

ユーザの休日における外出先 POI 情報に基づく プライバシーを考慮した居住地属性推定手法の一検討

湯木悠太¹ 米澤拓郎¹ 廣井慧¹ 河口信夫^{1,2}

キーワード : ユーザプロファイリング、ジオロケーション、モビリティ、POI

発表概要

近年スマートフォンの普及に伴って、アプリケーションやサービスの利用を通じ、個人の行動・活動履歴の収集や嗜好などの情報が抽出可能となってきている。特に、個人の Web サービスを利用する際の回遊履歴や、スマートフォンが接続している携帯基地局データおよび GPS 等のデータを参照することで、個人の特徴は容易に抽出しうる状況となっている。これら個々ユーザのプロファイルは、サービス提供者や開発者にとっては非常に有益な情報となり、またユーザ自身にとっても自分の嗜好にあったサービスを受けられる可能性が高まるため、利便性の向上につながると考えられる。一方、このような個人のデータの利活用を行う際には、プライバシーの保護が常に懸念される。特に、近年のプライバシーデータ保護に関しては、欧州の GDPR[3] 制定に代表されるようにその重要性が高まってきている。厳密な制度化が今後より進むにつれ、プライバシーデータを収集・保有する企業自身にとっても、データを扱うこと自体がリスクへとつながる懸念もなされている。

プライバシーデータの中で、特に慎重な取り扱いを要するデータとして、ユーザの位置情報が挙げられる。その中でも、ユーザの居住地や職場の位置などは最も秘匿すべき情報であると考えられる。一方で、居住地などの情報はレコメンドサービスなどを構築する際には非常に有益な情報となるため、今後は実際のユーザの居住地の位置情報を扱うことなく、いかに同等レベルのユーザのプロファイリングが達成されることが求められる。居住地は有用な情報なため、これまで SNS などを用いた推定[6]なども試みられている。

本研究では、上記の問題を解決するため、ユーザの居住地情報を直接用いることなく、居住地に関する属性を抽出可能な手法の構築を目指す。特に本稿では、休日の外出先 POI (Point of Interest) の遷移を分析することで、居住地の属性が推定可能であるか、の検討を行う。POI とは地図上の特定のポイントを指し、一般的には、目標物や施設を指す場合が多い[1]が、本研究においては、その施設などの一般名詞や属性を POI として定義し、実際の位置情報につながるような直接的な施設名は扱わない。すなわち、本研究

における POI を用いることで GPS のデータを抽象化し、個人に紐づかないような利活用を行えることが期待される。

本稿ではユーザの居住地 (生活圏) の属性が、休日におけるユーザの外出先 POI に影響を及ぼすという仮説を立て、休日の滞在 POI 遷移からユーザの居住地推定が可能であることを示すことを目的とした。居住地属性の中でも、本稿では特に居住地と周辺の駅の関係性が、休日行動に影響を与えるのではないかと仮説について検討を行った。

・分析対象データと分析内容

株式会社ログウォッチャーから購入した 6 月 1 日から 6 月 22 日までの愛知県岡崎市を通過したことがあるユーザのデータを用いて本仮説の検証を行った。本データはユーザが特定のアプリを利用している際に、利用規約同意の元で記録されたデータであり、各レコードには各ユーザの位置情報、時間などが含まれる。データのレコード総数は 7,478,613 件でユニークユーザ数は 29,866 人であった。本データは愛知県岡崎市を通過したユーザのデータを収集しているため出張等で新幹線を用いて通過したユーザもデータが含まれている。そのため愛知県近辺の北緯 34.5 度から 35.5 度まで、西経 136.5 度から 138 度までの範囲内で行動していたユーザに絞って仮説の検証を行った。また 1 日あたりのデータ数が少ないユーザについては、後述の滞在地の推定が適切に行えない可能性が高いためデータを除去して検討を行った。以上のデータクレンジングを実施した結果、データ数は 4,112,027 件、ユーザ数は 18,892 人となった。上述したように、データは高頻度で収集される状況と低頻度で収集される状況が混在している。滞在地推定手法として遠山ら[4]や遠藤ら[5]の提案する手法も存在するが高頻度で収集された GPS データを対象としているため本データでの推定には適していない。そのため、滞在地の推定は岩田ら[2]の疎な GPS を対象にした手法を利用した。しかし、岩田らの手法も 5 分間隔での測位データを対象としていたため岩田らの手法に加え以下の処理を実施した。

(1) データの圧縮

アプリ起動時の非常に高頻度で取得されたデータについては処理を簡便化するために 5 分間隔でまとめ、緯度経度について平均を用いた。統合する際には GPS の測位誤差

¹ 名古屋大学大学院 工学研究科
² 名古屋大学 未来社会創造機構

の大きい外れ値を含めてしまうと平均値が本来の値から大きくずれてしまう可能性を考慮し、5分間のデータの中で直前の測位点に対し、距離が1000m以上離れているものは外れ値とみなし、外れ値を含まない測位点の平均を用いた。

(2) スパースな測位時刻の分割

スパースな測位点に対して滞在地推定を連続で実施しないようにするため2時間以上時刻差のあるデータは分割し滞在地推定を実施した。

実際に仮説の検証をするにあたり生活圏周辺の駅の情報として生活圏内の周辺駅数、最寄り駅までの距離、生活圏内の駅と距離を考慮したスコアを用いた。スコアとして生活圏内周辺の駅までの距離の逆数の和を用いて計算を行った。また休日の外出先 POI 遷移の分析対象として、モビリティの活発さを検討した。モビリティの活発さの指標としては、ユーザの滞在地の総滞在時間、滞在地数、1日あたりの総移動距離とした。上記の生活圏情報と、休日の外出 POI 遷移との関係性を分析することにより、仮説の検証を行った。

・分析結果

図 1,2,3 に上記で述べた生活圏情報とユーザのモビリティの活発さに対するグラフを示す。また、表 1 に各周辺駅情報とモビリティの活発さの指標との相関係数を示す。

表 1 から、全ての周辺駅の要素に対して総移動距離は相関が見られなかった。周辺駅数と距離を考慮した周辺駅スコアでは総滞在時間に対し、弱い相関が得られ、最寄り駅までの距離からは総滞在地数に対して弱い相関が得られた。本研究では周辺 POI の中でも駅に着目しモビリティに対して検討を行ったが、駅以外の周辺 POI 情報と組み合わせることにより、得られた結果よりも強い相関が得られることが期待される。また、距離を考慮した周辺駅のスコアに関してはナイーブな計算による導出を実施した点やモビリティの指標の追加等、駅のための POI に対しての改善の余地が挙げられる。

今後の課題としては上記に述べたように他の POI 情報の抽出をどの方向性から進めていくかの検討や実証実験を行えるような環境整備を行っていくことなどが挙げられる。

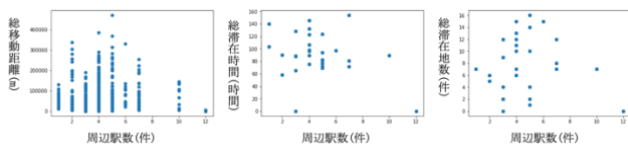


図 1 周辺駅数とモビリティの相関

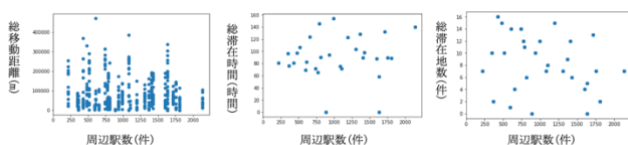


図 2 最寄り駅距離とモビリティの相関

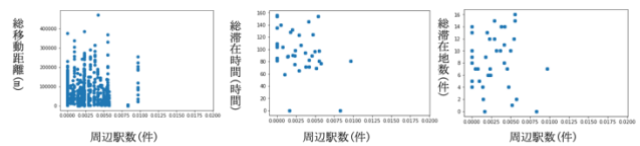


図 3 距離を考慮した周辺駅スコアとモビリティの相関
 表 1 各生活圏内周辺駅情報とモビリティとの相関係数

	総移動距離	総滞在時間	総滞在地数
周辺駅数	-0.0693	-0.306	-0.065
最寄り駅距離	-0.041	0.104	-0.237
周辺駅スコア	-0.145	-0.315	-0.008

謝辞

本研究の一部は、JST CREST JPMJCR1882 の支援の元で行われた。

参考文献

[1] “POI (Point of Interest) / エリアマーケティング・商圏分析・GIS用語集” .https://www.mapmarketing.co.jp/glossary/poi.html

[2] 岩田 紗瑛, 新田 知之, 高山 敏典, 柳澤 政生, 戸川 望, “疎な GPS 測位情報を対象にした測位精度と短時間滞在除去に基づく滞在地推定手法,” マルチメディア, 分散協調とモバイルシンポジウム 2017 論文集, vol. 2017, pp. 523–531, 2017.

[3] Data Protection in the EU: https://ec.europa.eu/info/law/law-topic/data-protection_en

[4] 遠藤 結城, 戸田 浩之, 小池 義昌, “移動手段推定のための表現学習を用いた GPS ログからの特徴抽出,” 情報処理学会論文誌データベース, vol.8, pp12-23, 2015

[5] 遠山 緑生, 服部 隆志, 萩野 達也, “携帯電話の測位機能を用いた有意位置の学習”, 情報処理学会論文誌, vol.46, pp2915-2924, 2005

[6] Lars Backstrom, Eric Sun, and Cameron Marlow. 2010. Find me if you can: improving geographical prediction with social and spatial proximity. In Proceedings of the 19th international conference on World wide web (WWW '10). ACM, New York, NY, USA, 61-70.