

# GPS データを用いた商業地域訪問者の特性抽出に関する研究

A Study on Extracting Visitors' Characteristic at Commercial Area from GPS Data

学籍番号 47116748  
氏名 羽田野 真由美  
(Hadano, Mayumi)  
指導教員 柴崎 亮介教授

## 1. はじめに

近年、大型商業施設の開業や顧客の嗜好の変化、後継者不足などにより、商店街を始めとする小規模な商業地域の衰退が問題になっている。

一般に、大型チェーン店や大型商業施設では、店舗の魅力向上のために、ID付きPOSデータなどから訪問者分析を行なってマーケティング施策を講じていることが多い。しかしながら、小規模な商業地域でID付きPOSデータを活用することはその導入・運用コストから難しいといえる。また、自治体などが実施する商圈アンケートも調査頻度が低く網羅性が低いため、小規模な商業地域が利用可能でかつ広域を対象としたデータセットがないというのが現状である。

そこで、本研究は南関東エリアの商業地域を規模に関わらず対象にして、GPSデータから商業地域訪問者の特性を抽出し、データセットとして整備することを目的とする。

GPSデータから買い物などの行動ラベルを付与する研究は過去に少数行なわれているが(Liao *et al.*, 2007)、アルゴリズムが複雑で大規模・長期のGPSデータに応用することが難しい。そこで本研究では滞在位置と滞在時間の2つの条件だけを利用した手

法で商業地域訪問を抽出した。

## 2. 商業地域訪問の抽出手法および検証

大規模GPSデータでの計算の前にまず、滞在目的が既知のGPSデータを用いて商業地域訪問の抽出手法の信頼性を検証する。以下にその手法および結果を述べる。

### 2.1 使用データ

商業地域の位置情報は2011年の商業集積統計(秋山ら, 2011)から得た。これは一定数以上の商店が集まって商店街を形成している地域をポリゴンデータとして整備したもので、商店街のような小規模な商業地域も含まれている。本研究では、2011年の南関東エリア(東京都, 埼玉県, 千葉県, 神奈川県)のデータ(8,774の商業地域)を用いた。

GPSデータは、金杉ら(2012)によって収集されたGPS履歴データを使用した。また検証用に同被験者の滞在目的付き行動日記データを利用した。本研究では123人の25日間にわたるデータを抜き出して使用した。

### 2.2 商業地域訪問の抽出手法

はじめにGPSデータから、一定時間以上、一定範囲の中に留まるという条件を指定して滞在点を抽出した。本研究では5分以上

のあいだ、半径 200 m 以内に留まることを滞在点抽出の条件とした。

次に、空間的に近接した滞在点同士をひとつにまとめるために、クラスタリングを行ない、滞在クラスタを抽出した。クラスタリングの際のアルゴリズムは DBSCAN を利用した。DBSCAN の距離のしきい値  $Eps$  と最小点数  $MinPts$  の 2 つのパラメータには、それぞれ 200 m と 2 を設定した。

最後に各クラスタをそれぞれ分析して、自宅・勤務地・商業地域訪問のラベルをつけた。各クラスタの滞在点が深夜の時間帯に一番多く属するクラスタを自宅とし、平日日中の時間帯に一番多く属するクラスタを勤務地とした。商業地域訪問のラベルは、I. 滞在点に付与、II. 滞在クラスタに付与という 2 通りの計算を行なうことで、手法の違いによる精度の差を調べた。両手法とも代表点が商業地域ポリゴンに含まれ、滞在時間が 10 分以上 4 時間未満という条件で商業地域訪問ラベルをつけた。滞在時間の下限しきい値は、鉄道駅での待ち時間・乗り換え時間を除去するために、上限しきい値は買い物行動を取っているとは考えづらい長時間の滞在 (例えば営業活動など勤務地以外の従業) を取り除くために設定した。

### 2.3 商業地域訪問の抽出結果

対象被験者の平均滞在点数と滞在クラスタの平均個数はそれぞれ 106.8 と 7.4 であり、平均滞在時間は 280 分であった。

商業地域訪問のラベルづけをした結果の例を図 1 に示す。各被験者の商業地域訪ラベル数の平均は、手法 I で 1.86、手法 II で 7.87 であった。手法 I は、複数の滞在点をひとつにまとめたクラスタにラベルを付与

しているため、手法 II よりも抽出数が少なくなっている。



図 1 ラベルづけの結果例

(大きい丸はクラスタ, 小さい丸は滞在点を表す)

### 2.4 滞在目的付き行動日記との比較

実際に買い物が行なわれた滞在範囲が、商業地域ポリゴンと重なる割合は、平均で 63.67% であった。

本手法で抽出した商業地域への訪問と、滞在目的付き行動日記データとを比較した結果、一致率は手法 I で 14.09%、手法 II で 41.41% であった。さらに、商業地域ポリゴン外で買い物が行なわれたことによる影響を除去して一致率を再計算すると、手法 I で 19.69%、手法 II で 60.56% であった。全体的に手法 I よりも手法 II の方が一致率が高く、買い物行動を精度よく抽出していることがわかる。これは、データ取得日数が 28 日と短く、クラスタリング時にノイズとして除去される滞在点が多いことや、自宅・勤務地クラスタの中に買い物行動が含まれることが原因と考えられる。手法 I の主な誤差要因は滞在時間の下限しきい値を 10 分と設定したことによって、短時間の買い

物行動を抽出できなかつたためと考えられる。なお、図 2 は手法 II における各被験者の一致率 (実線) と、商業地域内で買い物が行なわれた割合 (破線) を示している。

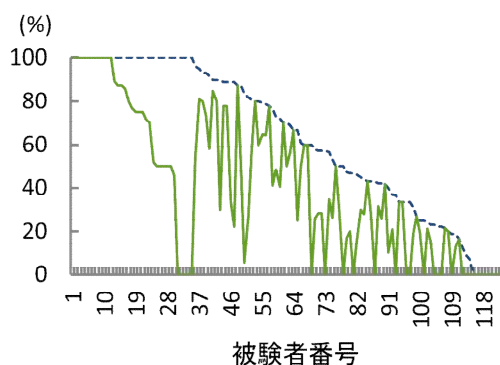


図 2 手法 II による各被験者の一致率 (実線)  
(破線: 商業地域内で買い物が行なわれた割合)

### 3. 商業地域訪問者のデータセットの作成

前章で検証された手法を、大規模・長期の GPS データに応用し、商業地域訪問者に関する基礎データセットを作成した。

#### 3.1 大規模 GPS データ

ゼンリンデータコム社の混雑統計 (R) データを用いた。これは承諾を得て取得した携帯電話の GPS データをもとにして、個人情報秘匿処理を行った統計データである。

本研究では対象期間を 2010 年 8 月 1 日からの 1 年間、対象場所を南関東エリア (埼玉県, 千葉県, 東京都, 神奈川県) として約 22 万人分 (221,269 人) のデータを利用した。

#### 3.2 商業地域訪問の抽出手法

前章で説明した手法を大規模 GPS データに応用した。前章で高い精度が得られたのは手法 II であったが、ここでの対象人数が約 22 万人と膨大であることから、計算時間

を考慮して手法 I を用いた。なお、大規模 GPS データの測位間隔が不規則かつ最低 5 分であることを考慮して、滞在の条件を 20 分以上、半径 150 m に変更した。

#### 3.3 実数への拡大処理

大規模 GPS データはサンプルデータなので、実数として利用する際は拡大処理を行う必要がある。本研究では、国勢調査の人口に対する大規模 GPS データの夜間人口の抽出率の逆数を拡大係数として設定した。なお、拡大係数を設定する集計単位は 3 次メッシュを利用した。

#### 3.4 年間出費額の推定

商業地域での滞在時間が長ければ長いほど、そこでの消費額も大きいと仮定して、年間出費額を推定した。具体的には、2011 年の家計調査から市区町村ごとの 1 人あたり出費額 (食費, 衣服費) を計算し、その値を年間総滞在時間で比例配分することによって推定した。

#### 3.5 結果

滞在クラスタ数は全体で 3,528,790 であり、そのうち商業地域訪問ラベルが付与されたクラスタ数は 536,982 (15.21%) であった。さらに、対象者 221,269 人のうち商業地域訪問ラベルが付与された人数は 164,102 人 (74.16%) であった。

結果を商業地域ごとに集計して年間ユニーク訪問者数を算出した結果、渋谷や新宿、池袋などの主要都市で大きい値を示した (図 4.6)。一番訪問者数が多かったのは渋谷駅井の頭通り沿いの商業地域で、およそ 370 万人であった。訪問者が多い地域は、鉄道

駅周辺に位置し、地域内店舗数が多いという特徴があった。

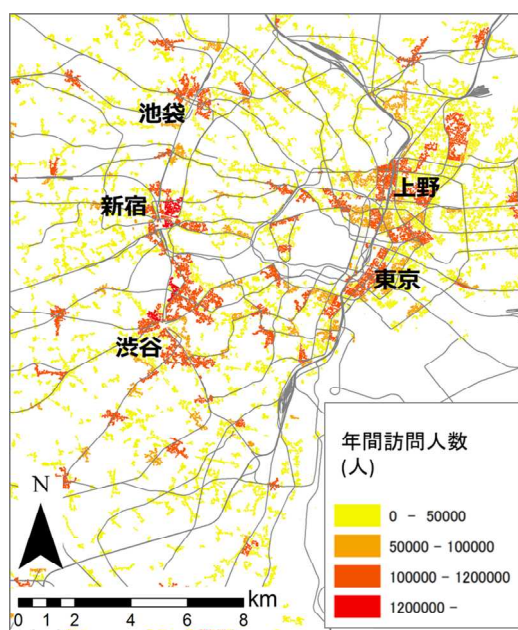


図3 首都圏における年間ユニーク訪問人数

本手法で計算された推定年間出費額を2004年の商業統計における年間販売額と比較して、手法の信頼性を検証した。その結果、食費の推定年間出費額と商業統計での年間販売額間の決定係数は0.48程度であり一定の相関がみられた(図4)。しかしながら、衣服費の推定年間出費額と実際の販売額の決定係数は0.22程度で過大評価の傾向が強かった。

今回の出費額の推定法は家計調査の値を総滞在時間帯で比例配分するという簡便なもので、訪問先の商業地域の構成業種を全く考慮しなかった。そのため、衣料品店がない地域でも衣服費を配分してしまい、過大評価の地域が多くみられるという結果になったと考えられる。しかしながら、食料品は比較的どの商業地域にも含まれ、ま

た最寄り品の傾向が強い品物であることから、今回のような方法でもある程度の精度の結果が得られたと考えられる。

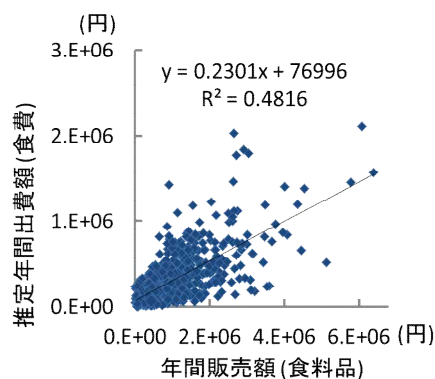


図4 商業統計の年間販売額と推定値の散布図

#### 4. おわりに

本研究では、GPSデータから商業地域訪問を抽出する手法を提案し、その手法を元にして商業地域訪問者の基礎データセットを整備した。

本研究によって今まで存在しなかった小規模な商業地域を含む広域エリアを対象に商業地域訪問者の基礎データセットを整備することができた。この成果は商業地域におけるマーケティングや戦略策定での活用が期待される。

#### 参考文献

- Liao, L., Fox, D., Kautz, H., 2007. Extracting Places and Activities from GPS Traces Using Hierarchical Conditional Random Fields, *The International Journal of Robotics and Research*, 26, 119-134.
- 秋山祐樹・仙石裕明・田村賢哉・柴崎亮介 (2011)：日本全土の商業統計ポリゴンデータの開発と商業集積地域ポリゴンデータの信頼性検証，第20回地理情報システム学会講演論文集 (CD-ROM, F-2-3)。
- 金杉洋・黒川茂莉・村松茂樹・関本義秀 (2012)：携帯電話基地局通信情報の行動分析への適用可能性把握，第32回交通工学研究発表会，309-315。