POI訪問ダイナミクスの可視化分析

Visualization Analysis of POI Visit Dynamics

藤井輝　熊野雅仁　小堀聡　木村昌弘 *
Teru Fujii, Masahito Kumano, Satoshi Kobori, Masahiro Kimura

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

Abstract: 位置情報ベースのソーシャルネットワークサービス (LBSN) の普及にともない，都市観光での人々の足跡に関する大量データが蓄積されつつあり，このような大規模な時空間データを有効活用することに期待が高まっている．本研究では，観光プランニング支援や観光マーケティングへの活用基盤として，観光都市における POI(point-of-interest) 群への人々の訪問ダイナミクスの特徴を視覚的に分析できるシステムの構築を目指し，ソーシャルメディアのアイテムに対するアテンション到着過程の確率モデルである RPP モデルと DPM-RPP モデルを援用した，POI 訪問ダイナミクスの可視化分析法を提案する．Foursquare データを用いた実験において，東京およびニューヨークでの POI 群に対する訪問ダイナミクス分析のケーススタディを通し，提案法の有効性を検証する．

1　はじめに

近年，Foursquare Swarm や Facebook Places などの位置情報ベースのソーシャルネットワークサービス (LBSN) が普及し，都市観光での人々の足跡に関する大量データが蓄積されつつある．LBSN では，その都市の観光スポットや有名店舗などが POI(point-of-interest) として登録されており，それら POI 群の位置情報および，ユーザが POI にチェックインした時間情報が観測可能である．そして，このような大規模な時空間データに基づいて，観光都市における人々の行動の傾向や特徴を理解し，観光推薦システムや観光マーケティングに有効活用することが期待されている [1], [2], [3].

一方，Twitter や料理レシピ共有サイトなどのソーシャルメディアでは，投稿されたアイテムが他のユーザ達から共有されポピュラリティを獲得していくという現象がよく起こる．そして，このようなオンラインアイテムへのアテンション到着過程のモデリングが，ポピュラリティ予測，トレンド予測，マーケティングなどの観点から，近年，ソーシャルメディアマイニングの分野で注目されている [4]．個々のアイテムのポピュラリティ予測については，既存の時系列解析法を用いる研究 [4], [5] や，特徴量を導入して既存の回帰問題や分類問題として扱う研究 [6], [7], [8] が，これまで主に行われてきた．最近 Shen ら [9] は，RPP モデル (Reinforced Poisson Process model) 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とその学習法を与え，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．ここに，RPP モデルは，アイテムの魅力度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エイジング特性を表す対数正規的な時間緩和関数 [10] および，“rich-get-richer” 現象を表す線形的強化関数 [11] という 3 つの構成要素からなる非一様 Poisson 過程である．我々 [12] 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モデルとして，Dirichlet 過程 [13] と RPP モデルを融合した DPM-RPP モデル (Dirichlet Process Mixture - RPP model) とその学習法を提案した．そして，料理レシピ共有サイトデータにおけるアイテムのポピュラリティ予測に対して，DPM-RPP モデルの精度が RPP モデルを含む従来法のものよりも平均的に見てかなり高いことを示した．さらに，DPM-RPP モデルを用いて，エイジングタイプの観点から料理レシピ群の構造を分析した．

本研究では，観光プランニング支援や観光マーケティングへの活用基盤として，観光都市における POI 群への人々の訪問ダイナミクスの特徴を視覚的に分析可能なシステムの構築を目指し，DPM-RPP モデルと RPP モデルを用いた POI 訪問ダイナミクスの可視化分析法を提案する．提案可視化法では，各 POI に対して，そのポピュラリティ予測の観点から DPM-RPP モデルと RPP モデルを比較し，予測精度が高い方のモデルを用いて訪問ダイナミクスに関するその魅力度とエイジング特性を可視化する．まず，DPM-RPP モデルによる POI のエイジング特性の可視化については，推定されたそのエイジングタイプを表示する．一方，RPP モデ

*連絡先：龍谷大学理工学部電子情報学科
〒 520-2194 大津市瀬田大江町横谷 1-5
E-mail: kimura@rins.ryukoku.ac.jp

ルによる POI のエイジング特性の可視化については，推定されたその対数正規的な時間緩和関数の特徴を効果的に表現する可視化法を新たに与え，それを表示する．そして，Foursquare データ [1] を用いた実験において，東京およびニューヨークでの POI 群に対する訪問ダイナミクス分析のケーススタディを通し，提案法の有効性を検証する．

2 分析法

対象とする観光都市における POI 全体の集合を D とし，期間 $[0,T)$ におけるそれら POI 群への人々の訪問ダイナミクスに対して，RPP モデル [9] と DPM-RPP モデル [12] を用いた可視化分析の枠組みを与える．ここに，$[0,T)$ は分析対象とする期間を表し，$T>0$ はその期間の長さを表している．任意の POI $d \in D$ に対して，$[0,T)$ における d への訪問者の到着時刻の系列を，

$$\boldsymbol{t}^d(T) = \left\{t_i^d;\, i=1,\ldots,n_d(T)\right\}$$

とする．ここに，$n_d(T)$ は $[0,T)$ における d への訪問者数であり，t_i^d は d への第 i 番目の訪問者の到着時刻を表し，$0 \le t_1^d \le \cdots \le t_{n_d(T)}^d < T$ を満たしている．我々は，観測データ，

$$\mathcal{T}(T) = \left\{\boldsymbol{t}^d(T);\, d \in D\right\}$$

に基づいて，$[0,T)$ における D への訪問ダイナミクスを分析する．

2.1 RPP モデル

まず，各 $d \in D$ に対して，$\boldsymbol{t}^d(T)$ を RPP モデルによりモデル化することを考える．RPP モデルはレート関数 $\lambda_d(t)$ の非一様 Poisson 過程であり，$\lambda_d(t)$ は，d の魅力度を表すフィットネス η_d，d のエイジング特性を表す対数正規的な時間緩和関数 $g(t;\mu_d,\sigma_d)$ および，“rich-get-richer” メカニズムを表す線形的強化関数 $r_d(t)$ の積，

$$\lambda_d(t) \;=\; \eta_d\, g(t;\mu_d,\sigma_d^2)\, r_d(t)$$

で表される．ここに，$\eta_d>0$, $\mu_d \in \mathbb{R}$, $\sigma_d>0$ はパラメータであり，

$$g(t;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}\,t}\exp\left\{-\frac{(\log t-\mu)^2}{2\sigma^2}\right\} \qquad (1)$$

であり，$r_d(t)$ は，$t_{i-1}^d \le t < t_i^d$ に対して，

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

である．ただし，$\nu>0$ はハイパーパラメータである．

RPP モデルによる分析では，$d \in D$ ごとに $\boldsymbol{t}^d(T)$ が完全に独立にモデル化され，モデルパラメータ値が $d \in D$ ごとに異なり得ることに注意しておく．

2.2 DPM-RPP モデル

次に，$\mathcal{T}(T)$ を DPM-RPP モデルによりモデル化することを考える．DPM-RPP モデルでは，POI 群 D は完全に独立ではなくなんらかの関係性を持つと考えている．そして，各 POI の魅力度を表すフィットネスについては，RPP モデルと同様に POI ごとに異なり得ると仮定するが，POI のエイジング特性を表す時間緩和関数については，POI ごとに完全に異なるのではなく，いくつかのエイジングタイプに分かれると仮定する．すなわち，RPP モデルに D のエイジングタイプ構造を組み込むことにより，DPM-RPP モデルは構築される．また，エイジングタイプの総数 K も観測データから学習するために，DPM-RPP モデルには Dirichlet 過程 [13] が組み込まれている．DPM-RPP モデルでは，各 $d \in D$ に対して，$\boldsymbol{t}^d(T)$ は以下の様に生成される．

まず，$k=1,2,3,\ldots$ に対して，エイジングタイプ k の時間緩和関数 $g(t;\mu_k,\sigma_k^2)$ (式 (1) 参照) のパラメータ値 (μ_k,σ_k^2) を，次の Normal-Inverse-Gamma 分布から生成する．

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

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

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

と生成する．ここに，$\mathrm{Beta}(1,\gamma)$ 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

$$G \;=\; \sum_{k=1}^{\infty}\pi_k\,\delta_{(\mu_k,\sigma_k^2)}$$

と定義する．ここに，$\delta_{(\mu_k,\sigma_k^2)}$ は位置 (μ_k,σ_k^2) に集中した確率測度を表す．このとき G は，密度関数 $p(\mu,\sigma^2 \mid \tau_1,\tau_2,\tau_3,\tau_4)$ 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 Dirichlet 過程 $\mathrm{DP}(\gamma, p(\mu,\sigma^2 \mid \tau_1,\tau_2,\tau_3,\tau_4))$ 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，

$$z_d \mid G \;\sim\; G$$

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 Gamma(α, β) から，

$$\eta_d \mid \alpha, \beta \ \sim \ \mathrm{Gamma}(\alpha, \beta)$$

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

$$\lambda_d(t) \ = \ \eta_d \, g(t; \mu_{z_d}, \sigma_{z_d}^2) \, r_d(t)$$

で定義し，レート関数 $\lambda_d(t)$ の非一様 Poisson 過程から $\boldsymbol{t}^d(T)$ を生成する．

2.3 学習法と予測法

まず，各 $d \in D$ に対して，観測データ $\boldsymbol{t}^d(T)$ から RPP モデルを学習する (詳細は [9] を参照)．時間緩和関数のパラメータの推定値 $\theta_d^* = (\mu_k^*, \sigma_d^{*2})$ および．フィットネスに対する事前分布のハイパーパラメータの推定値 α^*, β^* が得られる．RPP モデルを用いるとき，任意の $t > T$ に対して，d への期間 $[0.t)$ での訪問者総数の予測値 $\langle x_0^d(t) \rangle$ は，

$$\begin{aligned}
&\langle x_0^d(t) \rangle \\
&= \int_0^\infty x_0^d(t) \, p(\eta_d \mid \boldsymbol{t}^d(T), \theta_d^*, \alpha^*, \beta^*) \, \mathrm{d}\eta_d \\
&= (n_d(T) + \nu) \\
&\quad \times \left(\frac{X_d\left(\theta_d^*\right) + \beta^*}{X_d\left(\theta_d^*\right) - Y_d\left(\theta_d^*\right) + \beta^*} \right)^{n_d(T) + \alpha^*} - \nu
\end{aligned}$$

で与えられる．ここに，

$$x_0^d(t) = (n_d(T) + \nu) \exp\left\{ \eta_d Y_d(\theta_d^*) \right\} - \nu$$

であり，また，

$$\begin{aligned}
X_d(\theta) = &(n_d(T) + \nu) \int_0^T g(t; \theta) \, \mathrm{d}t \\
&- \sum_{i=1}^{n_d(T)} \int_0^{t_i^d} g(t; \theta) \, \mathrm{d}t
\end{aligned}$$

であり，

$$Y_d(\theta) = \int_T^t g(s; \theta) \, \mathrm{d}s$$

である．

次に，観測データ $\mathcal{T}(T)$ から DPM-RPP モデルを学習する (詳細は [12] を参照)．エイジングタイプの総数 K^*，各エイジングタイプ k の時間緩和関数のパラメータの推定値 $\theta_k^* = (\mu_k^*, \sigma_k^{*2})$，フィットネスに対する事前分布のハイパーパラメータの推定値 α^*, β^* および，各 $d \in D$ が各エイジングタイプ k に属する事後確率，

$$\phi_{d,k} = P(z_d = k \mid \boldsymbol{t}^d(T), \Theta^*, \alpha^*, \beta^*, \gamma)$$

が得られる．ただし，$\Theta^* = \{\theta_1^*, \ldots, \theta_{K^*}^*\}$ である．DPM-RPP モデルを用いるとき，任意の $t > T$ に対して，各 $d \in D$ への期間 $[0.t)$ での訪問者総数の予測値 $\langle x_1^d(t) \rangle$ は，

$$\begin{aligned}
&\langle x_1^d(t) \rangle \\
&= \sum_{k=1}^{K^*} \phi_{d,k} \int_0^\infty x_1^d(t) \, p(\eta_d \mid \boldsymbol{t}^d(T), z_d = k, \theta_k^*, \alpha^*, \beta^*) \, \mathrm{d}\eta_d \\
&= (n_d(T) + \nu) \\
&\quad \times \sum_{k=1}^{K^*} \phi_{d,k} \left(\frac{X_d\left(\theta_k^*\right) + \beta^*}{X_d\left(\theta_k^*\right) - Y_d\left(\theta_k^*\right) + \beta^*} \right)^{n_d(T) + \alpha^*} - \nu
\end{aligned}$$

で与えられる．ただし，

$$x_1^d(t) = (n_d(T) + \nu) \exp\left\{ \eta_d Y_d(\theta_{z_d}^*) \right\} - \nu$$

である．また，DPM-RPP モデルでは，各 $d \in D$ に対して，

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

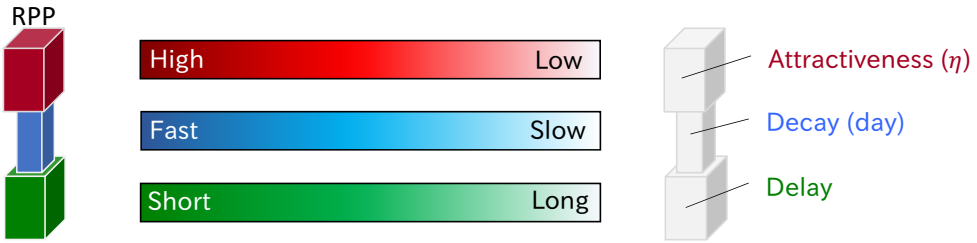
により，d のエイジングタイプ z_d^* を決定し，D のエイジングタイプ構造を抽出する．

2.4 可視化法

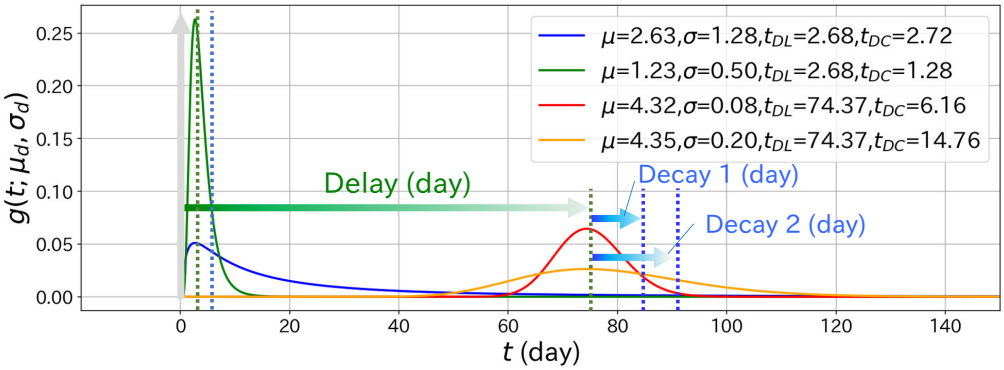
まず，$T' > T$ を固定し，各 $d \in D$ に対して，期間 $[0, T')$ での d への訪問者総数の予測を考える．そして，RPP モデルによる予測値 $\langle x_0^d(T') \rangle$ と DPM-RPP モデルによる予測値 $\langle x_1^d(T') \rangle$ を比較し，予測精度が高い方のモデルに基づいて d の魅力度とエイジング特性を可視化することにより，期間 $[0, T)$ における POI 群 D への訪問ダイナミクスの特徴を分析する．ここに，DPM-RPP モデルに従っている POI の訪問ダイナミクスは，エイジングタイプ構造の観点から D 内の他の POI 群の訪問ダイナミクスと関係していると考えられ，一方，RPP モデルに従っている POI は，D 内の他の POI 群との関係があまりない独自の訪問ダイナミクスをもっていると考えられる．

以下において我々は，POI 訪問ダイナミクスの主な特徴である魅力度とエイジング特性について，POI ごとにそれらを効果的に可視化する手法を与える．まず，DPM-RPP モデルに従っている POI に対して可視化する手法を与え，次に，RPP モデルに従っている POI に対して可視化する手法を与える．

(a) DPM-RPP モデルに従う POI の 3D glyph



(b) RPP モデルに従う POI の 3D glyph



(c) 時間緩和関数の例

図 1: POI のフィットネス（魅力度）とエイジング特性の可視化法

本研究では，各 POI が位置情報を持つことを有効活用し，地図システム上で相互の POI に付随する情報を比較可能とするため，POI に付随する情報を四角柱の組み合わせで表した立体的な 3D glyph として可視化する手法を考える．DPM-RPP と RPP モデルに従う POI を 3D glyph として表現する方法を図 1 に示す．

2.4.1 DPM-RPP モデルによる可視化

DPM-RPP モデルに従う POI 群は，同じ時間緩和関数を共有するいくつかのクラスタに分かれるが，地図上では相互に離れた位置に存在し得るため，同じクラスタに所属する POI であることを視認できるように，図 1a のような天井面の色が同じ四角柱を同一クラスタの POI として表す．また，DPM-RPP モデルに従う各 POI の魅力度を表すフィットネス η の高さを図 1a の濃い赤から白までのグラデーションで表す．フィットネスが高いほど四角柱の側面は濃い赤となり，フィットネスが低いほど白となる．これにより，同一クラスタに所属する DPM-RPP モデルに従う POI であっても，それぞれ異なるフィットネスを持つ POI として容易に識別することができる．なお，DPM-RPP モデルに従う POI では，クラスタごとに時間緩和関数が異なるが，クラスタ数は多くはないため，各クラスタの時間緩和関数は補助グラフにより参照可能にする．

2.4.2 RPP モデルによる可視化

RPP モデルに従う各 POI は，それぞれ異なる時間緩和関数を持つが，地図上に配置される POI の数が多いことから，各 POI の時間緩和関数の形状を地図上で容易に視認可能にする方法として図 1b のように，四角柱を三段縦に積み重ねる方法を用いる．図 1c は時間緩和関数の例を示している．緑色の時間緩和関数は遅延日数が短く急激に減衰し，青色の時間緩和関数は緑色のものと同じく遅延日数が短いが減衰は緩やかとなっている．また，赤色の時間緩和関数は遅延日数が長く急激に減衰し，黄色の時間緩和関数は赤色のものと同じく遅延日数が長いが減衰は緩やかとなっている．

それらの時間緩和関数の違いを可視化で視認可能とする上で，まず，時間緩和関数が極大となる日，

$$t_{DL} = \exp(\mu - \sigma^2)$$

を基準日 t=0 からの遅延（Delay）と定義し，3D glyph の下から 1 段目にある四角柱で表す．基準日からの日数が短いほど基準日から近い頃に訪問者が急増する傾向を示すことを意味し，基準日からの日数が長いほど日数が経過してから POI への訪問者が急増する傾向を示すことになる．言い換えれば，遅延日数が短いほど人気は早期に得られる傾向を示し，遅延日数が長いほど人気は晩期に得られる傾向を示している．ここでは，遅延日数が短いほど濃い緑となり，遅延日数が長いほど白となるように配色した．次に，人々が POI を訪問する行動がピークを迎えてから減衰する傾向を表すうえで，時間緩和関数が極大となる日から急激に減衰したり，ゆっくり減衰することを特徴として視認可能とするため，3D glyph の下から 2 段目の四角柱で減衰傾向を表す．時間緩和関数において，それが極大となる日から右側の変曲点までの日数，

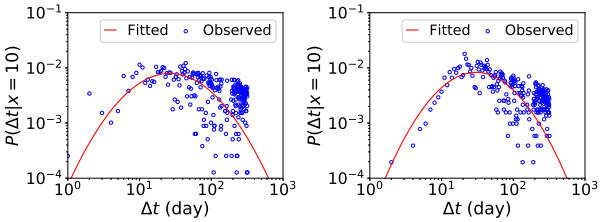
$$t_{DC} = \exp\left(\mu + \frac{-3\sigma^2 + \sigma\sqrt{\sigma^2 + 4}}{2}\right) - t_{DL}$$

をその減衰度（Decay）と定義する．t_{DC} が小さいということはピークからの速い減衰を表しており，それが大きいということはピークからゆっくりとした減衰を表している．ここでは，減衰が速いほど濃い青となり，ゆっくりであるほど白くなるように配色した．最後に，各 POI の魅力度を表すフィットネスは，最上段にある四角柱の側面および天井面を用いて図 1a と同様に，濃い赤から白からまでの色を配色して表す．

3 実験

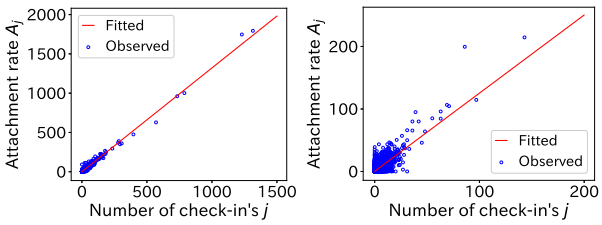
3.1 実験データ

Foursquare の check-in データに含まれる，東京とニューヨークにおける 2012 年 4 月 3 日から 2013 年 2 月 16 日までの実データを用いた．活発な POI に焦点を当てて動的分析を行うため，check-in 数が 10 件以上の POI を対象とし，データセットを構築した．そのデータセットをここでは Foursquare データセットと呼ぶ．図 2 は Foursquare データセットにおけるエイジング特性の観点から POI 訪問行動に関する動的分析を行った結果である．エイジング特性の検証においては初日の設定が必要となる．Foursquare の POI 訪問行動においては，史跡や海岸，洞窟の他，最初の日を決定し難い分析対象が多い．そこで，本研究では基準日を定めるうえで，Foursquare データセットの期間における初日をすべての POI 共通の基準日とした．先行研究 [10] と同様に，エイジング特性の観点から東京（図 2a），ニューヨーク（図 2b）の結果を見ると，ともに対数正規分布に従う傾向があることが見てとれる．本研究では時間緩和関数として対数正規分布を用いることが妥当であると考える．また，図 3 は”rich-get-richer”性の検証において，先行研究 [11] の手法を Foursquare データに適用した結果である．Attachment 比は，2012 年 4 月から 6 月の POI 訪問行動を調べた場合の check-in 数に対するその直後の一ヶ月間（7 月）の check-in 数との関係を示している．check-in 数に対する Attachment 比が明瞭な線形性を示していることから，Foursquare データの東京（図 3a）およびニューヨーク（図 3b）においても”rich-get-richer”性を確認することができる．



(a) 東京データ　(b) ニューヨークデータ

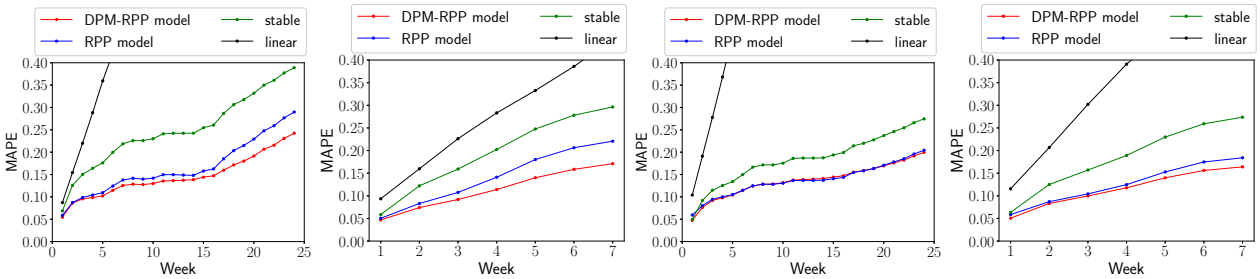
図 2: Foursquare データにおけるエイジング特性



(a) 東京データ　(b) ニューヨークデータ

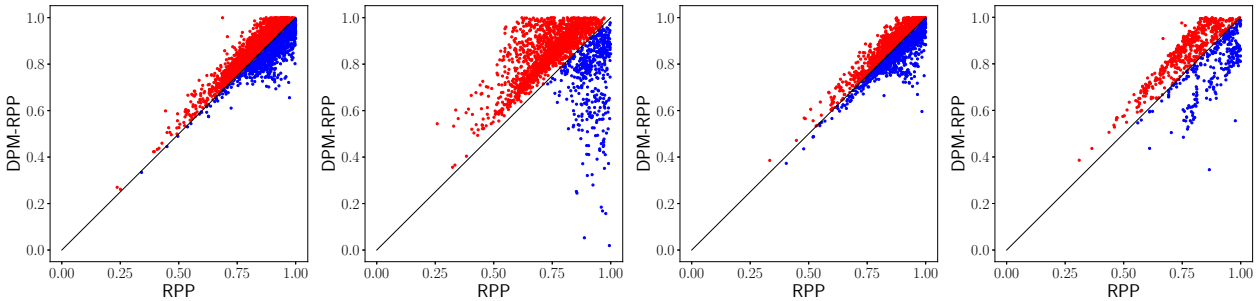
図 3: Foursquare データにおける”rich-get-richer”性

本研究では，動的分析における季節性の有無を調べるため，東京とニューヨークのデータに対して，4 月から 6 月までの春と 10 月から 12 月までの秋についての 2 つのデータセットを構築した．なお，東京データの春データセットにおける POI 数は 3,339，check-in 総数は 162,861，また，東京データの秋データセットにおける POI 数は 2,165，check-in 総数は 97,740 となった．次に，ニューヨークデータの春データセットにおける POI 数は 2,386，check-in 総数は 62,329，また，ニューヨークデータの秋データセットにおける POI 数は 801，check-in 総数は 16,829 となった．また，基準日については，春と秋それぞれのデータ上の初日とした．なお，先行研究 [12] に従ってパラメータ γ を 1 に設定し，ν を 10 に設定した．



(a) 東京-春データ (b) 東京-秋データ (c) ニューヨーク-春データ (d) ニューヨーク-秋データ

図 4: ポピュラリティ予測性能の比較



(a) 東京-春データ (b) 東京-秋データ (c) ニューヨーク-春データ (d) ニューヨーク-秋データ

図 5: DPM-RPP と RPP による POI のポピュラリティ予測精度

表 1: DPM-RPP に従う POI 数と RPP に従う POI 数.

	東京		ニューヨーク	
	春	秋	春	秋
DPM-RPP（赤）	2,151	1,678	1,236	527
RPP（青）	1,188	481	1,150	274

3.2 ポピュラリティ予測の結果

全ての POI に対して，期間 $[0, T')$ 内の訪問ダイナミクスを用いて，将来の t $(\geq T)$ 時間後までに獲得する check-in 総数 $x^d(t)$ の予測を行い，その性能を評価した．DPM-RPP および RPP モデル以外に，単純法として学習期間の最終時点から過去 1 週間の変化量に基づいて線形予測を行うモデル（linear）と，学習期間の最終時点から POI 訪問は変化しないと見なす定常予測モデル（stable）との比較を行う．我々は，先行研究 [12] と同様にハイパーパラメータを設定し，ポピュラリティ予測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．図 4 に，4 つのモデルに基づく MAPE を指標として用いたポピュラリティ予測の結果を示す．ニューヨークの春データについては DPM-RPP と RPP の予測性能に顕著な差は現れなかったものの，東京の春と秋のデータおよびニューヨークの秋データの結果においては $T' - T$ が大きくなるほど DPM-RPP は RPP よりも予測性能が高くなる傾向が見てとれる．

3.3 可視化分析の結果

可視化分析においては，各 POI ごとに DPM-RPP モデルと RPP モデルを適用したポピュラリティ予測において，より高い精度を示した方のモデルに対応する 3D glyph を使って可視化を行うことに注意しておく．なお，予測期間の長さが短いニューヨーク-秋データに合わせ，$T' - T$ を 6 週間とした．図 5 は，各 POI に対して，横軸に RPP モデルによる予測精度，縦軸に DPM-RPP モデルによる予測精度をとったポピュラリティ—予測精度の POI 分布であり，図 5a と図 5b はそれぞれ東京の春と秋のデータに関する結果，図 5c と図 5d はそれぞれニューヨークの春と秋のデータに関する結果である．それぞれの図において，赤い点は DPM-RPP の精度が高かった POI，青い点は RPP の精度が高かった POI を表している．表 1 は，図 5 において DPM-RPP の予測精度が RPP に対して上回った赤い POI の数と RPP の予測精度が DPM-RPP に対して上回った青い POI の数を東京およびニューヨークの春と秋のデータで比較したものである．いずれも，DPM-RPP に従う POI の数が多いことから，RPP に従う POI よりも DPM-RPP に従う POI が多いことが

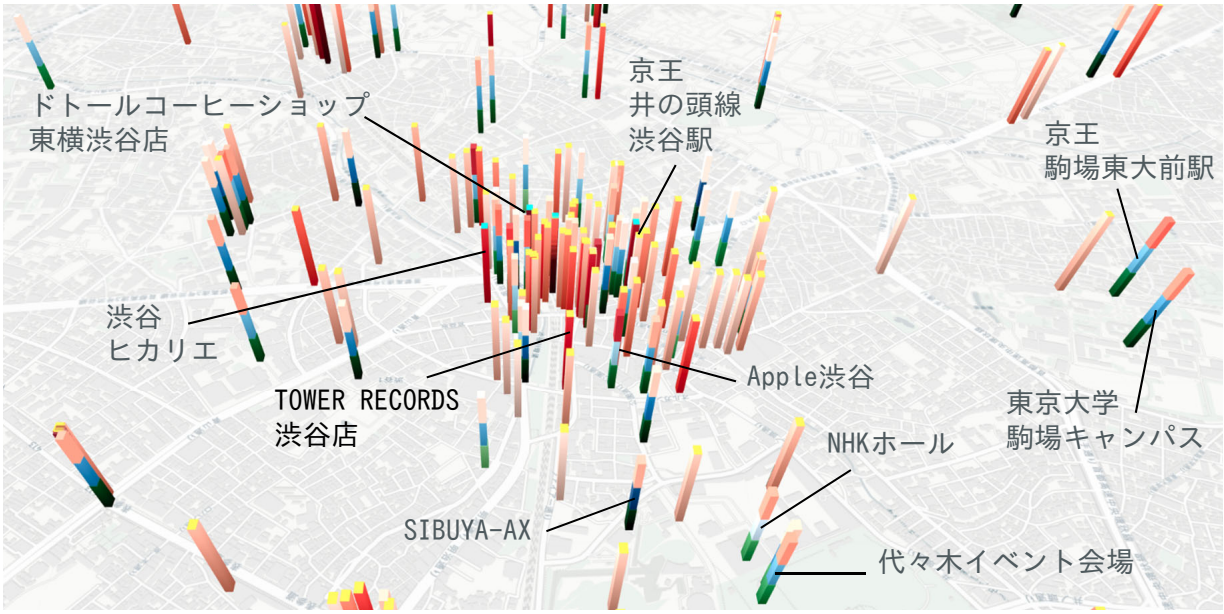


図 8: 東京-春データ（渋谷周辺）の可視化結果

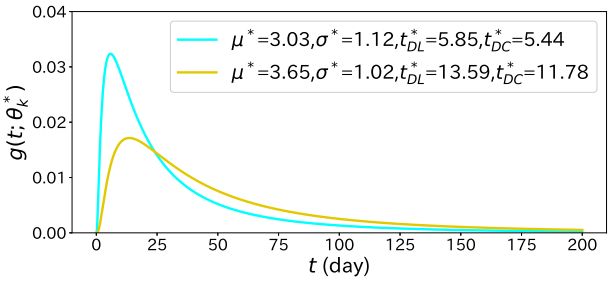


図 6: 主要クラスタの時間緩和関数（東京-春データ）

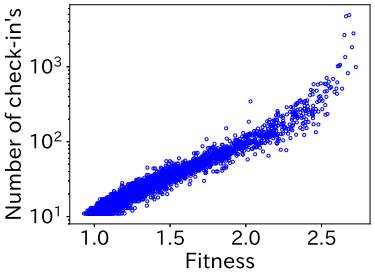


図 7: check-in 数とフィットネスとの関係（東京-春データ）

わかる．ここでは頁数の制限から，東京-春データに着目した可視化結果を用いて地図上での POI の地域分析を試みる．なお，3D glyph を描画する地図システムとして，OSM Buildings を用いた．

3.3.1 東京-春データの結果

図 6 は，DPM-RPP モデルに基づいて推定された異なるエイジングタイプに対応する時間緩和関数である．東京-春データでは，2 つの主要なエイジングタイプが得られた．黄色のエイジングタイプに従う POI 数は 2,990，水色のエイジングタイプに従う POI 数は 348 であった．また，黄色のエイジングタイプの遅延 t_{DL} が 13.59 日であるのに対して，水色のエイジングタイプでは遅延 t_{DL} が 5.85 日であり，時間緩和関数が極大となるまでの日が半分程度と遅延が短い傾向があることがわかる．また，DPM-RPP モデルに従う 2 つのエイジングタイプ構造の減衰傾向としては，水色の方が黄色の方よりも急峻に減衰することがわかる．すなわち，水色のエイジングタイプに従う POI 群は，もう 1 つのエイジングタイプに対して高いアテンションを獲得するがすぐに忘れ去られる傾向があることを示唆している．図 7 は，東京-春データに含まれる各 POI ごとに check-in 数とフィットネスの関係を示した散布図である．各 POI への訪問（check-in）数（人気度）はフィットネスと関係性があることがわかる．すなわち，フィットネスを魅力度と見なすのは妥当であると考える．

DPM-RPP モデルに従う 2 つのエイジングタイプそれぞれに所属する POI 群や RPP モデルに従う POI 群が実世界上でどのように分布しているかを確かめるために，各 POI に対応する 3D glyph を地図システム上に描画する．東京-春データに基づいて，渋谷周辺を可視化した結果が図 8 である．渋谷の中心街周辺には DPM-RPP における黄色のエイジングタイプに従う POI 群が多く，それらの多くの魅力度が低めであることがわか

る．ただし，TOWER RECORDS 渋谷店は黄色のエイジングタイプであるが魅力度が非常に高いことがわかる．京王井之頭線渋谷駅や渋谷ヒカリエ，ドトールコーピーショップ東横渋谷店は DPM-RPP における水色のエイジングタイプに従い，魅力度が高く，エイジングタイプ構造の観点から D 内の他の POI 群の訪問ダイナミクスと関係していると考えられる．一方，SIBUYA-AX というライブ会場は RPP に従い，3D glyph の 1 段目が濃い緑，すなわち遅延がかなり短く，2 段目が濃い青，すなわち減衰もかなり速い傾向があり，他の POI 群との関係があまりない独自の訪問ダイナミクスをもっていると考えられる．また，SIBUYA-AX に対して東京大学駒場キャンパスや京王駒場東大前駅，代々木イベント会場は 3D glyph の 1 段目が少し濃い緑，すなわち遅延が少し長めで，2 段目が水色，すなわち減衰も少し緩やかであるが，NHK ホールは 1 段目が緑，すなわち遅延がさらに長めで，2 段目が白，すなわち減衰は非常にゆっくりであることがわかる．

以上のように，提案法による可視化分析により，POI 群のそれぞれがどのような POI 訪問ダイナミクスを示すかという動的な観点から地域分析を可能にし得ることが示唆された．

4 まとめ

本研究では，観光プランニング支援や観光マーケティングへの活用基盤として，観光都市における POI 群への人々の訪問ダイナミクスの特徴を視覚的に分析できるシステムの構築を目指し，ソーシャルメディアのアイテムに対するアテンション到着過程の確率モデルである RPP モデルと DPM-RPP モデルを援用した，POI 訪問ダイナミクスの可視化分析法を提案した．Foursquare データを用いた実験において，東京およびニューヨークでの POI 群に対する訪問ダイナミクス分析のケーススタディを通し，動的なアテンション獲得過程について行動データ分析を行える可能性を示した．これらの結果は観光マーケティングにつながると考えている．

参考文献

[1] Yang, D., Zhang, D., Zheng, V.W., and Yu, Z.: Modeling user activity preference by leveraging user temporal characterisitics in LBSNs, *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, Vol. 45, No. 1, pp. 129–142 (2015).

[2] Chen, S., Yuan, X., Wang, Z., Guo, C., Liang, J., Wang, Z., Zhang, X.L., and Zhang, J.: Interactive visual discovering of movement patterns from sparsely sampled geo-tagged social media data, *IEEE Transactions on Visualization and Computer Graphics*, Vol. 22, No. 1, pp. 270–279 (2016).

[3] 相原健郎: ビッグデータを用いた観光動態把握とその活用：動体データで訪日外客の動きをとらえる, 情報管理, Vol. 59, No. 11, pp. 743–754 (2017).

[4] Szabo, G. and Huberman, B.: Predicting the popularity of online content, *Communications of the ACM*, Vol. 53, No. 8, pp. 80–88 (2010).

[5] Yang, J. and Leskovec, J.: Patterns of temporal variation in online media, *Proceedings of WSDM'11*, pp. 177–186 (2011).

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

[7] 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).

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

[9] 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).

[10] Wang, D., Song, C. and Barabási, A.-L.: Quantifying long-term scientific impact, *Science*, Vol. 342, No. 6154, pp. 127–132 (2013).

[11] Redner, S.: Citation statistics from 110 years of Physical Review, *Physics Today*, Vol. 58, pp. 49–54 (2005).

[12] 松谷貫司, 木村昌弘: 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, 情報処理学会論文誌「数理モデル化と応用」, Vol. 10, No. 2, pp. 14–24 (2017).

[13] 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).