

NGA2015におけるユーザの情報収集と回遊行動の分析

風間 一洋^{†,a} 谷 直樹^{†,b} 榭 剛史^{‡,c} 吉田 光男^{‡,d}

† 和歌山大学大学院システム工学研究科 ‡ 株式会社ホットリンク ‡ 豊橋技術科学大学情報・知能工学系

a) kazama@ingrid.org b) s161032@center.wakayama-u.ac.jp c) sakaki@ipr-ctr.t.u-tokyo.ac.jp

d) yoshida@cs.tut.ac.jp

概要 日本酒ゴーアラウンドは、日本酒固めが中心になって10月1日の「日本酒の日」に開催しているはしご酒イベントである。このイベントには多くの飲食店が参加するために、参加者は事前に公開される飲食店と酒蔵に関する情報を閲覧して巡回プランを作成しておき、当日はその時の状況に合わせてプランを修正しながら飲食店を巡回する必要がある。本稿では、NGA 2015に参加する飲食店と酒蔵に関する事前の情報収集とイベント当日の回遊行動を支援するために作成したシステムである NGA ナビについて紹介すると共に、実際に2015年9月24日からイベント当日の10月1日までの8日間運用して収集した利用履歴を情報閲覧と回遊行動の観点から分析した結果について報告する。さらに、作成したシステムと収集したデータの妥当性や問題点について考察する。

キーワード 日本酒ゴーアラウンド, 情報収集, 回遊行動, 位置情報, Web 利用マイニング

1 はじめに

「日本酒ゴーアラウンド」は、知名度が低かった10月1日の「日本酒の日」を広め、日本酒の楽しさを伝えるために、大阪の「日本酒固め」¹が2008年に開始したはしご酒イベントである。当初は大阪だけだったが、2015年は札幌・東京・京都・大阪・広島・福岡の6カ所で同時開催され、122軒が参加した。

日本酒ゴーアラウンドでは、参加飲食店は特定の酒蔵とペアを組む。参加者は、日本酒啓蒙のためのゆるキャラ「さかずきん」のバッジを参加証として事前購入し、当日はそのバッジをつけて周囲に「日本酒の日」であることをアピールしながら飲み歩くと、各飲食店では1杯の日本酒が無料でサービスされ、このイベントのために用意された日本酒や料理を味わうことができる。

このイベントが誕生した大阪では、参加飲食店数は48軒と多く、北区、福島区、都島区、中央区、天王寺区、浪速区と広範囲に渡って分布しており、さらに飲食店によって営業時間や営業形態、提供されるお酒や料理などが大きく異なるので、参加者は事前に公開される飲食店と酒蔵に関する情報に基づいて訪問したい店を選択して、それらの巡回プランを作成しておき、当日は予想滞在時間と実滞在時間の差や混雑度合いなど、その時の状況に合わせて巡回プランを修正しながら行動する必要がある。さらに、飲食店が未訪問の場合は、その飲食店の位置や最適な経路も調べなければならない。

そこで、大阪と札幌の二つの地域に対して、日本酒ゴーアラウンドに参加する飲食店と酒蔵に関する事前の情報収集と当日の回遊行動を支援する NGA ナビを作成

して、2015年9月24日からイベント当日の10月1日までの8日間運用して、その利用履歴を収集した。

本稿では、NGA ナビについて紹介すると共に、このシステム運用によって得られた利用履歴を、特に情報収集と回遊行動という二つの観点から分析した結果について報告する。さらに、これらの分析結果に基づいて、今回作成したシステムと収集した利用履歴データの妥当性や問題点について考察する。

2 関連研究

本研究の目標は事前の情報収集と当日の回遊行動の両方を支援するシステムの実現であるが、今年是最初の実績を作るためにシンプルなシステムを作成した。しかし、今後の研究の展開を考えると、以下のような技術要件が必要となる。

まず、仮想的な情報空間における情報収集と、距離や時間などの物理的制約を受ける実空間における回遊行動は同じではないが、互いに関連していることから、情報収集時のユーザの利用履歴を分析した結果を変換し、回遊行動の支援に利用することを目指す。その点において Web 利用マイニング技術 [1] を、実空間の物理的制約を含めて扱えるように拡張する必要がある。

次に、当日の回遊行動を把握するために GPS や Wi-Fi を使って測位できるスマートフォンの利用を前提とする。例えば、Hemminki らの研究 [2] ではスマートフォンを使って細かい粒度で位置情報を収集しているが、Web で利用するサービスとして実現した場合には基本的にアクセスした時点の荒い粒度の位置情報しか収集できない。そこで、はしご酒イベントでは回遊する飲食店の位置や特性が既知であることと、事前の情報収集履歴や関連情報を着目した回遊行動推定技術を確立する必要がある。

Copyright is held by the author(s).

The article has been published without reviewing.

¹<http://blog.livedoor.jp/manjigatame3/>

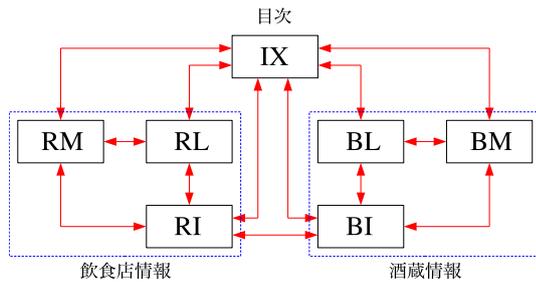


図 1: 閲覧モード

最後に、本システムでは飲食店の混雑状況を推定するために、現実世界の状況を知る手段を持たねばならないが、個々のユーザの利用履歴の分析だけでは実現は難しい。すでに各店舗の Web サイト, Twitter, Facebook と連携する目的で収集しているアカウント名や URL を利用して、例えば Sakaki らの Twitter をソーシャルセンサとして用いた研究 [3] のように、ソーシャルメディアの情報を分析して推定することを検討している。

3 NGA ナビと履歴

3.1 NGA ナビ

NGA ナビは、単一コードでパソコンからスマートフォンまでサポートできるように、本体を Python 2.7, ユーザインタフェースを HTML 5 と JavaScript, CSS, Google Maps JavaScript API で記述し, OS X 10.9.5 上で Apache 2.2.29 と mod_wsgi 4.4.11 を使って運用した。

ユーザ登録しなくてもクライアントを区別できるように、アクセス時には HTTP Cookie のクライアント識別子の有無を確認し、付与されていない場合に一意な識別子を生成して割り当てる。

ユーザのアクセス位置と誤差は、HTML5 の Geolocation API を使って取得した。精度向上のために enableHighAccuracy は true に、測位に GPS と Wi-Fi の両方が利用できない場合のレスポンス悪化を回避するために timeout は 10 秒にした。

図 1 に、NGA ナビの 7 種類の情報閲覧モードを示す。IX モードは目次で、RL, RM, RI モードはそれぞれ飲食店のリスト、地図、情報を、BL, BM, BI モードはそれぞれ酒蔵のリスト、地図、情報を表示する。飲食店情報と酒蔵情報は完全に分離しているわけではなく、他方の情報の一部も同時に表示され、また飲食店名・酒蔵名をクリックすれば他方に切り替わる。

IX モードから他のモードに移行するリンクは JavaScript を使用して、ユーザの位置情報を取得してから各種パラメータと一緒に HTTP POST で渡す。つまり、JavaScript を解釈しない一般的な Web ロボットは他のモードに移行できず、JavaScript を解釈する高機能な Web ロボッ

トは robots.txt でアクセスを禁止して影響を排除した。

飲食店情報に関する RL, RM, RI モードの画面を、図 2 に示す。RL, RM モードでは、開催地、時刻、座席、喫煙・禁煙、冷や・燗酒などの条件で絞り込むことができる。位置情報が有効ならば、RL モードでは飲食店を近い順にソートし、RM モードでは現在位置が青い点として表示される。さらに、RI モードでは近隣飲食店の訪問を促進するために、1km 以内の飲食店を最大 5 件表示した。目指す飲食店の経路検索は、Google Maps を公共交通機関を指定して呼び出して対処した。

3.2 履歴の記録

クライアントアクセス時には、クライアント識別子、情報閲覧モード、アクセス時刻、IP アドレス、ユーザエージェント、緯度・経度、位置の精度、飲食店・酒蔵の検索条件、飲食店・酒蔵の情報 ID を JSON 形式で記録した。なお、情報閲覧、公式ブログや、飲食店と酒蔵の Web ページ, Twitter, Facebook, 食べログなどのリンクを多数埋め込んでいるが、今回はそれらは監視対象にはしていない。

4 分析

4.1 システム利用状況の分析

本システムは 8 月 31 日に限定公開し、参加飲食店による試用やデータ登録・修正をある程度終えてから、2015 年 9 月 24 日 15 時に公式ページで告知を開始した。2015 年 9 月 24 日からイベント当日の 10 月 1 日までの 8 日間のアクセス数は 30,734 回、クライアント識別子数は 4,005 である。

図 3 に、8 日間のアクセス数の時間変動を示す。横軸が集計時間帯、縦軸がアクセス数である。この結果から、イベント前日の 9 月 30 日は昼から深夜に掛けてアクセス数が顕著に増加していることがわかる。イベント当日の 10 月 1 日の開催時間は 12:00 から 24:00 までだが、前日までと比較すると顕著にアクセス数が増大し、開始時刻の 12 時台に最大のピークがあり、次に終業時刻直前の 16 時台に第 2 のピークがある。なお、9 月 28 日の 7 時台と 12 時台のピークは「おはよう朝日です」でこのイベントについて放送された影響である。

次に、利用ユーザ数を分析する。本稿では、あるクライアント識別子で 2 回以上アクセスし、IX モード以外へのアクセスがあった場合に、HTTP Cookie や JavaScript を正しく処理できる Web ブラウザを人間が操作しているとみなして、ユーザと呼ぶ。ただし、同一ユーザがパソコンと iPhone を使った場合には、別のユーザとして扱われることに注意する必要がある。8 日間のユーザ数は 1,900 ユーザであり、図 4 にユーザ数の時間変動を示す。横軸が集計時間帯、縦軸がユーザ数である。図 3 と

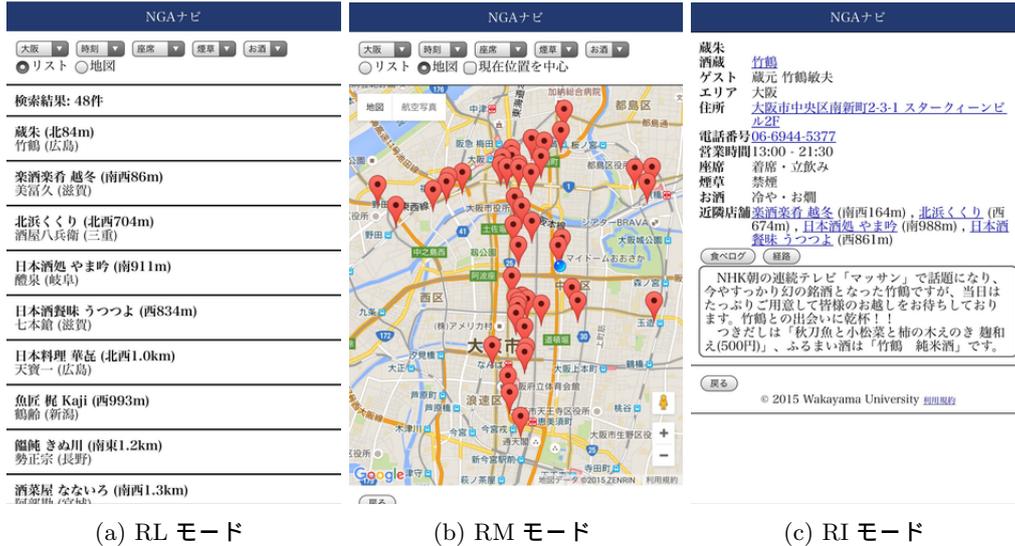


図 2: NGA ナビの飲食店情報画面

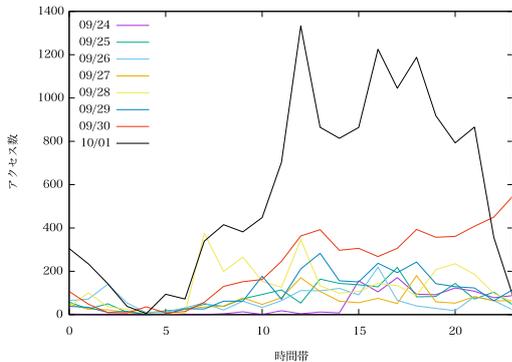


図 3: アクセス数の時間変動

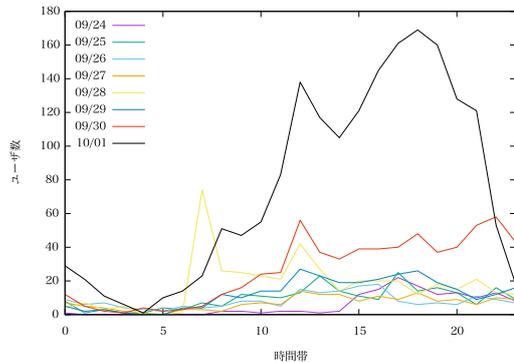


図 4: ユーザー数の時間変動

比較すると、開始直後の 12 時台では 1 人当たりのアクセス回数が多く、18 時台では参加しているユーザー数が多いと推測できる。

参考のために、札幌・大阪地区でイベント営業している飲食店数の時間変動を、図 5 に示す。大阪は昼から営業したが、札幌は夕方から営業したなどの地域による違いはあるが、12 時頃のピークを除けば、イベント当日のユーザー数の変動曲線と形状が類似している。

4.2 情報閲覧モードの分析

システム利用が顕著に多かったイベント前日(9月30日)と当日(10月1日)に絞って、ユーザーが図 2 のどのモードを利用したかを調べる。情報閲覧モードの利用頻度を表 1 に、1 時間ごとの利用頻度の積み重ね折れ線グラフを図 6 に示す。横軸が集計時間帯、縦軸が各情報閲覧モードのアクセス数の積み重ねである。この結果から、飲食店情報の閲覧(RI モード)が全体の半分近くを占めており、イベント当日は飲食店地図(RM モード)の利用率は明確に増加しているが、逆に回遊行動に直接

関係しない酒蔵情報の利用率は減少している。

4.3 利用デバイスの分析

情報収集や回遊行動時にユーザがどのようなデバイスを利用しているかを知るために、イベント前日と当日のデバイスの利用状況の積み重ね折れ線グラフを図 7 に示す。横軸が集計時間帯、縦軸が各機器のアクセス数の積み重ねである。なお、デバイスの種類は、ユーザーエージェント名に“iPhone”、“iPad”、“Android”、“Macintosh”、“Windows NT”の文字が含まれているかで判定した。この結果から、イベント前日と当日は利用デバイスの種類が変化していることがわかる。例えば、前日は Android が 20.4%、iPhone が 32.4% に対して Windows は 40% と多いが、イベント当日は Android と iPhone が逆転し、特に 18 時以降の 6 時間は Android が 27.0%、iPhone が 59.0% に対して Windows は 3.3% に大幅に減少している。つまり、情報収集段階でパソコンを使うが、回遊中はモバイルデバイスに切り替えているユーザが多いと思われる。さらに、本稿で提案したユーザに限定したことで

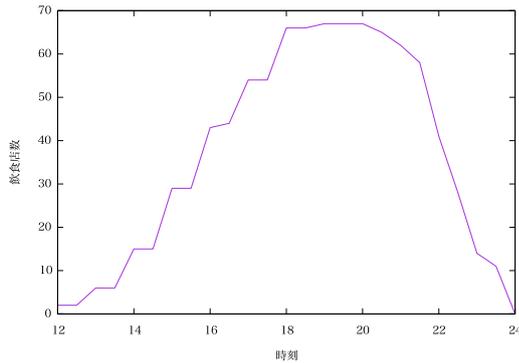
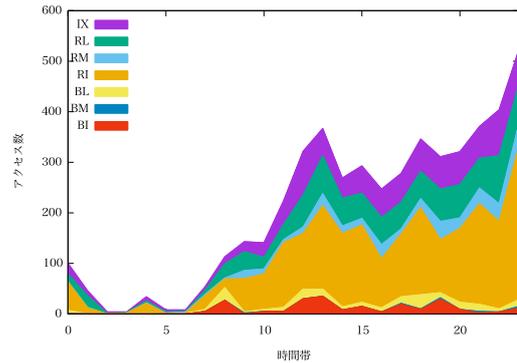


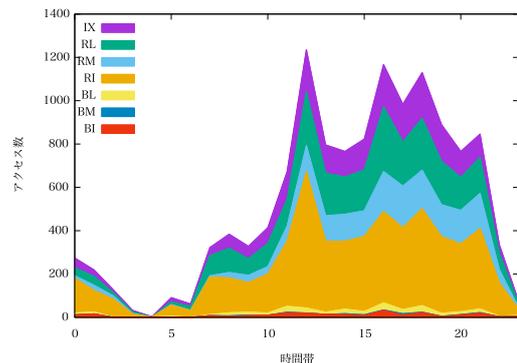
図 5: 飲食店数の時間変動

表 1: モードの利用頻度

モード	9/30	10/10
IX	931 (18.9%)	2099 (16.5%)
RL	939 (19.0%)	2805 (22.0%)
RM	332 (6.7%)	1790 (14.0%)
RI	2291 (46.5%)	5477 (43.0%)
BL	204 (4.1%)	295 (2.3%)
BM	16 (0.3%)	33 (0.3%)
BI	217 (4.4%)	242 (1.9%)



(a) 9月30日



(b) 10月1日

図 6: 情報閲覧モードの時間変動

Web ロボットの影響を排除できていることがわかる。

4.4 位置情報の分析

まず、イベント当日に Android または iPhone ユーザが、どの程度位置情報を付与していたのかを調べた。その結果、iPhone の場合は 449 ユーザ中 187 ユーザ (41.7%)、Android の場合は 307 ユーザ中 59 ユーザ (19.2%)、合わせると 756 ユーザ中 246 ユーザ (32.54%) であった。位置情報を付与したユーザの割合が少ないが、位置情報を有効にする利点がユーザに正しく理解されていなかったこと、HTML5 の Geolocation API を呼び出した場合に 2 回表示される位置取得許可のダイアログで拒否しているユーザが多かったことが考えられる。

次に、位置情報を付与した Android または iPhone ユーザのアクセス数の分布を、両対数グラフにプロットして図 8 に示す。横軸はアクセス数 (f)、縦軸はそのアクセス数以上である累積分布関数 $P(\geq f)$ の値である。例えば、5 回以上アクセスしたユーザは全体の 57.5%、10 回以上アクセスしたユーザは全体の 29.8% であった。

さらに、ユーザと情報閲覧した飲食店またはユーザに最も近い飲食店の間の距離の変化を分析する。図 9 に、Android または iPhone ユーザのアクセス時刻と距離の分布を示す。横軸はアクセス時刻、縦軸は距離である。図 9a から、イベント開催直後は、右斜め上に向かう直線状の分布で表されるように近い順に表示される検索結

果を順番に閲覧しているユーザが見られるが、18 時以降は近い少数の飲食店に絞って閲覧している傾向がわかる。図 9b から、8 時頃から 18 時頃までは、水平の直線状の分布で示される同一場所からの集中的なアクセスが見られるが、18 時以降は飲食店の近傍からアクセスしている傾向がある。なお、検索した飲食店から 100m 以内のアクセスは 12 時以降から始まり、18 時以降になると飲食店との距離が全体的に短くなる傾向がある。

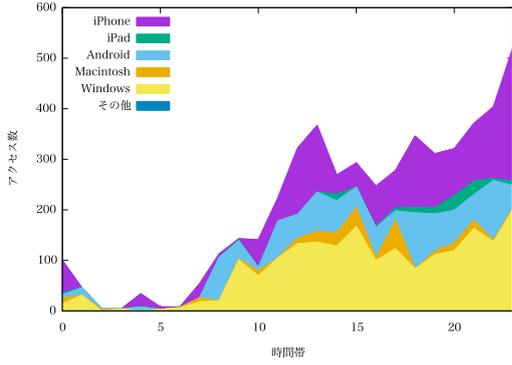
実際にどこからアクセスしているかを調べるために、誤差が 100m 以内の Android または iPhone ユーザからの大阪周辺のイベント当日のアクセス位置を、Yahoo! Open Local Platform² を使ってヒートマップとして可視化した結果を図 10 に示す。なお、赤い丸印は飲食店の位置である。この結果を見ると、大阪中心部では飲食店内またはその周辺におけるアクセスが多いが、中心部から離れると飲食店周辺からのアクセスがほとんどない地域もあるなど、偏りがあることがわかる。

4.5 情報閲覧の分析

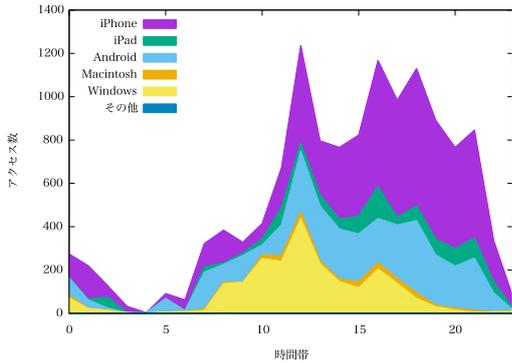
次に、イベント前日と当日の各飲食店の情報を閲覧したユーザ数を調べた。

飲食店ごとの閲覧ユーザ数を、図 11 に示す。横軸は各飲食店をイベント当日の閲覧数で順位付けした順位で

²<http://developer.yahoo.co.jp/webapi/map/>

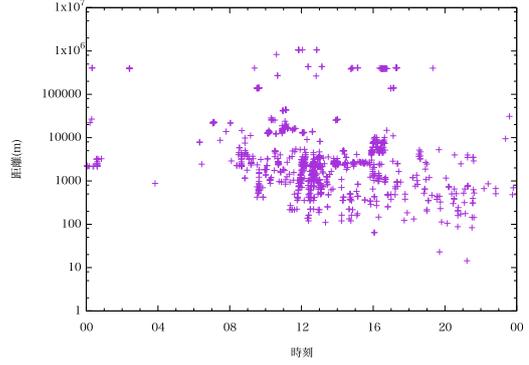


(a) 9月30日

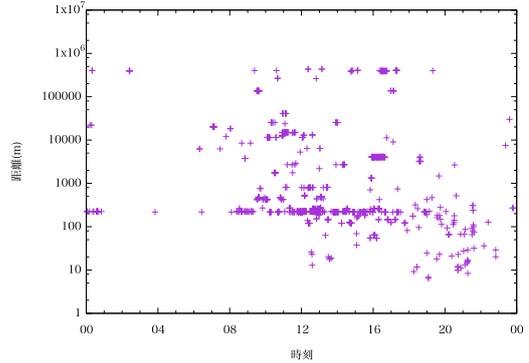


(b) 10月1日

図7: デバイス利用状況の時間変動



(a) ユーザと閲覧飲食店間



(b) ユーザと最近傍飲食店間

図9: 距離の時間変動

あり、縦軸は1日の閲覧数である。なお、48位までは大阪、49位以下は札幌の飲食店と明確に分離していた。イベント前日と当日の閲覧ユーザ数の相関係数は0.728であり、中程度の相関があった。

そこで、分析対象をユーザが活発な大阪に限定して、回遊行動を反映しているイベント当日の閲覧数 y を目的変数、事前の情報収集時の人気度を反映しているイベント前日の閲覧数 x_1 、1km以内に存在する飲食店数 x_2 、営業時間 x_4 、営業時間中の営業飲食店数の平均値 x_4 を説明変数として、Rを使ってステップワイズ法で重回帰分析した。この結果、表2に示す結果が得られ、

表2: 重回帰分析の結果

	回帰係数	標準誤差	t 値	p 値
定数項	12.2006	5.4827	2.225	0.03118
x_1	1.1820	0.1701	6.949	1.21×10^{-8}
x_2	1.6422	0.3226	5.090	6.82×10^{-6}

決定係数 R^2 は0.7015、自由度調整済みの決定係数 R^{*2} は0.6882と高い精度が得られた。また、 x_1, x_2 共に p 値は1%以下であり、説明変数として妥当であった。

つまり、当日の回遊行動は、事前の人気度に加えて、近傍にすぐ行くことができるような他の飲食店が存在するかどうか大きく影響し、これが図11におけるイベント前日と当日の閲覧順位の差に繋がるとされる。また、営業時間を長時間化したり、他店がまだ営業していない時間帯から営業することは、当日の回遊行動にあまり影響を与えていないことがわかった。

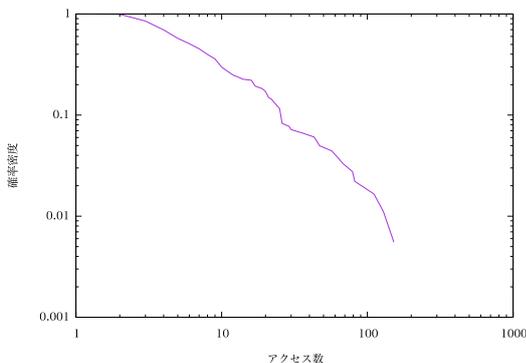


図8: 位置情報付与ユーザのアクセス数の分布

5 考察

システムへのアクセス履歴は事前の情報閲覧と当日の回遊行動を反映していると仮定して、イベント前日と当日に分けて分析した結果、情報閲覧時と回遊行動時ではアクセスパターンや、利用デバイス、位置関係に固有の特徴が現れることがわかった。ただし、平日開催という

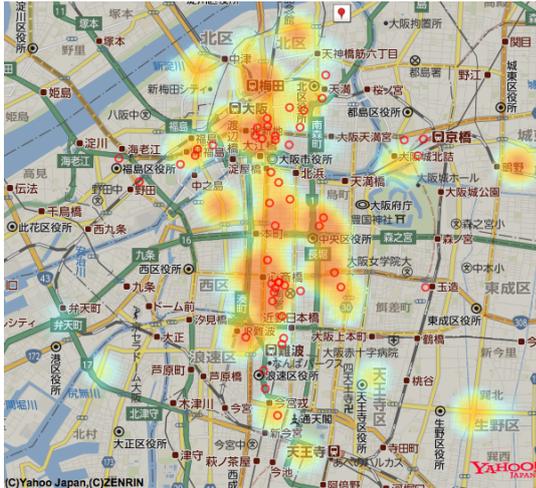


図 10: ヒートマップ (10月1日)

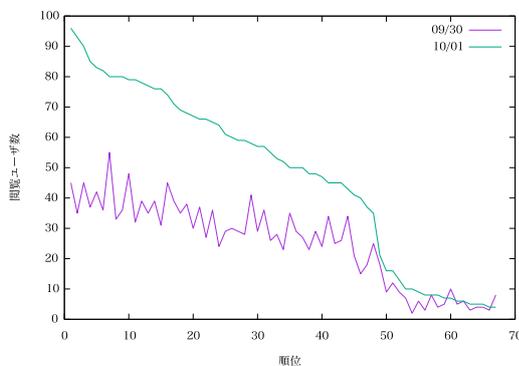


図 11: 飲食店情報の閲覧ユーザ数

こともあり、夕方勤務終了後に回遊しているユーザも多く、詳細な分析のためには情報閲覧と回遊行動の分離が必要である。

位置情報に関しては、イベント中はGPSを備えたスマートフォンの使用率が高くなるが、位置取得を拒否しているユーザの割合が大きかった。また、回遊行動時にはある飲食店に滞在中に次の飲食店情報を閲覧すると仮定したが、実際には店外からと思われるアクセスが多く、当初計画していた離散的な位置情報からの回遊行動の推定は困難であることがわかった。そこで、位置情報付与率を増加させるようなシステム改良と、店舗訪問の有無をユーザ本人に入力してもらうチェックイン機能との組み合わせを検討している。

また、情報閲覧と回遊行動の関係を分析した結果、当日の回遊行動は事前の人気度に加えて、飲食店の地理的關係も影響していた。この結果は、事前の情報閲覧の履歴を使って当日の回遊行動を支援できることを示唆すると共に、ユーザの嗜好に合っているが地理的に離れている飲食店への適切な誘導が必要になる。

飲食店やユーザから混雑状況推定の要望があったが、

今回はシステム利用が飲食店数の開店よりもかなり早いペースで増加していることが混雑に結びついていることは把握できた程度であり、より詳しく実世界の状況を把握するためには、ソーシャルメディア分析などの他の手法との組み合わせが必要である。

6 おわりに

本稿ではNGA 2015のために運用したNGAナビの利用履歴について分析した結果、事前の情報収集と当日の回遊行動を示す特徴を確認できることを示すと共に、位置情報に関して位置情報付与率の増加と店舗訪問の確実な把握などの課題があることを明らかにした。

イベントの開催日である10月1日からの経過時間が少ないこともあり、本稿では作成したシステムと収集したデータの妥当性を立証するための分析と、今年度のシステムが抱える問題点の洗い出しに重点を置いたが、今後は、データのさらに詳細な分析を進める予定である。

来年の10月1日は土曜日であることから今年度以上の参加者が見込まれると同時に、昼から回遊行動を開始できることで長時間の回遊履歴を収集できると思われる。今年は着実な実績を作るために、安定性を重視して機能を絞り込んだシステムを設計・開発したが、研究用のデータ収集のための条件が整った来年に向けて、目標とする事前の情報収集と当日の回遊行動の両方を支援するシステムの実現に必要なとされる技術要件を盛り込んだ次期システムの検討をおこなう予定である。

謝辞

NGAナビの実現にあたって、大阪エリアの情報を提供して頂いた日本酒卍固め、札幌エリアの情報を提供して頂いたkome Ezo酒事務局、そして各エリアの情報の取りまとめと仲介をして頂いた蔵朱・大西正哉氏と酒井啓氏に感謝する。

参考文献

- [1] Jaideep Srivastava, Robert Cooley, Mukund Deshpande, and Pang-Ning Ta. Web usage mining: Discovery and applications of usage patterns from web data. *ACM SIGKDD Explorations Newsletter*, Vol. 1, No. 2, pp. 12–23, 2000.
- [2] Samuli Hemminki, Petteri Nurmi, and Sasu Tarkoma. Accelerometer-based transportation mode detection on smartphones. In *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems (SenSys '13)*, pp. 13:1–13:14. ACM, 2013.
- [3] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. Earthquake shakes twitter users: Real-time event detection by social sensors. In *Proceedings of the 19th International Conference on World Wide Web*, pp. 851–860, 2010.