha\in\mathcal{A}$ に対して K^* 次元離散確率分布，

$$\boldsymbol{\theta}_\alpha = (\theta_{\alpha,1},\ldots,\theta_{\alpha,K^*})$$

が得られ，$\Theta = \{\boldsymbol{\theta}_\alpha; \alpha\in\mathcal{A}\}$ とする．また，事後確率，

$$\xi_{n,i} = P(x_n = i \,|\, \mathcal{T}(T), \Theta, \boldsymbol{\mu}^*, \boldsymbol{\gamma}^*, W^*)$$

が推定できる ($n=1,\ldots,N(T)$, $i=0,1,\ldots,n-1$)．したがって，各 $n=1,\ldots,N(T)$ に対して n 次元離散確率分布，

$$\boldsymbol{\xi}_n = (\xi_{n,0},\xi_{n,1},\ldots,\xi_{n,n-1})$$

を得る．

Algorithm 1 CHP モデルによる期間 $[T,T_1)$ での共有イベント系列の生成

Input: Observed sequence $\mathcal{T}(T)$, and estimated parameters $\Theta,\boldsymbol{\mu}^*,\boldsymbol{\gamma}^*,W^*$
Output: Share-event sequence during $[T,T_1)$, $\mathcal{S}=\{(t_n,\alpha_n); n=N(T)+1,\ldots,N(T_1)\}$

Initialize $\mathcal{S}\leftarrow\emptyset$, $n\leftarrow N(T)$ and $s\leftarrow T$
Compute $\lambda^*_\alpha(s\,|\,\Theta)$ for $\forall\alpha\in\mathcal{A}$ by Eq. (12)
Set $\Lambda_0 \leftarrow \sum_{\alpha\in\mathcal{A}} \lambda^*_\alpha(s \mid \Theta)$
loop
  Sample $\Delta \sim \mathrm{Exponential}(\Lambda_0)$
  Set $s \leftarrow s+\Delta$
  if $s \geq T_1$ **then**
    break
  end if
  Compute $\lambda_\alpha(s\,|\,\Theta)$ for $\forall\alpha\in\mathcal{A}$ by Eq. (12)
  Set $\Lambda_1 \leftarrow \sum_{\alpha\in\mathcal{A}} \lambda^*_\alpha(s \mid \Theta)$
  Sample $d \sim \mathrm{Uniform}(0,1)$
  if $d \leq \Lambda_1/\Lambda_0$ **then**
    Set $t_{n+1}\leftarrow s$
    Sample $\alpha_{n+1} \sim \mathrm{Multinomial}\left(\bar{\boldsymbol{\lambda}}^*(t_{n+1})\right)$
    Set $\mathcal{S}\leftarrow\mathcal{S}\cup\{(t_{n+1},\alpha_{n+1})\}$
    Set $n\leftarrow n+1$
  end if
  Set $\Lambda_0\leftarrow\Lambda_1$
end loop

§6 協調構造の抽出

共有イベントの観測系列 $\mathcal{T}(T)$ から推定された CHP モデルを用いて，ポピュラリティダイナミクスの観点から $\mathcal{A}$ における協調構造を次のように抽出する．各アイテム $\alpha\in\mathcal{A}$ に対して $z(\alpha)$ を，

$$z^*(\alpha) = \operatorname*{argmax}_{1\leq k\leq K^*} \theta_{\alpha,k}$$

で推定し，各 $k=1,\ldots,K^*$ に対して，

$$\mathcal{A}_k = \{\alpha\in\mathcal{A}; z^*(\alpha)=k\}$$

とする．5·4 節において我々は，料理レシピ共有サイトの実データにおける協調構造 $Z^*=\{z^*(\alpha); \alpha\in\mathcal{A}\}$ の抽出を試みる．

4·2 予測の枠組み

共有イベントの観測系列 $\mathcal{T}(T)$ から推定された CHP モデルを用いて，近い将来 $[T,T_1)$ において発生する共有イベントの系列を予測する枠組みを与える．ここに，$T_1 > T$ である．ここでは特に，シミュレーションに基づく手法を採用する．

我々のシミュレーションアルゴリズムは期間 $[T,T_1)$ 内の共有イベント系列，

$$\mathcal{S} = \mathcal{T}(T_1)\setminus\mathcal{T}(T) = \{(t_n,\alpha_n); n=N(T)+1,\ldots,N(T_1)\}$$

を生成するものである．まず，推定されたパラメータ Θ, $\boldsymbol{\mu}^*, \boldsymbol{\gamma}^*, W^*$ に対して CHP モデルの強度関数は，

$$\lambda^*_\alpha(t \mid \Theta) = \mu^*_\alpha + \sum_{(t_n,\alpha_n)\in\mathcal{T}(t)} \sum_{k=1}^{K^*} w^*_{\alpha,k}\theta_{\alpha_n,k} \exp\{-\gamma^*_k(t-t_n)\} \tag{12}$$

で与えられることに注意する．我々は，多変量 Hawkes 過程の有効なシミュレーション法としてよく知られている Ogata のアルゴリズム [Ogata 81] に基づいて，$[T,T_1)$ 内で CHP モデルをシミュレートし，目標の共有イベント系列 $\mathcal{S}$ を生成する．Algorithm 1 にその生成手順の要約を示す．5 行目では，平均 $1/\Lambda_0$ の指数分布から時間間隔 Δ を生成する．12 行目では，台 $[0,1]$ の一様分布から d を生成する．15 行目では，$|\mathcal{A}|$ 次元離散確率分布，

$$\bar{\boldsymbol{\lambda}}^*(t_{n+1}) = \left(\bar{\lambda}_\alpha(t_{n+1} \mid \Theta)\right)_{\alpha\in\mathcal{A}}$$

から α_{n+1} を生成する．ここに，各 $\bar{\lambda}^*_\alpha(t \mid \Theta)$ は，

$$\bar{\lambda}^*_\alpha(t \mid \Theta) = \frac{\lambda^*_\alpha(t \mid \Theta)}{\sum_{\alpha'\in\mathcal{A}} \lambda^*_{\alpha'}(t \mid \Theta)} \tag{13}$$

で定義され，総和が 1 となるように正規化された強度関数を表している．$\bar{\lambda}^*_\alpha(t \mid \Theta)$ は，共有イベントが発生する時刻 t が与えられたとき，共有されるアイテムとして α が選択される確率を表していることに注意しておく．

5. 評 価 実 験

本章では，人工データおよび料理レシピ共有サイト Cookpad の実データを用いて，CHP モデルの評価を行う．まず，両データにおけるポピュラリティ予測性能に関して，CHP モデルを従来の Hawkes 過程モデルである HP モデルと MHP モデルと比較する．次に，CHP モデルに基づいて，ポピュラリティダイナミクスの観点から Cookpad データにおける協調構造の分析を試みる．

5·1 評 価 指 標

予測期間 $[T,T_1)$ において予測するべき共有イベント系列を $\mathcal{T}([T,T_1)) = \mathcal{T}(T_1) \setminus \mathcal{T}(T)$ とする．Hawkes 過程モデルの予測性能の評価によく用いられる方法に従い，予測したい共有イベントの発生時刻までのすべての共有イベントを与えて予測性能を評価する．実験では，以下の一般的な 3 つの評価指標を用いた．

- **PL (Prediction log-likelihood) 指標:** PL 指標では，将来の共有イベント群 $\mathcal{T}([T,T_1))$ の尤度，すなわち，予測したい共有イベントにおいて実際に発生した時刻とアイテムが選ばれる尤度を測定する [Farajtabar 17, Iwata 13, Zhou 13a, Zhou 13b]．任意のポアソン過程モデルの強度関数 $\lambda_\alpha(t)$ に対して，PL 指標は，

$$PL = \sum_{(t_n,\alpha_n)\in\mathcal{T}([T,T_1))} \ln\lambda_{\alpha_n}(t_n) - \int_T^{T_1} \sum_{\alpha\in\mathcal{A}} \lambda_\alpha(t)dt$$

と定義される．これはポアソン過程における $\mathcal{T}([T,T_1))$ の対数尤度と同じ形であることに注意しておく (式 (5) 参照)．

- **LI (Log-likelihood of items) 指標:** PL 指標と異なり LI 指標では，任意の将来の共有イベント $(t_n,\alpha_n) \in \mathcal{T}([T,T_1))$ に対して，その時刻 t_n が与えられたときにアイテム α_n が選ばれる確率，すなわち，将来の共有イベントの時刻が与えられた際にどのアイテムが共有されるかを予測する性能を測定する [Iwata 13]．LI 指標は強度関数 $\lambda_\alpha(t)$ を用いて，

$$LI = \sum_{(t_n,\alpha_n)\in\mathcal{T}([T,T_1))} \ln\frac{\lambda_{\alpha_n}(t_n)}{\sum_{\alpha'\in\mathcal{A}}\lambda_{\alpha'}(t_n)}$$

と定義される．ここに，対数の真数はそれらの総和が 1 となるように正規化されている (式 (13) 参照)．

- **AR (Average rank) 指標:** 将来の共有イベントの時刻が与えられたとき，LI 指標がその時刻に実際に発生した真のアイテムの生起確率を測定するのに対し，AR 指標ではより厳格に，その時刻で予測モデルが真のアイテムを選ぶ順位を測定する [Farajtabar 17]．AR 指標は，

$$AR = \frac{1}{|\mathcal{T}([T,T_1))|} \sum_{(t_n,\alpha_n)\in\mathcal{T}([T,T_1))} rank(\alpha_n;t_n)$$

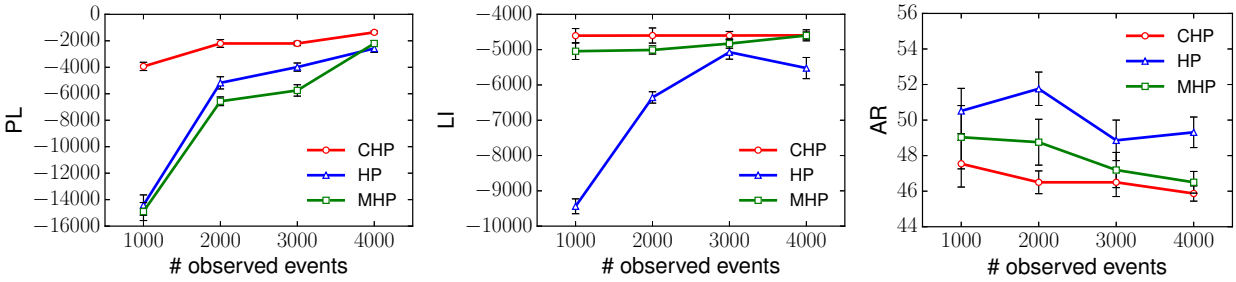
と定義される．ここに，$rank(\alpha;t)$ は強度関数 $\lambda_\alpha(t)$ に従って時刻 t における α の順位を返す関数である．

予測実験では，CHP モデルの強度関数として式 (12) を用いる．

5·2 人工データによる予測性能の評価

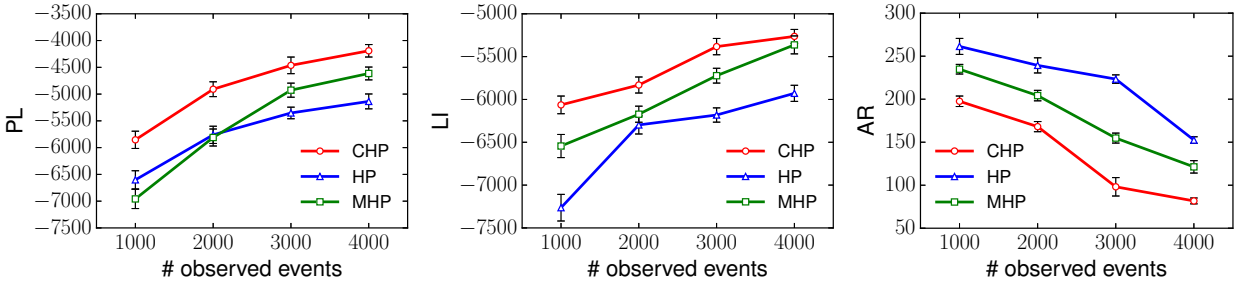
CHP モデルを導入する意義とその学習法の有効性を検証するために，まず，CHP モデルから生成したアイテム群の共有イベント系列の人工データを用いて，CHP モデルを 3 章で述べた既存の Hawkes 過程モデルである HP モデルおよび MHP モデルと予測性能の観点から比較した．

アイテム数を $|\mathcal{A}| = 100$，協調グループ数を $K = 10$ と設定し，Algorithm 1 に基づいてアイテム群の共有イベント系列の人工データを生成し，4 つのデータセットを構築した．ここに，観測データ $\mathcal{T}(T)$ は $|\mathcal{T}(T)| = 1000, 2000, 3000, 4000$ であり，それぞれに対し予測データ $\mathcal{T}([T,T_1))$ は $|\mathcal{T}([T,T_1))| = 1000$ である．HP モデルおよび MHP モデルの学習は，CHP モデルの学習と同様な手法を用いた．また，これら 3 つのモデルともパラメータ推定におけるサンプリングでは，200 回の burn-in を含む 1,000 回の反復を行った．



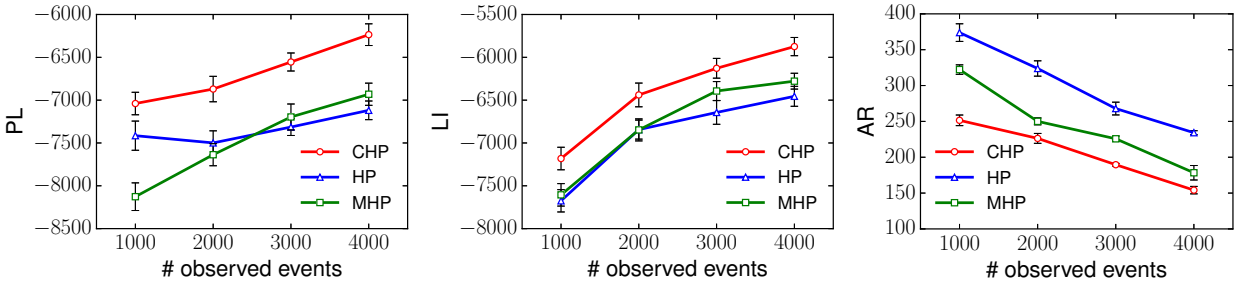
(a) PL 指標に関する結果　(b) LI 指標に関する結果　(c) AR 指標に関する結果

図 3　人工データにおける予測性能の比較



(a) PL 指標に関する結果　(b) LI 指標に関する結果　(c) AR 指標に関する結果

図 4　Cookpad の期間 1 データセットにおける予測性能の比較



(a) PL 指標に関する結果　(b) LI 指標に関する結果　(c) AR 指標に関する結果

図 5　Cookpad の期間 2 データセットにおける予測性能の比較

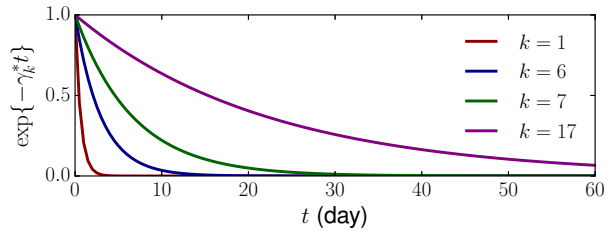
PL 指標，LI 指標，AR 指標により，CHP モデル，HP モデルおよび MHP モデルの予測性能を比較した．5 回の試行における結果を図 3 に示す．既存モデルは観測イベント数が少ない場合に性能が劣化したのに対し，提案モデルは常に最も性能が高かった．このことは，提案する CHP モデルは既存の Hawkes 過程モデルでは捉えるのが困難な新たな性質を有していること，および開発した CHP モデルの学習法は有効であることを示している．ところで，観測イベント数 4,000 のデータセットに対しては CHP モデルと MHP モデルの性能差は小さくなったが，これは，アイテム数に対して十分に多数の共有イベントが観測されたならば，原理的には CHP モデルを包含している MHP モデルが，観測データから CHP モデルを同定できる可能性があることを示唆している．しかしながら，そのような大量データを獲得することは，一般には困難であることに注意しておく．

5·3　実データによる予測性能の評価

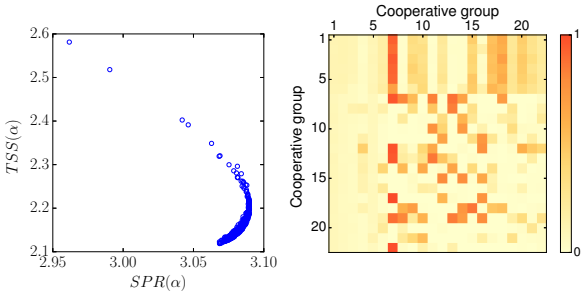
次に，実データを用いて，提案モデルである CHP モデルと既存モデルである HP モデルおよび MHP モデルを予測性能 (PL, LI, AR) の観点から比較した．実験では，料理レシピ共有サイト Cookpad [*1]の実データを用いた．Cookpad においてユーザは創作した料理レシピを投稿でき，また，別のユーザはそれらのレシピに賛意を表すメッセージ「つくれぽ」を投稿できる．ここでは，料理レシピとそれに対する賛意メッセージをそれぞれアイテムとその共有イベントとみなした．

共有イベント発生回数の日々の変動に関する定常性を考慮して，Cookpad の 2007 年データを対象とした．2007 年を期間 1 (1 月 1 日から 6 月 30 日) と期間 2 (7 月 1 日から 12 月 31 日) の 2 つの期間に分割し，以下の手順でデータセットを構築した．まず，各期間において，最初の 1 カ月間 (期間 1 では 1 月 1 日から 31 日，期間 2 では 7 月 1 日から 31 日) に 5 回以上共有されたアイテム

*1　https://cookpad.com/



(a) 抽出された協調グループの指数減衰関数



(b) $SPR(\alpha)$ と $TSS(\alpha)$ の関係　(c) $R=(R_{k,\ell})$ の可視化結果

図 6　協調構造の分析結果

表 1　抽出された協調グループの代表レシピ

グループ	タイトル
$\mathcal{A}_1$	ナンプラーそうめん@タイ風
	そうめん☆彡
	夏バテ知らずの☆そうめん
$\mathcal{A}_6$	レンジで簡単ピクルス
	とうもろこしは皮ごと簡単レンジでチン！
	ノンオイル☆ノンフライ☆ポテトチップス
$\mathcal{A}_7$	揚げない炒めない合せ調味料なし☆麻婆なす
	おろし玉ねぎでしょうが焼き
	豚バラに、甘酢ネギ胡麻だれ。
$\mathcal{A}_{17}$	超気持ちイイ レタスの芯のとり方♪
	オーブンいらずのふわふわパン
	無駄のない グレープフルーツの剥き方♪

表 2　影響を受けやすいレシピ

タイトル	$z^*(\alpha)$	$\mathrm{argmax}_k w^*_{\alpha,k}$	手順数
梅ツナうどん	5	$k=6$	2
簡単！失敗知らず【本格】カルボナーラ	3	$k=6$	6
本当に簡単すぎます！ココアおからケーキ	5	$k=6$	2

を対象とした．期間 1 データセットおよび期間 2 データセットのアイテム数は，それぞれ 620 および 1,095 であった．次に，各期間 $j=1,2$ に対して，最初から順に対象アイテムの共有イベントを調べていくことにより，4 つのデータセット $\mathcal{D}_1^j,\mathcal{D}_2^j,\mathcal{D}_3^j,\mathcal{D}_4^j$ を構築した．ここに，人工データによる実験と同様，観測データ $\mathcal{T}(T)$ は $|\mathcal{T}(T)|=1000, 2000, 3000, 4000$ であり，それぞれに対し予測データ $\mathcal{T}([T,T_1))$ は $|\mathcal{T}([T,T_1))|=1000$ である．例えば $\mathcal{D}_4^2$ は，期間 2 での 1,095 アイテムに対するイベント数 4,000 の観測データ $\mathcal{T}(T)$ とイベント数 1,000 の予測データ $\mathcal{T}([T,T_1))$ から構成されている．各モデルの学習については，人工データによる実験の場合と同じ設定で行った．

期間 1 データセットと期間 2 データセットに対する 5 回の試行結果をそれぞれ，図 4 と 図 5 に示す．CHP モデルは，HP モデルと MHP モデルよりも常に高性能であった．特に，3 つの指標の中で最も基本となるものであり，将来の共有イベントの尤度を評価する PL 指標に注目すると，観測イベント数が少ない場合やアイテム数に比べて観測イベント数が少ないような期間 2 データセットにおいては，CHP モデルと MHP モデルの性能差はより顕著であることが観察される．これらは，Cookpad における共有イベント発生過程が相互エキサイテーション性を有し，CHP モデルがそうした性質を過学習することなく捉えられること，そして，協調構造を組み込むことに意義があったことを示唆している．したがって，アイテム間の相互作用の構造を適切に組み込んだ CHP モデルの有効性が実証された．

5·4　実データにおける協調構造の分析

CHP モデルに基づいて，ポピュラリティダイナミクスの観点から Cookpad における協調構造を分析した．紙数の関係上，ここでは Cookpad の期間 2 データセットに対する結果のみを報告する．

協調グループ数の推定値 K^* は 22 であった．分析の容易化のために，協調グループのインデックス $k=1,\ldots,22$ を γ_k^* が大きい順に並べ替えた．γ_k^* の値が小さくなるほど，指数減衰関数における減衰は緩やかになることに注意しておく．抽出された協調グループ群において，4 つの代表的な指数減衰関数 ($k=1,6,7,17$ における指数減衰関数) を図 6 (a) に示す．また，これらの協調グループ $\mathcal{A}_1,\mathcal{A}_6,\mathcal{A}_7,\mathcal{A}_{17}$ に対して，$\theta_{\alpha,k}$ の値に基づくランキングの上位 3 位までの代表的レシピを表 1 に示す．まず，$\mathcal{A}_1$ に属するレシピは日本の夏の人気メニューである，そうめんに関連したものであり，それらの影響は最も急速に減衰していた．$\mathcal{A}_6$ に属するレシピは電子レンジで簡単かつ短時間で調理されるレシピに関連しており，それらの影響は急速に減衰するものの，$\mathcal{A}_1$ よりは緩やかであった．$\mathcal{A}_7$ に属するレシピは日本の家庭料理の惣菜に関連したものであり，それらの影響は $\mathcal{A}_6$ よりもさらに緩やかに減衰していた．$\mathcal{A}_{17}$ に属するレシピは効率よく料理をするための新規技法と関連したものであり，それらは長期的に影響を与え続けていた．

次に，共有イベントの相互エキサイテーション性の観点から，各料理レシピ α の影響の受けやすさについて分析した．分析に際し，$w^*_{\alpha,1},\cdots,w^*_{\alpha,22}$ を用いた 2 つのスコアを導入する．まず，異なる協調グループからの影響のばらつきを測定するスコア，$SPR(\alpha)=-\sum_{k=1}^{K^*}\bar{w}^*_{\alpha,k}\ln\bar{w}^*_{\alpha,k}$

表 3 影響力の強いレシピの抽出結果

タイトル	出次数
無駄のない グレープフルーツの剥き方♪	694.7
超気持ちイイ レタスの芯のとり方♪	436.5
豚と茄子の生姜マヨ炒め	390.1

を導入する．ただし，$\bar{w}^*_{\alpha,k} = w^*_{\alpha,k} / \sum_{\ell=1}^{K^*} w^*_{\alpha,\ell}$ である．次に，影響の受けやすさの度合いを測定するスコア，$TSS(\alpha) = \sum_{k=1}^{K^*} w^*_{\alpha,k}$ を導入する．図 6 (b) は，各レシピ α に対して $SPR(\alpha)$ と $TSS(\alpha)$ の値をプロットしたものである．興味深いことに，ほとんどのレシピでは $SPR(\alpha)$ と $TSS(\alpha)$ の間に正の相関が見られたが，$TSS(\alpha)$ の値が大きい少数のレシピ (図の左上に位置するレシピ) では $SPR(\alpha)$ の値が特に小さかった．すなわち，特定の協調グループのみから集中的に影響を受けるレシピは，過去の共有イベントから影響を受ける度合いが高い (すなわち，過去の共有イベントに敏感である) という傾向が示唆される．そこで，このような性質を持つレシピについてさらにその特徴を調べた (表 2 参照)．これらのレシピは，少数の手順で簡易に料理することを目指したもので，紹介文においても簡単さが強調されていた．またこれらは，$\mathcal{A}_6$ から最も強い影響を受けていた．

さらに我々は，$K^* \times K^*$ 行列 $R = (R_{k,\ell})$ を導入し，共有イベントの相互エキサイテーション性の観点から協調グループ間における影響の傾向を分析した．ここに，$R_{k,\ell}$ は $\mathcal{A}_\ell$ から $\mathcal{A}_k$ のレシピが受ける影響度の平均値を表し，$R_{k,\ell} \propto \sum_{\alpha\in\mathcal{A}} \theta_{\alpha,k} w^*_{\alpha,\ell} \ / \sum_{\alpha\in\mathcal{A}} \theta_{\alpha,k}$ で定義される．ただし，$\max_{1\le k,\ell\le K^*} R_{k,\ell} = 1$ と正規化する．図 6 (c) に，$R = (R_{k,\ell})$ の可視化結果を示す．k 行 ℓ 列目の成分の色が $R_{k,\ell}$ の値を表している．$R_{k,\ell}$ は $(k,\ell) = (17,7)$ のとき最大であり，$\mathcal{A}_{17}$ の新規の効率的な料理技法に関するレシピは，$\mathcal{A}_7$ の家庭料理の惣菜に関するレシピから影響を受けやすいという傾向が示唆される．

5·5 実データにおけるアイテム間の影響分析

5·4 節では，協調グループ間の影響関係を分析した．本節では，X の事後分布の推定値 $\boldsymbol{\xi}_1, \cdots, \boldsymbol{\xi}_{N(T)}$ を用いて，個々のアイテム間の影響関係を分析する．ここでも同様に，Cookpad の期間 2 データセットに対する結果のみを報告する．

推定値 $\xi_{n,i}$ は第 i 共有イベント (t_i, α_i) がその後の第 n 共有イベント (t_n, α_n) を誘発した確率を表しているので，それはレシピ α_i からレシピ α_n への影響を定量化していると考えられる．したがって，レシピ α' からレシピ α への影響スコア $INF(\alpha;\alpha')$ を，$INF(\alpha;\alpha') = \sum_{(t_n,\alpha_n)\in\mathcal{T}(T)} \mathbb{I}[\alpha_n = \alpha] \sum_{i=1}^{n-1} \xi_{n,i} \mathbb{I}[\alpha_i = \alpha']$ で定義する．我々は，レシピをノードとしレシピからレシピへの影響スコアをリンク重みとする重みつき有向ネットワークによって，レシピ間の影響関係のネットワークを構築し，ノードの出次数を用いて影響力が強いレシピを抽出した．表 3 に影響力が強いレシピの抽出結果を示す．このように，CHP モデルを用いることで，ポピュラリティダイナミクスの観点からソーシャルメディアサイトにおける影響力の強いアイテムを抽出することができる．

6. ま と め

本論文では，ソーシャルメディアサイトを対象とし，共有イベント系列に基づいてオンラインアイテム群の協調構造を抽出するために，新たな確率過程モデルである CHP モデルを提案し，各アイテムの将来ポピュラリティを予測することを試みた．CHP モデルは，Hawkes 過程にディリクレ過程を新たなやり方で組み込むことにより構築され，アイテムに対する共有イベントの時系列を生成する．我々は，共有イベントの観測系列から CHP モデルを推定する効率的なベイズ学習法を開発し，さらに，多変量 Hawkes 過程に対する Ogata のアルゴリズムを拡張することにより，CHP モデルの下で将来の共有イベントを予測する有効な枠組みを与えた．

人工データおよび Cookpad データを用いた実験において，PL，LI，AR の 3 つの指標の観点から，CHP モデルとアイテム間に相互作用のない Hawkes 過程 (HP) モデルおよび多変量 Hawkes 過程 (MHP) モデルとの予測性能を比較した．そして，CHP モデルは HP モデルと MHP モデルよりも予測性能が高いこと，特に，観測された共有イベント数が少ない場合には MHP モデルとの性能差がより顕著になることを実証し，協調構造を考慮する CHP モデルの有効性を示した．また，Cookpad データを用いた実験では，CHP モデルに基づいて，ポピュラリティダイナミクスの観点から，Cookpad における料理レシピ間の協調構造を同定し，料理レシピの協調グループに関するいくつかの特徴的な性質を明らかにした．

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 JP17K00433 の助成を受けたものである．

◇ 参 考 文 献 ◇

[Aalen 08] Aalen, O., Borgan, O., and Gjessing, H.: *Survival and Event Histrory Analysis: A Process Point of View*, Springer (2008)

[Bandari 12] Bandari, R., Asur, S., and Huberman, B.: The pulse of news in social media: Forecasting popularity, in *Proceedings of ICWSM'12*, pp. 26–33 (2012)

[Chen 13] 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)

[Cheng 14] Cheng, J., Adamic, L., Dow, P., Kleinberg, J., and Leskovec, J.: Can cascades be predicted?, in *Proceedings of WWW'14*, pp. 925–936 (2014)

[Daneshmand 14] Daneshmand, H., Gomez-Rodriguez, M., Song, L., and Schölkopf, B.: Estimating diffusion network structures: Recovery conditions, sample complexity & soft-thresholding algorithm, in

Proceedings of ICML'14, pp. 793–801 (2014)
[De 16] De, A., Valera, I., Ganguly, N., Bhattacharya, S., and Rodriguez, M. G.: Learning and forecasting opinion dynamics in social networks, in *Proceedings of NIPS'16*, pp. 397–405 (2016)
[Farajtabar 14] Farajtabar, M., Du, N., Gomez-Rodriguez, M., Valera, I., Zha, H., and Song, L.: Shaping social activity by incentivizing users, in *Proceedings of NIPS'14*, pp. 2474–2482 (2014)
[Farajtabar 17] Farajtabar, M., Wang, Y., Gomez-Rodriguez, M., Li, S., Zha, H., and Song, L.: COEVOLVE: A joint point process model for information diffusion and network evolution, *Journal of Machine Learning Research*, Vol. 18, No. 41, pp. 1–49 (2017)
[Gao 15] Gao, S., Ma, J., and Chen, Z.: Modeling and predicting retweeting dynamics on microblogging platforms, in *Proceedings of WSDM'15*, pp. 107–116 (2015)
[Gomez-Rodriguez 10] Gomez-Rodriguez, M., Leskovec, J., and Krause, A.: Inferring networks of diffusion and influence, in *Proceedings of KDD'10*, pp. 1019–1028 (2010)
[Hawkes 71] Hawkes, A.: Spectra of some self-exciting and mutually exiting point process, *Biometrika*, Vol. 58, No. 1, pp. 83–90 (1971)
[Iwata 13] Iwata, T., Shah, A., and Ghahramani, Z.: Discovering latent influence in online social activities via shared cascade poisson processes, in *Proceedings of KDD'13*, pp. 266–274 (2013)
[Kempe 03] Kempe, D., Kleinberg, J., and Tardos, E.: Maximizing the spread of influence through a social network, in *Proceedings of KDD'03*, pp. 137–146 (2003)
[Linderman 14] Linderman, S. and Adams, R.: Discovering latent network structure in point process data, in *Proceedings of ICML'14*, pp. 1413–1421 (2014)
[Neal 00] 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)
[Ogata 81] Ogata, Y.: On Lewis' simulation method for point processes, *IEEE Transactions on Information Theory*, Vol. 27, No. 1, pp. 23–31 (1981)
[Pinto 13] Pinto, H., Almedia, J., and Goncalves, M.: Using early view patterns to predict the popularity of youtube videos, in *Proceedings of WSDM'13*, pp. 365–374 (2013)
[Shen 14] Shen, H., Wang, D., Song, C., and Barabási, A.-L.: Modeling and predicting popularity dynamics via reinforced Poisson processes, in *Proceedings of AAAI'14*, pp. 291–297 (2014)
[Szabo 10] Szabo, G. and Huberman, B.: Predicting the popularity of online content, *Communications of the ACM*, Vol. 53, No. 8, pp. 80–88 (2010)
[Wang 13] Wang, D., Song, C., and Barabási, A.-L.: Quantifying long-term scientific impact, *Science*, Vol. 342, No. 6154, pp. 127–132 (2013)
[Yang 11] Yang, J. and Leskovec, J.: Patterns of temporal variation in online media, in *Proceedings of WSDM'11*, pp. 177–186 (2011)
[Zhao 15] Zhao, Q., Erdogdu, M., He, H., Rajaraman, A., and Leskovec, J.: SEISMIC: A self-exciting point process model for predicting tweet popularity, in *Proceedings of KDD'15*, pp. 1513–1522 (2015)
[Zhou 13a] Zhou, K., Zha, H., and Song, L.: Learning social infectivity in sparse low-rank networks using multi-dimensional Hawkes processes, in *Proceedings of AISTATS'13*, pp. 641–649 (2013)
[Zhou 13b] Zhou, K., Zha, H., and Song, L.: Learning triggering kernels for multi-dimensional Hawkes processes, in *Proceedings of ICML'13*, pp. 1301–1309 (2013)

〔担当委員：奥 健 太〕

2017 年 10 月 5 日　受理

著 者 紹 介

松谷　貫司

2016 年龍谷大学理工学部電子情報学科卒業．2018 年同大学院理工学研究科電子情報学専攻修士課程修了．同年，西日本電信電話株式会社入社，現在に至る．

熊野　雅仁（正会員）

1991 年立命館大学理工学部基礎工学科卒業．1991 年龍谷大学理工学部実験助手．2008 年龍谷大学理工学部実験講師．現在に至る．メディア論，社会ネットワーク可視化分析の研究と教育に従事．博士 (工学)[神戸大学大学院工学研究科情報知能学専攻]．情報処理学会，電子情報通信学会，認知科学会，ACM, IEEE，他各会員．

木村　昌弘（正会員）

1989 年大阪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，2005 年退社．現在，龍谷大学理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，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，電子情報通信学会各会員．

斉藤　和巳（正会員）

1985 年慶應義塾大学理工学部数理科学科卒業．同年，日本電信電話株式会社入社．2007 年静岡県立大学経営情報学部教授．機械学習，複雑ネットワーク等の研究に従事．博士（工学）．電子情報通信学会，日本神経回路学会，日本応用数理学会，日本データベース学会，日本行動計量学会，観光情報学会各会員．

大原　剛三（正会員）

1995 年大阪大学大学院基礎工学研究科博士前期課程修了．1996 年日本学術振興会特別研究員 DC2．1997 年大阪大学産業科学研究所助手，同助教を経て，2009 年青山学院大学理工学部情報テクノロジー学科准教授，2017 年同教授．データマイニング，機械学習，社会ネットワーク分析の研究に従事．博士（工学）．IEEE，AAAI，情報処理学会，電子情報通信学会各会員．

元田　　浩（正会員）

1967 年東京大学大学院工学系研究科原子力工学専攻修士課程修了．同年，株式会社日立製作所に入社．原子力システムの設計，運用，診断，制御に関する研究に従事．1995 年退社．1996 年大阪大学産業科学研究所教授，2006 年退職．現在，米国空軍科学技術局東京オフィス科学顧問．大阪大学名誉教授．機械学習，知識獲得，知識発見，データマイニング，社会ネットワーク解析の研究に従事．工学博士．