

Twitter の位置情報付きツイートを利用したイベントにおける ユーザーの行動分析

中里 主哉† 小林 えり‡ 齊藤 和巳† 風間 一洋†† 吉田 光男‡‡

† 静岡県立大学 経営情報学部 ‡ 静岡県立大学大学院 経営情報イノベーション研究科

†† 和歌山大学 システム工学部 ‡‡ 豊橋技術科学大学 情報・知能工学系

1 はじめに

Twitter では、自分がどこでツイートをしたかを情報として付与して発信する位置情報付きツイートを行うことができる。位置情報を基に「あるイベントに参加したユーザーがどのような行動をとっているのか」を解明することで、イベントでだけでなくその前後の交通、宿泊、飲食などのさまざまな点で効果的なサービスの提供が期待できるようになると考えられる。

本研究では、あるイベントに着目し、イベント参加者のツイートをを用いて、イベント参加者がよく訪れるスポットをツイート回数とユーザー数を基準として抽出する手法を提案する。提案法は、時系列頻度データから特徴的な時間区間や場所の抽出を試みる方法 [Allan02, Kleinberg03] の一つに位置づけられ、最適観光ルート案内等の研究への応用が期待できる。

2 提案手法

提案アルゴリズムを以下に示す。

1. 対象のイベントの情報として場所を含む矩形領域を表す 2 地点の座標とデータの取得期間、イベントの開催期間を入力する。
2. 入力された場所の座標内でイベント開催期間に一度でもツイートしたユーザーを抽出し、抽出ユーザーの取得期間のツイートを収集する。
3. mean-shift クラスタリング [Comaniciu02] により、収集したツイート位置情報からスポットを抽出する。
4. Z-score により、有意に多くツイートされているスポット、及びその区間を抽出する。

以降、ステップ 4. の詳細について述べる。まず、データの収集期間を一定の間隔で分割する。例えば、1 日を

1 時間間隔で分割すれば 24 の区間ができる。これら区間を整数 $n(1 \leq n \leq N)$ で表す。一方、抽出したスポットを整数 $m(1 \leq m \leq M)$ で表す。

区間 n のスポット m のツイート数を $a(n, m)$ で表すと、また、区間 n のツイート数を $b(n) = \sum_{m=1}^M a(n, m)$ 、スポット m のツイート数を $c(m) = \sum_{n=1}^N a(n, m)$ 、全ツイート数を $L = \sum_{n=1}^N \sum_{m=1}^M a(n, m)$ となる。いま、周辺分布 $b(n)$ と $c(m)$ を固定すれば、区間 n のスポット m でのツイート確率は $P(n, m) = (b(n) \times c(m))/L^2$ となる。よって、実際のツイート数 $a(n, m)$ と、この確率に基づく期待値 $L \times P(n, m)$ から、区間 n のスポット m で有意に多くツイートされたかを以下の Z-score $z(n, m)$ で求める。

$$z(n, m) = \frac{a(n, m) - L \times P(n, m)}{\sqrt{L \times P(n, m)(1 - P(n, m))}} \quad (1)$$

上記では $a(n, m)$ をツイート数としたが、これをツイートしたユーザー数に置き換えられる。本稿では、前者をツイート数基準、後者をユーザー数基準と呼び、 $z(n, m)$ が有意に大きいものを抽出し分析する。

3 評価実験

本研究では、2015 年 8 月 14 日から 16 日の 3 日間開催されていたコミックマーケット (コミケ) を用いた。なおでは、コミケ開催中の 3 日間に、開催地である東京ビックサイト (緯度:35.63~35.63 経度:139.79~139.80 の区間) で位置情報付きツイートをしたユーザーを参加者とした。ただし、これらのユーザーの開催期間前後の行動を分析するため、前後 2 日間を加えた、合計 7 日間のツイートを集計した。2015 年 8 月 12 日から 8 月 18 日の 7 日間の位置情報付きツイート 51,806 件、該当ユーザー数 2,801 人を実験対象とする。

本データでは $N = 7 \times 24 = 168$, $M = 1,845$, ツイート基準では $L = 51,806$ 、ユーザー基準では $L = 2,801$ となる。

3.1 抽出スポットの可視化

提案法より抽出したスポットをプロットした結果を図 1 に示す。会場周辺での動きを確認するため、東京

User behavior analysis using tweets with the position information in events

†Yukiya NAKASATO ‡Eri KOBAYASHI †Kazumi SAITO

††Kazuhiro KAZAMA ‡‡Mitsuo YOSHIDA

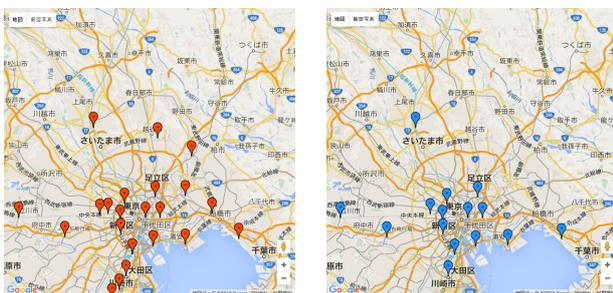
†University of Shizuoka

都心周辺のみを記載した。図1(a)にツイート数基準、図1(b)にユーザ数基準での結果である。

図1(a)より、ツイート数基準の場合、駅やその周辺の商店が多く抽出され、特に「秋葉原駅」や「新宿駅」で多くツイートされている。一方、図1(b)よりユーザ数基準では、上記に加え、線路沿いのスポットが多く抽出されたことから、イベント参加者の主な移動手段は電車や新幹線であると考えられるが、サービスエリアも抽出されたことから、車やバスで移動したユーザもいたことがわかった。また、ユーザ数基準での抽出より、ツイート数での抽出のほうが多く地図上に示されたことから、コミケ参加者には、単位時間あたりのツイート数が極端に多い人が多いということが考えられる。

抽出時間区間の多いスポットをみていくと、ツイート数・ユーザ数ともに「秋葉原駅」や「大崎駅」が抽出された。「秋葉原駅」では、関連店舗が多く、コミケ参加ユーザになじみ深い場所であるため買い物や打ち上げなどで訪れていることが考えられる。「大崎駅」ではコミケ開催期間中に「大崎コミックシェルター」を設置しており、それを目的としたユーザが訪れていると考えられる。

ツイート数基準のほうにのみ「岩手県平泉町」が抽出されたのは、あるユーザがツイートすると自動で指定した位置情報が付与される特殊なツイートをしているためである。また、「伏見稲荷大社」は「夜狐八重奏」の影響によるものと考えられる。「夜狐八重奏」は、iOS向けのTwitterクライアントであり、ツイートのジオタグに京都府の伏見稲荷大社の座標を自動で追加する機能によって多くのツイートが本来の位置情報と関係なく伏見稲荷大社の位置に観測される。[Morikuni15]



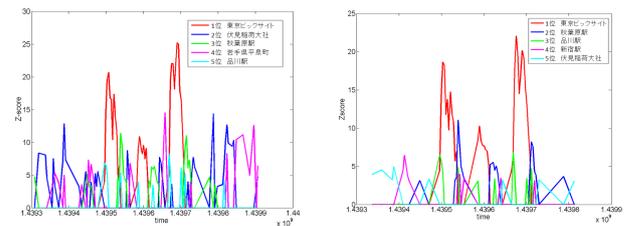
(a) ツイート数基準 (b) ユーザ数基準

図1: スポット抽出

3.2 Z-score の推移

提案法により抽出されたスポットのうち、出現時間区間の多い上位5件のZ-scoreの推移を表すグラフを

図2に示す。横軸は時間 (UNIXTIME 表記)、縦軸に Z-score を示し、ツイート数・ユーザ数基準での上位5スポットをグラフの凡例に示す。両結果とも、会場である東京ビックサイトでは12時台をピークに5時台から16時台に多くツイートされ、16時以降からは秋葉原でのツイート数が急激に増えている。よって、秋葉原を訪れるユーザとコミケに参加するユーザーには連動性があり、利用層が被っていることが考えられる。イベントの直前や直後にユーザ数が多い品川駅は、新幹線での移動などで利用するユーザが多いことが考えられる。



(a) ツイート数基準 (b) ユーザ数基準

図2: 上位5スポットのZ-scoreの推移

4 おわりに

本研究では、Twitterの位置情報付きツイートを用いて、ツイート回数とユーザ数を基準としてイベント参加者のよく訪れるスポットを抽出する方法を提案した。実験結果より、多少のばらつきはあるものの、参加者が多く集まるスポットや、多く集まる時間帯など、ユーザの行動を把握することができた。一方で、まだ収集データの精密さ不十分さや、曖昧さが目立った。特に、場所が大きく離れている場所は観光を考えるとという点で適切ではないように思える。今後は、より具体的な位置情報の設定や自動ツイートなどの投稿データを排除したより精密なデータ収集や、ほかのデータを交えたユーザ行動の検証を進める予定である。

謝辞 本研究は、総務省SCOPE(No.142306004)、及び、科学研究費補助金基盤研究(C)(No.26330345)の助成を受けた。

参考文献

[Allan02] J. Allan, "Topic detection and tracking: event-based information organization", Kluwer Academic Publishers, 2002.

[Kleinberg03] J.Kleinberg, "Bursty and hierarchical structure in streams", Data Mining and Knowledge Discovery, 7: 373-397, 2003.

[Comaniciu02] D.Comaniciu and P.Meer, "Mean Shift: A Robust Approach Toward Feature Space Analysis," IEEE Transactions on PAMI 24:603-619, 2002.

[Morikuni15] 森國泰平, 吉田光男, 岡部正幸, 梅村恭司: "ツイート投稿位置推定のためのフィルタリング手法", 情報処理学会論文誌, データベース Vol.8, No.4, pp.16-26, 2015.