

Foursquare データに基づいた観光地アクティビティの動的分析

森田紗椰^a 熊野雅仁^b 木村昌弘^c

龍谷大学理工学部電子情報学科

a) *t140181@mail.ryukoku.ac.jp* b) *kumano@rins.ryukoku.ac.jp* c) *kimura@rins.ryukoku.ac.jp*

概要 近年、非一様ポアソン過程に基づいて、科学論文の引用イベントやオンラインアイテムの共有イベントの到着過程を分析する研究が注目されている。ここでは、アイテムが新たなアテンションを獲得する能力に対して、時間減衰効果と rich-get-richer 現象のような強化効果の分析が主要な課題となっている。本研究では、そのような研究を拡張し、Foursquare データを用いて観光地へのアテンションの到着過程を分析することを考える。そして、観光地アクティビティの時間減衰効果および強化効果の性質を明らかにするとともに、それらを観光地の人気度、カテゴリおよび地理的位置の観点から総合的に分析し、観光マーケティングへの応用を模索する。

キーワード ソーシャルメディアマイニング, アテンションダイナミクス, 観光マーケティング

1 はじめに

近年、Facebook や Foursquare 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 (SNS) が普及し、都市での人々の足跡に関する大量データが蓄積されてきている。そして、このような大規模な時空間情報に基づいて、ユーザ行動の傾向や特長を理解し推薦システムを構築することについての関心が高まっている [1]。ところで、SNS で足跡を残すという行為はユーザの自由意志に任されているため、個々のユーザの足跡系列パターンは一般にスパースであり密度変化もランダムであり得るので、そのような時系列の分析自体はあまり意味がない場合が多い。我々は、個人の足跡系列ではなく、多くの人々の足跡を総合することによって得られる観光地へのアテンションの時系列を分析し、都市の観光マーケティングを支援することを目指している。

都市における観光地と同様、ソーシャルメディアやジャーナルに投稿されたオンラインアイテムや論文は、他から共有されたり引用されたりするなどのアテンションを得ることで人気を獲得していく。ソーシャルメディアにおけるこのような現象は社会のトレンド形成にも影響を及ぼし得るので、近年、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研究が注目されている [2, 3]。特に論文引用ダイナミクスについては、Redner [4] は “rich-get-richer” 現象が存在することを示し、Wang ら [2] はアイテムの新アテンション獲得能力に対するエイジング効果 (時間減衰効果) が対数正規分布に従うことを示した。さらに、Wang ら [2] や Shen ら [3] は、アイテムの魅力を表すフィットネス、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 つの要素から構成された非一様ポアソン過程である RPP (Reinforced Poisson Process) モデルを提案し、それが将来の引用数予測において従来法よりも精度が良いことを実験で示した。

本論文では、RPP モデルを拡張し観光地アテンションダイナミクスの有効なモデルを構築することを目指して、Foursquare の check-in データセット [1] に基づき、東京とニューヨークにある観光地へのアテンション到着過程を分析する。まず、Wang ら [2] の手法と Redner [4] の手法を用いて、観光地アクティビティにおけるエイジング効果と “rich-get-richer” 現象の存在について調べ、論文やオンラインアイテムのアテンション到着過程の場合と比較する。次に、観光地へのアテンション到着の時間間隔 Δt の分布を指数分布でモデル化して、指数分布の減衰率パラメータ λ を最尤推定し、それと人気度との関係について分析する。さらに、観光地のカテゴリや地理的位置との関係についても総合的に分析し、観光マーケティングへの応用を模索する。

2 実験設定

Foursquare の check-in データセット [1] に含まれる、ニューヨークと東京における 2012 年 4 月から 2013 年 2 月までの実データを用いた。エイジング効果の検証においては初日の設定が必要となる。論文など公開日を初日とできる分析対象とは異なり、Foursquare の観光地は、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、その日を含む n 日以内に check-in がなかった観光地を対象として、 $n+1$ 日目以降の最初の check-in 獲得を一人目として c 人目が check-in するまでの時間を分析対象の範囲に設定した。ニューヨークデータの観光地総数 5,135、check-in 総数 227,428、東京データの観光地総数 7,873、

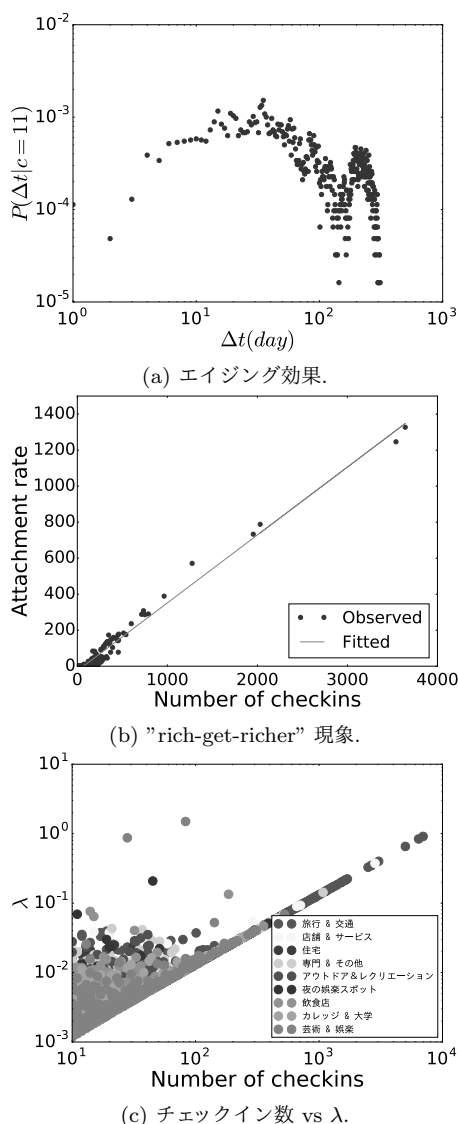
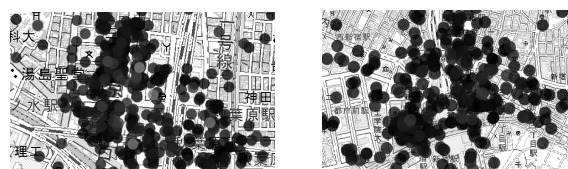


図 1: Foursquare の東京データにおける動的分析の結果

check-in 総数 573,703 であった。基準日は 2012 年 4 月 3 日 (水) とし, $n=4$ に設定した。その日は 4 月 6 日 (金) であり平日の末日となる。 c は $c=11$ と設定した。

3 実験結果

紙数の関係上, 東京データの結果のみを示す。図 1 は, エイジング効果に関する観光地アクティビティに関する動的分析の結果である。先行研究とは異なり, 対数正規分布ではなく, 特に二峰性の混合分布であるようにも見てとれる。図 1b は, "rich-get-richer" 現象の検証において, 先行研究の手法を適用した結果であり, 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 の東京データにおいても "rich-get-richer" 現象を確認することができる。図 1c は, 各観光地について check-in 数に対する減衰率パラメータ λ の分布を示した結果であり, check-in 数が少ない場合に λ



(a) 秋葉原駅周辺.

(b) 新宿駅周辺.

図 2: 観光地の減衰率パラメータ λ の可視化.

が大きくなる観光地があるが, check-in 数が多いとき正の相関をもつことがわかる。なお, 描画点の色は, 各観光地が属するカテゴリーの Top 階層 (10 種) に対応している。check-in 数 500 以上は鉄道駅が多く, 最も多い観光地は「旅行&交通」カテゴリーの秋葉原駅, 次が新宿駅であったが, 「店舗&サービス」カテゴリーのヨドバシカメラも含まれていた。check-in 数は少ないが λ の大きい Top 2 は「芸術&娯楽」カテゴリーのコミックマーケットやタイフェスティバルなどのイベントであった。図 2 は, 観光地の分布を示したものである。check-in 数が 500 以上の観光地はピンクとしたが, 我々はより多くの色 (カテゴリー) が含まれる check-in 数が 500 未満, λ が 10^{-1} 未満に注目しており, その範囲において λ が高いほど赤, 低いほど黒で表した。黒いものよりも赤いものはより短い間隔で check-in を獲得する活発な観光地である。秋葉原駅は新宿駅周辺よりもより活発な観光地が多いことがわかる。一方, Δt が長く (check-in が無い日がより長く続く場合がある) λ が低い観光地が新宿に多いこともわかる。

4 まとめ

Foursquare を対象として, アテンション到着過程を分析した。近年, 人々のリアル行動データを観光地創生に生かす手法が注目されている。本研究では, 足跡の量だけでなく, 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. *et al.*, "Modeling user activity preference by leveraging user spatial temporal characteristics in lbsns," *IEEE Trans. SMC*, vol. 45, no. 1, pp. 129–142, 2015.
- [2] Wang, D., Song, C., and Barabási, A.-L., "Quantifying long-term scientific impact," *Science*, vol. 342, pp. 127–132, 2013.
- [3] Shen, H., Wang, D., Song, C. *et al.*, "Modeling and predicting popularity dynamics via reinforced poisson processes," in *AAAI'14*, 2014, pp. 291–297.
- [4] Redner, S., "Citation statistics from 110 years of Physical Review," *Physics Today*, vol. 58, no. 6, pp. 49–54, 2005.