

2012 年度 修 士 論 文

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

羽田野 真由美
Hadano, Mayumi

東京大学大学院新領域創成科学研究科
環境学研究系 社会文化環境学専攻

目次

第 1 章	はじめに	1
1.1	研究の背景・目的	1
1.1.1	小規模商業地域の現状	1
1.1.2	商業地域訪問者に関する既存データセット	2
1.1.3	本研究の目的・独自性	2
1.2	本論文の構成	3
第 2 章	既往の研究事例	4
2.1	商業地域訪問者の分析事例	4
2.1.1	アンケート調査	4
2.1.2	ID つき POS データ	5
2.1.3	モバイル空間統計	5
2.1.4	GPS データ	6
2.2	GPS データを用いたひとの行動特性抽出事例	7
2.2.1	ひとの行動抽出に関する事例	7
2.2.2	行動ラベルの付与に関する事例	7
2.3	まとめ	8
第 3 章	GPS データを用いた商業地域訪問の抽出手法	9
3.1	使用データ	9
3.1.1	商業集積統計	9
3.1.2	GPS データおよび行動日記データ	10
3.2	商業地域訪問の抽出手法	11
3.2.1	滞在点の抽出	11
3.2.2	滞在クラスターの抽出	11
3.2.3	自宅・勤務地・商業地域のラベル付け	12
3.2.4	比較検証方法	14
3.3	結果	15

3.3.1 滞在地の抽出結果	15
3.3.2 滞在クラスターの抽出結果	15
3.3.3 自宅・勤務地・商業地域訪問のラベル付けの結果	16
3.3.4 比較検証の結果	17
3.4 まとめ	20
第4章 商業地域訪問者に関する基礎データセットの作成	21
4.1 使用データ	21
4.1.1 GPS データ	21
4.1.2 収入・支出に関する統計データ	21
4.2 商業地域訪問者の基礎データセット作成手法	23
4.2.1 商業地域訪問の抽出	23
4.2.2 データセット整備のための諸項目の計算	24
4.2.3 実数への拡大計算	25
4.3 結果	28
4.3.1 全体の結果	28
4.3.2 商業地域ごとの集計結果	29
4.3.3 商業統計との比較	37
4.4 まとめ	40
第5章 結論	41
5.1 本研究の成果	41
5.2 本研究の課題	42
5.2.1 商業地域訪問の抽出手法における課題	42
5.2.2 基礎データセット作成手法における課題	43
5.3 本研究の展望	44
5.3.1 大規模 GPS データの属性の偏りを考慮に入れた拡大処理	44
5.3.2 他地域への応用	44
5.3.3 買い物行動モデルの作成	44

図目次

図 1.1	本論文の構成	3
図 2.1	第 2 章の構成の概念図	4
図 3.1	商業集積統計による商業地域の空間的分布と商業地域内店舗数	9
図 3.2	GPS データから商業地域への訪問を抽出するフローチャート	11
図 3.3	DBSCAN のアルゴリズム概念図	12
図 3.4	商業地域訪問ラベルの付与条件概念図	13
図 3.5	正誤判定条件の概念図	14
図 3.6	GPS 履歴・滞在地・滞在クラスタの分布例	15
図 3.7	手法 I によるラベル付けの結果例 (上図と下図はそれぞれ違う被験者)	16
図 3.8	商業地域への訪問抽出と行動日記データの一致率	17
図 3.9	買い物時の滞在時間の分布	19
図 4.1	家計調査を基に計算した世帯の年間総出費額	22
図 4.2	商業地域訪問者に関するデータセット作成のフローチャート	23
図 4.3	拡大係数の算出法および概念図	26
図 4.4	拡大係数を用いて計算した推定昼間人口と国勢調査の昼間人口の散布図	27
図 4.5	滞在クラスタ数と商業地域訪問ラベル付きクラスタ数の頻度分布	28
図 4.6	東京都心部における年間総訪問者数	30
図 4.7	各主要都市における年間総訪問者数	31
図 4.8	各時間帯の訪問割合 (AM 2:00 – PM 2:00)	34
図 4.9	各時間帯の訪問割合 (PM 2:00 – AM 2:00)	35
図 4.10	東京都心部における訪問頻度	36
図 4.11	東京都心部における訪問者の自宅からの距離	36
図 4.12	推定年間出費額と年間販売額の空間分布	38
図 4.13	推定年間出費額と年間販売額の散布図	39
図 5.1	イオンレイクタウンおよび商業集積統計データの位置関係	42

表目次

表 2.1	商業地域訪問者の分析における利用データの特性比較	6
表 3.1	行動日記データの滞在目的	10
表 3.2	各手法の平均一致率及び平均調整一致率	18
表 3.3	各手法の違い一覧	18
表 4.1	住宅土地統計調査における世帯の年間収入階級	22
表 4.2	時間帯とその項目名	25
表 4.3	拡大係数を用いた推定人口の決定係数と RMSE	27
表 4.4	各商業地域の年間総訪問者数 (上位 20 位まで)	32
表 4.5	各商業地域の 1 店舗当たり年間総訪問者数 (上位 20 位まで)	33

第1章 はじめに

1.1 研究の背景・目的

1.1.1 小規模商業地域の現状

近年、大型商業施設の開業や、オンラインショッピングの誕生、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。中小企業庁が実施する商店街実態調査報告書によると、全国を対象に抽出された商店街のうち1970年では、39.5%が「繁栄している」と回答しているのに対し（松隈久昭，2006），2009年にそのように回答したのは1.0%だけであった（中小企業庁，2010）。

商店街に代表される小規模な商業地域は、単なる商業機能だけでなく地域の防災・防犯、地域文化の保存や継承などの機能をあわせ持っている。また、自動車を利用出来ない高齢者にとって、自動車利用が前提の郊外型ショッピングセンターでの買い物は難しいため、徒歩移動が可能な距離に商店街があることは、いわゆる「買い物弱者¹」の防止にも効果があるといえる。したがって、商店街の衰退を食い止めることは、地域コミュニティの保全のためにも重要な課題であるといえる。

では、現況の商店街が抱える問題とは一体何だろうか。2009年の商店街実態調査報告書によると、問題点として後継者難（51.3%）の次に多く選択された項目は、魅力ある店舗が少ないこと（42.7%）であった（中小企業庁，2010）。魅力のある商店街づくりとして近年注目を集めている例として、様々なアートイベントの開催²や、地域キャラクターの開発³、大学生によるカフェの運用⁴などが挙げられる。これらの施策は一定の効果を上げた地域もある一方で、地域の戦略がない状況で他地域の成功例をそのまま真似しても持続的な活性化は期待できないという指摘もなされている（木下，2009）。

そこで本研究では、魅力ある店舗を形成するために行う戦略立案の際に、商業地域訪問者の分析を行うことを提案する。一般に大規模なチェーン店や大型商業施設では、訪問者分析をもとにしてマーケティング施策を講じていることが多い。特に、最近ではCRM（顧客関係管理）の観点から、顧客の生活パターンや嗜好を考慮に入れたきめ細やかなマーケティングによって競争力をあげる例が増えてきた⁵。したがって、商店街をはじめとする小規模な商業地域においても十分な訪問者分析を行なうことが出来れば、その情報に基づいた

¹ 流通機能や交通網の弱体化とともに、食料品等の日常の買い物が困難な状況に置かれている人々のこと。経済産業省では買い物弱者数を600万人程度と推計している。

² 例えば、青森県八戸市商店街のうわさプロジェクトなど。

³ 東京都押上・業平橋地区5商店会のキャラクター「おしなりくん」など。

⁴ 愛知県名古屋市中区日比野商店街や栃木県宇都宮市ユニオン通り商店街の例などが挙げられる。

⁵ カルチュア・コンビニエンス・クラブ株式会社によるTポイントプログラムが有名。

第1章 はじめに

戦略を立案することができ、魅力ある店舗の形成に活用できると考えられる。

1.1.2 商業地域訪問者に関する既存データセット

商業地域訪問者の分析を行なうための既存のデータセットは網羅性が低いという問題がある。たとえば、自治体が実施する商業実態調査は、来街者または商店主にアンケートをとる形式が一般的であるため、自治体ごとに調査項目が異なり、定期的に行なわれていることは稀である。また大型チェーン店などでは、POS データやポイントカード履歴を利用して、訪問者の分析を行うことができるが、このようなデータは小規模な商業地域をカバーしない。つまり、商業地域訪問者に関する既存データセットには、商店街をはじめとする小規模な商業地域を含みながら広域をカバーするものは存在しないといえる。

一方で近年、携帯電話内蔵の GPS データを大量に集計することによって、時々刻々と変化する都市の動態や人の移動が明らかになりつつある。携帯電話内蔵の GPS データは、定期的に広域の情報を取得する面でメリットがあり、交通や観光分野では、人の行動調査に携帯電話内蔵の GPS データを利用するものが増えている。

1.1.3 本研究の目的・独自性

そこで、本研究は小規模な商業地域を含む南関東エリアを対象に、大規模・長期の GPS データを用いて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。ここでの商業地域訪問者の特性とは、商業地域での滞在時間や、滞在時間帯、滞在頻度、自宅との距離などのことをいう。また、GPS データから得られた推定自宅位置と他統計データを組み合わせることによって、サンプルデータを実数に変換し、個人の年間出費額（食費と衣服費）を推定する手法を提案する。

大規模・長期の GPS データを利用した研究はまだ少なく、本研究のように商業地域訪問者の特性抽出のために使用した例はまだない。また、GPS データと他統計データを統合して年間出費額を推定するような研究例も過去に行なわれていない。

1.2 本論文の構成

本論文の構成は以下の通りである (図 1.1)。ここまでの第 1 章では本研究の背景や目的、独自性について説明した。背景では、商店街をはじめとする小規模な商業地域の現状や課題、対策などを述べ、商業地域訪問者に関する情報の重要性を説明した。また本研究の目的は、小規模な商業地域を含む南関東エリアを対象に、長期・大量の GPS データを用いて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である。

第 2 章では、本研究に関連した既往研究を述べる。前半は、過去に行なわれた商業地域訪問者の分析事例として、行政や大型チェーン店が行なった例などを挙げ、使用データによる特性の違いを比較する。後半は、GPS データを利用したひとの行動分析事例を説明しよく用いられる手法などを述べた。

第 3 章では、GPS データから商業地域への訪問を抽出する手法を説明する。また、その手法の信頼性を複数の抽出条件下で比較検証を行なうことによって説明する。

第 4 章では、第 3 章で検証された手法を、大規模・長期の GPS データに応用して作成した商業地域訪問者に関する基礎データセットについて述べる。GPS データを全数に拡大する際の手法についてもここで述べる。また、得られた結果を分析し訪問者の地域特性などを考察するとともに、商業統計と比較することでその信頼性を検証する。

第 5 章では、結論として本研究で得られた成果を総括し、課題や展望について述べる。

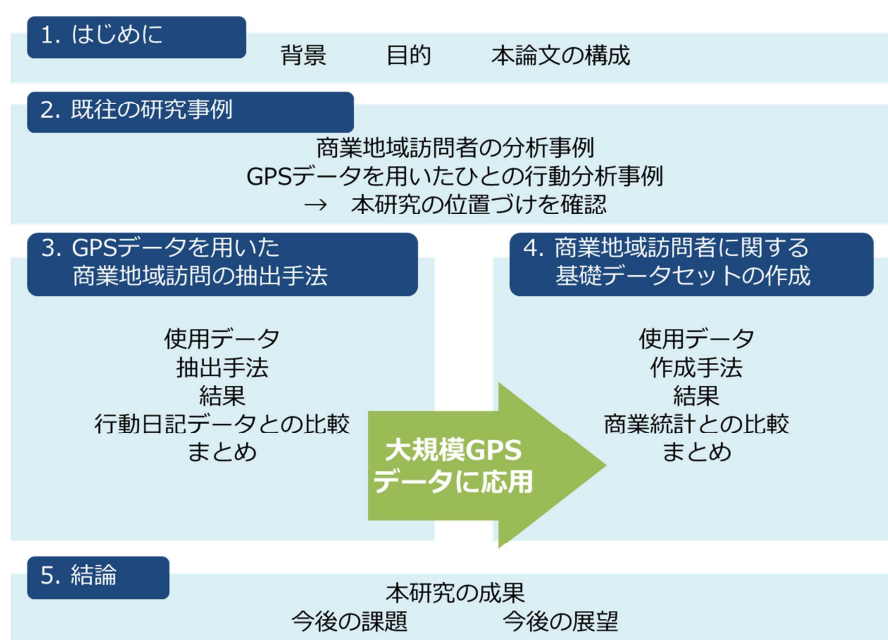


図 1.1 本論文の構成

第2章 既往の研究事例

ここでは、本研究に関連した既往研究を述べる。2.1 では、過去に行なわれた商業地域訪問者の分析事例をその利用データごとに説明する。2.2 では、GPS データを利用した一般的な行動分析に関する研究例を紹介する (図 2.1)。2.3 ではそれらを踏まえ、本研究が置かれている位置についてまとめる

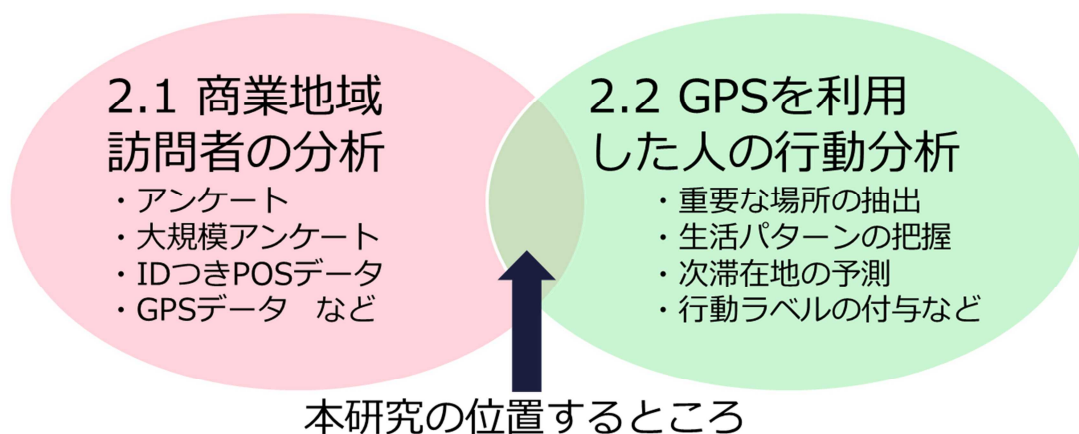


図 2.1 第2章の構成の概念図

2.1 商業地域訪問者の分析事例

ここでは、商業地域訪問者の分析を行なった既往研究について、その使用データごとにレビューする。また、表 2.1 に以下で述べるデータの特性を比較したものをまとめた。

2.1.1 アンケート調査

商業地域訪問者に対するアンケート調査は、自治体による商業実態調査や都市計画分野の研究などで昔からよく行なわれてきた。例えば最近では羽藤ら(2002) が松山市で、梅井ら (2005) が福岡都心部で行った例や、自治体によって商業実態調査を行った例 (東京都文京区, 2012) が挙げられる。しかしながらこれらの例は全て対象エリアが狭く、調査項目に統一性が欠け、また定期的に実施されることが少ないといえる。

一方で、対象エリアが都市圏レベルの大規模なアンケート調査 (パーソントリップ調査) 結果を利用して訪問者を分析した例も存在する。パーソントリップ調査は一日の移動に関する情報を都市圏人口の数%を対象に調査したもので、主な調査項目は、個人属性、出発地、到着地、移動時間、移動目的、移動手段などである。パーソントリップ調査を利用した研

究例は、滞在時間の長さと購買結果の関係を分析したものや (島崎ら, 2009), 大型ショッピングセンターの収益予測の乖離要因について考察したもの (鈴木ら, 2012) が挙げられる。パーソントリップ調査は、対象エリアが広域、非集計データが利用可能であるという利点がある一方で、調査頻度が10年程度と長いことや、ほとんどの対象日が平日のみ、買い回り行動が記述されづらいなどの欠点も存在する。

2.1.2 IDつきPOSデータ

以上の2つの例はアンケート調査結果をもとに訪問者の分析を行なっていたが、ICTの発達によって、わざわざアンケートを書かなくても、ポイントカードや会員カードのIDをPOSデータ⁶と結びつけることによって、顧客の属性や購買履歴を自動的に収集することができるようになった。ポイントカードや電子クーポンが広く普及している大型チェーン店では、このIDつきPOSデータを利用して、顧客をセグメントごとに分類し、きめ細やかなマーケティングを行なっている (安岡ら, 2012)。IDつきPOSデータを利用した研究例としては、行動パターンの抽出 (関, 2003) や競合店間での購買品の関連性分析 (渡辺ら, 2005)、嗜好に基づいたレコメンデーション (小沢, 2006)、顧客行動予測システムの構築 (石垣ら, 2011) などが挙げられる。また高松ら (2013) は交通系ICカードの一種であるPASMOカードのデータと、PASMO IDが付与されたPOSデータを利用して、乗降時刻や鉄道利用頻度も併せた訪問者分析をしている。

しかしながら、導入や運用、データ分析にかかるコストが原因で、個人経営などの小規模な店舗ではPOSシステムが導入されていないことが多い。最近では、タブレット端末がそのままPOSレジになるアプリケーション (例えば Square Register, ユビレジなど) が月額1万円以内 (一部無料) で公開されており、低コストになったといえるが、現時点では一部の普及に留まっている。

2.1.3 モバイル空間統計

携帯電話の位置情報をもとにした「モバイル空間統計」という新しい統計データを利用して商業地域訪問者の分析を行なった例もある (千葉県柏市, 2012)。これは、NTTドコモの携帯電話ネットワークの運用データから、時々刻々と変化する人口をメッシュ単位や市区町村単位に集計したデータである (寺田ら, 2012)。運用データは毎日集計されているので、任意の期間を対象にすることができ、性別や年代別の集計データを利用出来ることが

⁶ なお、POSとはpoint of salesの略であり、POSデータは店舗で商品販売時に毎回記録する販売情報のことである。

利点である。実際に柏市の例では、曜日、時間帯毎の商業地域内における人口変化を分析し、「柏駅周辺は男女ともに20代に支持され、平日では60歳代の女性の人口が多い」というような結果を得ている。

しかしながら、モバイル空間統計は2次メッシュなどの大きなメッシュ単位では高い信頼性をもつものの、より細かい集計単位では人口のばらつきが大きい地域において推計値の扱いに留意が必要となる可能性が大藪ら(2012)によって指摘されている。また、モバイル空間統計は集計データなのである人の一日の行動を追っていくことはできない。

2.1.4 GPS データ

一方で携帯電話のGPSデータをアンケート調査の代替として用いることで商業地域訪問者の回遊行動を分析した例も存在する(藤田ら, 2003)。GPSデータを利用するため、モバイル空間統計とは違い、人の一日の行動を追うことできる。従来、GPSデータを利用する研究は、被験者に承諾を得なくてはならないため、対象人数や対象エリアが限られてしまうという欠点があった。しかしながら、GPSデータセットの蓄積にともない、秋山ら(2012)は17万人分という大規模の匿名GPSデータ(混雑統計(R)データ)を利用して、東京都におけるある1日の商業地域訪問者の特性分析を行なっている。また、大規模なGPSデータはデータ量が膨大で扱いが難しいという欠点はあるものの、広域で定期的に非集計データをもとにした分析を行えるという点で、大きな可能性を秘めているといえる。

表 2.1 商業地域訪問者の分析における利用データの特性比較

	調査 頻度	対象 エリア	集計/非集計 データ	欠点
アンケート	不定期な ことが多い	狭い (商店街 単位など)	非集計	調査項目に 統一性がない
大規模アンケート (パーソントリップ調査)	10年程度	都市圏	非集計	買い回り行動が 記述されづらい
IDつきPOSデータ	毎日, 毎時間	店舗	非集計	導入・運用・ データ分析コスト
モバイル空間統計	毎日, 毎時間	全国	集計	過疎地域での 信頼性が低い
携帯電話GPSデータ	不定期な ことが多い	狭いことが 多い	非集計	GPSの誤差
大規模携帯電話 GPSデータ (混雑統計(R)データ)	毎日, 毎時間	全国	非集計	データが膨大 研究例が少ない

2.2 GPS データを用いたひとの行動特性抽出事例

2.1 では、商業地域訪問者の分析に関する既往研究について説明し、特に大規模 GPS データが有力であることを述べた。そこで、ここでは GPS データを利用して行動特性を抽出した研究事例について説明する。

2.2.1 ひとの行動抽出に関する事例

GPS データを利用したひとの行動抽出に関する代表的な研究例は、その人がよく訪れるような重要な場所の抽出 (Ashbrook *et al.*, 2003; Zheng *et al.*, 2009) や、生活パターンの把握 (Ye *et al.*, 2009; Liao *et al.*, 2005; Jiang *et al.*, 2012; 西野ら, 2008), 移動交通手段の推定 (Schüssler & Axhausen, 2008), 次に訪れる場所の予測 (Ashbrook *et al.*, 2003; 山田ら, 2010) などが挙げられる。

これらの研究の多くが、GPS データから距離や時間の閾値をつかって滞在点を抽出し、k-means 法をはじめとするクラスタリングアルゴリズムを利用して、何度も訪れるような重要な場所を抽出するところから分析を始めている (Ashbrook *et al.*, 2003; Zheng *et al.*, 2009; Ye *et al.*, 2009)。

2.2.2 行動ラベルの付与に関する事例

また、ひとの行動抽出に関する事例の中で特に本研究と関連が強いものとして、GPS データに自宅や勤務地、買い物などの行動ラベルを付与する例が挙げられる。

Liao ら (2007) は条件付き確率場を利用して、GPS 利用者の行動を学習し、労働や自宅での睡眠、娯楽、乗車/降車などの行動を推定するモデルを構築している。さらにこのモデルを検証した結果、85%の精度が得られている。しかしながら、この手法は1人分・1週間のデータを処理するのに標準的なデスクトップ PC で1分かかると書いてあり、大規模・長期の GPS データ使用する本研究に応用することは難しいと考えられる。

大規模・長期の GPS データを利用して行動ラベルを付与した例としては、秋山ら (2012) の例が挙げられる。秋山ら (2012) は、同一商業地域での滞在時間にしきい値を設け、それ以下であれば商業地域訪問者、それ以上であれば労働者という簡便な条件を利用して、各商業地域の商業的な集客力と雇用総出力を分析している。しかしながら、この研究は他統計との比較検証などが行なわれていない。

2.3 まとめ

第2章では、本研究に関連した既往研究を説明し、本研究が置かれている位置を確認した。本研究は商業地域訪問者の分析と、GPS データを利用したひとの行動特性抽出との分野が重なる部分に位置する。

従来の商業地域訪問者の分析で使用されたデータセットは様々な種類があるが、対象エリアが小さく、小規模商業地域を網羅しないという課題があった。しかしながら、近年利用出来るようになった大規模な GPS データを用いる方法は、データ量が膨大で扱いが難しいという欠点はあるものの、広域で定期的に非集計データをもとにした分析を行えるという点で、大きな可能性を秘めているといえる。

また、GPS データを利用して行動特性を抽出した研究事例には、様々なものがあるが、その多くが、GPS データから距離や時間の閾値をつかって滞在点を抽出し、クラスタリングアルゴリズムを利用して、何度も訪れるような重要な場所を抽出するところから分析を始めていた。GPS データから買い物などの行動ラベルを精度良く付与する既往例はあるものの、アルゴリズムが複雑であるので大規模・長期の GPS データに応用することは難しい。簡便な方法で商業地域訪問者と労働者を判別した研究例はあるが、これは他データとの比較検証が行なわれていないという問題点がある。

このような背景を踏まえると本研究は、大規模・長期の GPS データからなるべく簡便な方法で商業地域訪問のラベルを付与し、既存統計との比較検証を行った初めての研究であるといえる。

第3章 GPS データを用いた商業地域訪問の抽出手法

この章では GPS データから抽出する商業地域への訪問を抽出する手法を説明し、その手法が実際に信頼できるものなのかを比較検証を行なうことで説明する。

3.1 使用データ

3.1.1 商業集積統計

本研究では商業集積の分布と位置情報を把握するために 2011 年の東京都の商業集積統計を用いた。商業集積統計とは秋山ら (2011) によって開発されたデータセットである。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の分布をポリゴン表したものであり、もちろん小規模の商業地域も含まれている。本研究では、2011 年の南関東 (東京都、埼玉県、千葉県、神奈川県) のデータ (全 8,774 カ所) を用いた。各商業集積は、形状 (ポリゴン)、商店数、業種毎の商店数、チェーン店の数などの情報を保持している。

商業集積統計データの一部を図 3.1 に示す。池袋や、新宿、渋谷などの都心部周辺で大規模な集積が起こっているのが分かるほか、鉄道駅の前に列を成して集積が起こっているのが確認出来る。

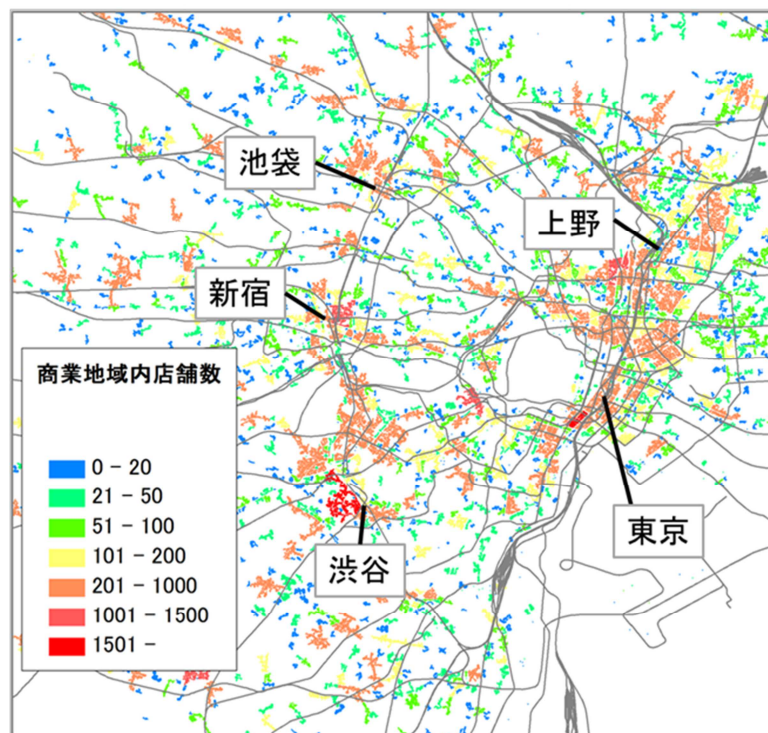


図 3.1 商業集積統計による商業地域の空間的分布と商業地域内店舗数

3.1.2 GPSデータおよび行動日記データ

GPSデータから商業地域訪問を抽出し、その結果を検証するためには、GPSデータと滞在目的が一体になっているデータが必要である。そこで本研究では金杉ら(2012)によって収集された184人分の25日間にわたるGPS履歴データと行動日記データを利用した。GPS履歴データは、AndroidスマートフォンのGPS測位機能を利用したアプリケーションを被験者が使用することによって得た。

行動日記データは、被験者がWeb画面から滞在地とその滞在目的を登録することによって得たものである。滞在目的の選択肢一覧を表3.1に示す。なお、本研究では行動日記データの中から、滞在範囲(矩形)と滞在時間、滞在目的の3つの情報を利用した。GPS履歴データと行動日記データの詳細については金杉ら(2012)を参照されたい。

本研究では、全184人分のデータから滞在目的の項目番号が4, 5, 6を含み、かつその滞在場所が東京都内にあるという条件で、123人を抜き出して使用した。

表 3.1 行動日記データの滞在目的

項目番号	滞在目的
1	勤務先(帰社を含む)
2	通学先(帰校を含む)
3	自宅
4	買い物
5	食事・社交・娯楽(日常生活圏内)
6	観光・行楽・レジャー(日常生活圏外)
7	その他私用(塾・習い事・通院・送迎等)
8	その他業務

3.2 商業地域訪問の抽出手法

ここでは GPS データから商業地域の訪問を抽出する手法を説明する．全体のフローチャートを図 3.2 に示す．まず GPS データから距離や時間のしきい値をつかって滞在点を抽出した．次にクラスタリングアルゴリズムを利用して，何度も訪れるような重要な場所を滞在クラスタとして抽出する．その後，自宅・勤務地・商業地域訪問の順にラベルをつけ，商業地域への訪問を抽出するという手続きをとる．最後にこの手法で得た推定データと，行動日記データを比較して，手法の信頼性を検証する．以下に，この手法の詳細を述べる．

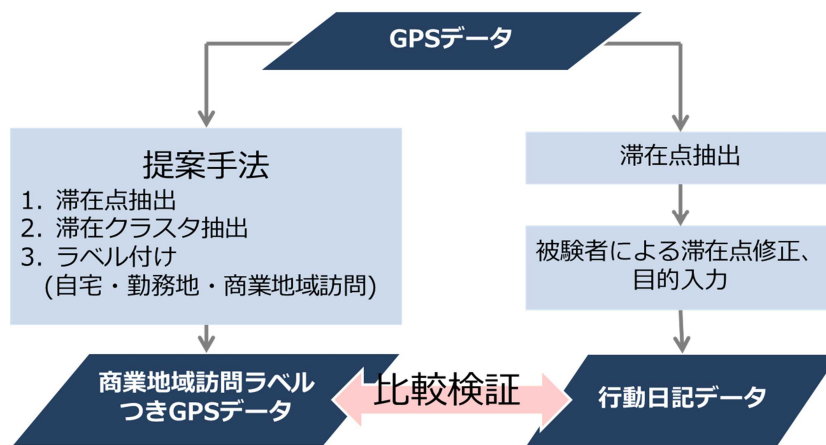


図 3.2 GPS データから商業地域への訪問を抽出するフローチャート

3.2.1 滞在点の抽出

はじめに，GPS データの各位置座標から滞在点の抽出を行なった．ここでの滞在とは，利用者がある一定時間，一定範囲の中に留まることをいい，滞在点は構成する GPS データの代表点のことをいう．なお，各滞在点は構成する GPS 履歴の最大 (最小) 緯度・経度情報を持つほか，到着時間や，出発時間，滞在時間という時間情報も保持する．本研究では滞在の条件を 5 分以上，半径 200 m の範囲に留まることとした．

3.2.2 滞在クラスタの抽出

本研究では，複数日の GPS データを利用しているため，例えば自宅付近の滞在点は複数存在する．そこで，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った．クラスタリングの際のアルゴリズムには DBSCAN を利用した．なお，滞在クラスタは構成する滞在点の中心の緯度・経度情報を持つほか，滞在時間の平均や，滞在時間帯，滞在頻度の情報も保持する．

DBSCAN は入力データ中の密度が高い集団をそれぞれひとつのクラスタとして取り出すアルゴリズムであり、西野ら(2008) も GPS データのクラスタリングに使用している。複雑な形状のクラスタも抽出可能であり、K-means 法のような代表的なクラスタリング方法に比べると、あらかじめクラスタ数を設定しなくてよいという特長がある。DBSCAN は距離の閾値 Eps と対象数の閾値 $MinPts$ という 2 つのパラメータを用い、半径 Eps 以内にある対象の接続関係を図 3.3 のように定義し、接続している対象が $MinPts$ 個以上あるとき、同じクラスタに分類するというものである。図 3.3 では、どの点からも Eps 以上の距離がある点 P は、ノイズとして除去される。詳しいアルゴリズムに関しては、Ester *et al.* (1996) を参照のこと。本研究では、 Eps を 200 m、 $MinPts$ を 2 と設定して計算した。

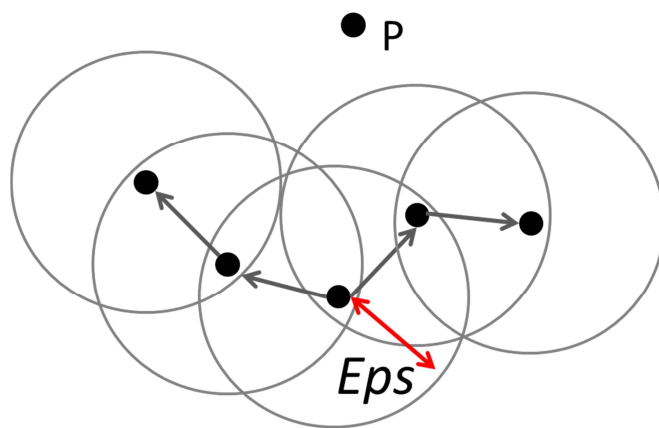


図 3.3 DBSCAN のアルゴリズム概念図

3.2.3 自宅・勤務地・商業地域のラベル付け

各クラスタをそれぞれ分析し、滞在点を持つ情報に応じて自宅・勤務地・商業地域訪問のラベルをつけた。本研究では、自宅・勤務地・商業地域訪問の順にラベルづけした。なお、商業地域訪問のラベルについては、Ⅰ. 滞在点に付与、Ⅱ. 滞在クラスタに付与という 2 通りの計算を行なった。以下にその詳細を述べる。

始めに各クラスタの滞在点が一番多く深夜の時間帯(AM 00:00 から AM 03:59) に属するクラスタを自宅とした。次に、勤務地は平日の滞留点の中で一番多く日中の時間帯 (AM 12:00 から PM 3:59) に属するクラスタとした。

最後に、商業地域訪問のラベルを付与した。商業地域訪問は自宅や勤務地での滞在と比較すると訪問頻度が少ないことを考慮して、Ⅰ. 滞在クラスタに付与、Ⅱ. 滞在点に付与という 2 通りの計算を行なった。どちらの計算でもそれぞれの位置情報と滞在時間の情報を利用して、商業地域訪問のラベルをつけた。具体的には、

- ・ 代表点 (Ⅰの場合は滞在クラスタの中心点, Ⅱの場合は滞在点を構成する GPS データからなる矩形) が商業地域ポリゴンに含まれる
- ・ 滞在時間 (Ⅰの場合, クラスタを構成する滞在点の滞在時間の平均) が 10 分以上 4 時間未満

という2つの条件をどちらとも満たすとき, 商業地域訪問ラベルをつけた (図 3.4). 滞在時間の条件をもうけた理由を以下に説明する. まず, 商業集積ポリゴンの位置情報を利用するだけでは, 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) まで含んでしまう可能性がある. さらに, 商業集積ポリゴンは鉄道駅付近に集中しているため, ポリゴンに含まれるという条件だけでは, 駅での乗り換え行動や乗車前の待機行動を誤って抽出してしまう可能性がある. そこで, 滞在時間の上限閾値と下限閾値とを設定することで, これらの誤差原因を除去した. 本研究では, 滞在時間の上限閾値として4時間, 下限しきい値として10分を用いた. 上限しきい値は, 同一商店街での回遊時間は長くても概ね4時間以内という研究 (大岩ら, 2005) から, 4時間に設定した. 下限しきい値は, 電車の待ち時間・乗り換え時間はおおむね10分以内だとする統計データ⁷を参考にして今回は10分と設定した.

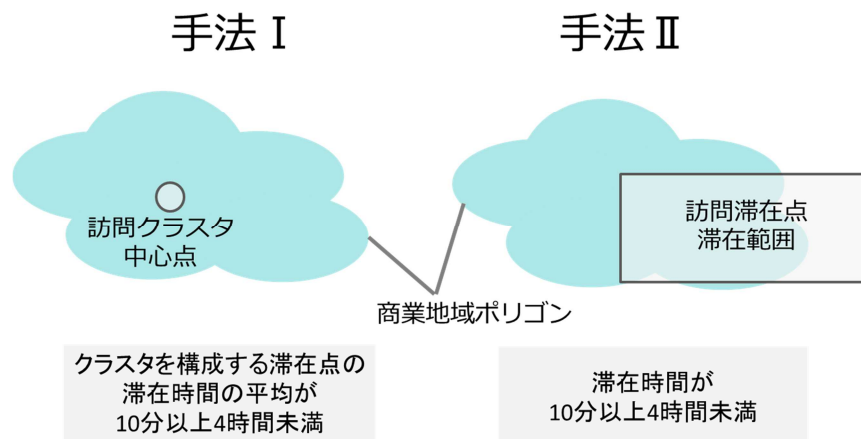


図 3.4 商業地域訪問ラベルの付与条件概念図

⁷ 仙台都市圏パーソントリップ調査(宮城県)では待ち時間が10分以内の人が全体の約90%であった。また、京阪神都市圏パーソントリップ調査(京阪神都市圏交通計画協議会, 2000)によると鉄道利用トリップの平均待ち時間は約8分であった。

3.2.4 比較検証方法

以上の手法によって GPS データから商業地域訪問を抽出し、その結果を行動日記データと比較することで手法の検証を行なった。比較の際の正誤判別条件を以下で述べ図 3.5 に表す。

- ・ 手法Ⅰ（商業地域訪問ラベルを滞在クラスタに付与）

行動日記データにおける買い物の滞在範囲と、商業地域ラベルがついた滞在クラスタの中心点が一定の距離（50 m）以内であり、どちらも同じ商業地域ポリゴンと重なる。

- ・ 手法Ⅱ（商業地域訪問ラベルを滞在点に付与）

行動日記データにおける買い物の滞在範囲と、商業地域訪問ラベルがついた滞在地の範囲が重なり、どちらも同じ商業地域ポリゴンと重なっている

本研究では、各被験者の行動日記データにおける買い物行動のうち、以上の条件を満たしたものを一致率として計算した。また、商業地域ポリゴン外での買い物行動による影響を除外するために調整一致率 r' を式のように計算した。

$$r' = r/R$$

ただし、行動日記データにおいて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を R とし、3.3.4 で説明した比較検証手法で得られた各被験者の一致率を r とする。

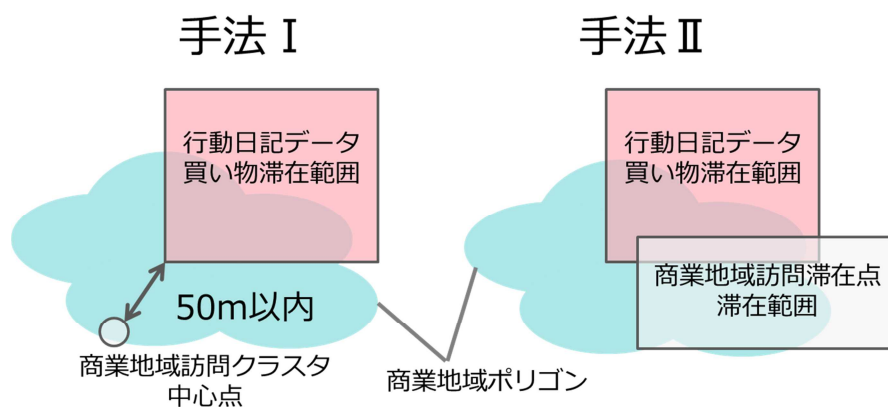


図 3.5 正誤判定条件の概念図

3.3 結果

3.3.1 滞在地の抽出結果

滞在地の計算結果例を図 3.6 に示す。この図において桃色の点が GPS 履歴，黄色のピンが滞在点の中心座標を表す。滞在地を計算した結果，移動途中の GPS データが省かれ，滞在地が抽出出来たことが分かる。なお対象被験者の平均滞在点数は 106.8 個，平均滞在時間は 280 分であった。

3.3.2 滞在クラスタの抽出結果

滞在クラスタの計算結果例を図 3.6 に示す。滞在クラスタを計算した結果，距離的に近い滞在地同士が一つのクラスタとしてまとめられたことがわかる。また，滞在地の密度が低い部分ではクラスタが形成されずにノイズとして除去された様子が見て取れる。なお，対象被験者の滞在クラスタの平均個数は 7.4 個であった。

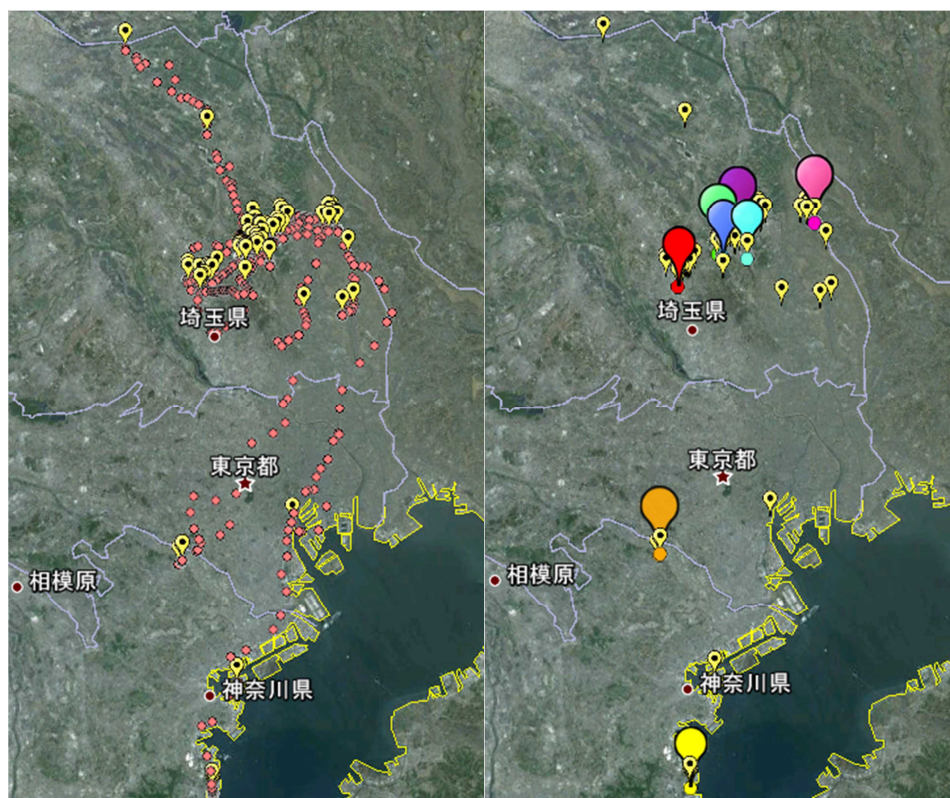


図 3.6 GPS 履歴・滞在地・滞在クラスタの分布例
(左図：桃色点は GPS 履歴を，黄色ピンは滞在点を表す)

(右図：黄色ピンは滞在点を，各色のバルーンはそれぞれ滞在クラスタを表す)

3.3.3 自宅・勤務地・商業地域訪問のラベル付けの結果

手法Ⅰによって自宅・勤務地・商業地域訪問のラベルづけをした結果例を図 3.7 に示す。丸印は滞在点を示し、色はクラスごとに異なるように表した。商業地域訪問ラベルが上図のでは 1 カ所、下図では 4 カ所抽出できた。また、手法Ⅱでは前者は 7 カ所、後者は 28 カ所抽出できた。全体でみると、抽出された商業地域訪問の数は、手法Ⅰで平均 1.86、手法Ⅱで平均 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 3.7 手法 I によるラベル付けの結果例 (上図と下図はそれぞれ違う被験者)

3.3.4 比較検証の結果

商業地域への訪問を抽出した結果と、行動日記データでの買い物の行動の一致率を図3.8に示す。一致率は全体的に手法Ⅰよりも手法Ⅱの方が高く、買い物行動を精度よく抽出していることがわかる。

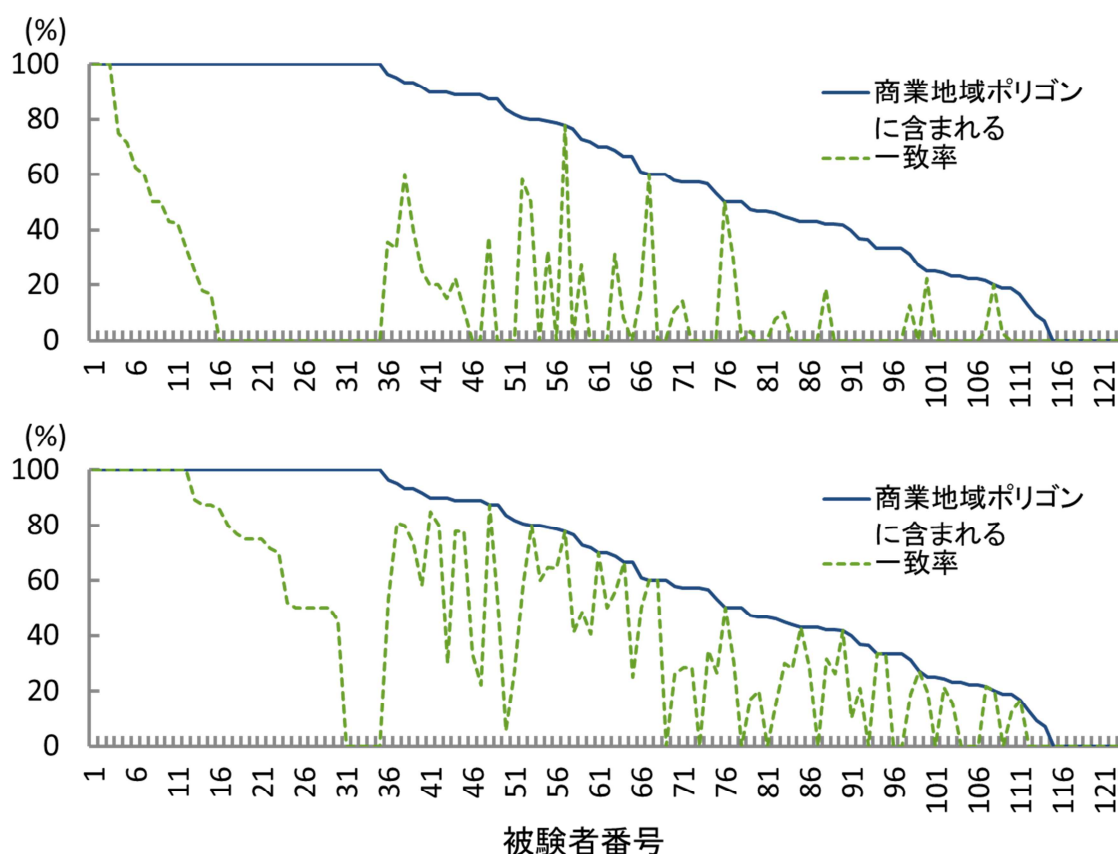


図 3.8 商業地域への訪問抽出と行動日記データの一致率
(上図：手法Ⅰ，下図：手法Ⅱ)

行動日記データにおいて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、平均して 63.67%であった。すべてというわけではないが、6 割以上の買い物行動が商業地域ポリゴン内で行なわれていることがわかる。

また、各手法の平均一致率、平均調整一致率を表 3.2 にまとめた。手法Ⅰの平均調整一致率は 19.69%であり、手法Ⅱの 60.56%と比較するとかなり低い。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ためだと考えられる。というのも本研究では DBSCAN のパラメータである *MinPnt* を 2 と設定しているため、対象期間中に滞在点が 2 点以上ないものはノイズになってしまうのである。さらに、自宅・

勤務地付近での買い物であっても、誤って自宅・勤務地のラベルを付けてしまった可能性も高い。そこでこれらの影響による誤差を調べるために手法Ⅱを一部変更した手法Ⅲ、Ⅳ、Ⅴを設定した。各手法における商業地域訪問フラグをつける対象は表 3.3 の通りである。

表 3.2 をみると、自宅・勤務地フラグやノイズを対象外にした手法Ⅲ～Ⅴは全ての滞在点を対象とした手法Ⅱよりも精度が下がっており、特にノイズを対象外にすることよりも自宅・勤務地の滞在点を対象外することの方が影響は大きかった。さらに、手法Ⅰと手法Ⅴを比較すると、複数の滞在点を一つにまとめて商業地域訪問ラベルをつけることによってさらに精度が下がるということが分かる。

表 3.2 各手法の平均一致率及び平均調整一致率

手法	一致率 (%)	調整一致率 (%)
I	14.09	19.69
II	41.41	60.56
III	28.40	44.07
IV	34.56	49.29
V	20.98	32.04

表 3.3 各手法の違い一覧

手法	商業地域訪問ラベルをつける対象		
	滞在点/滞在クラスター	自宅・勤務地	ノイズ
I	滞在クラスター	×	×
II	滞在点	○	○
III	滞在点	×	○
IV	滞在点	○	×
V	滞在点	×	×

次に、一番精度の良い手法Ⅱでの誤差要因を考察するために、買い物行動における滞在時間の頻度分布を調べた。その結果を図 3.9 に示す。買い物行動の滞在時間の頻度分布は全体的に右裾が長い山形であり、ピークは5分以上10分未満であることから短時間の買い物が多いことが分かる。本研究で設定している滞在時間の条件（10分以上、4時間未満）を満

たすものは全体の 68.34%であった。つまり、手法Ⅱにおける最大の誤差要因は、買い物行動における滞在時間の条件（特に下限閾値）であることがわかる。

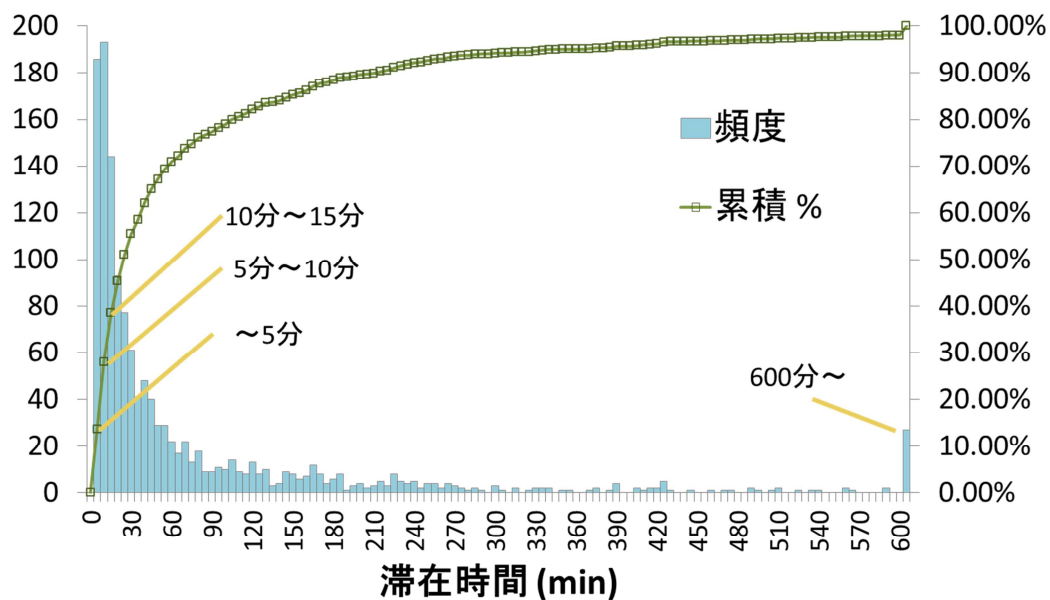


図 3.9 買い物時の滞在時間の分布

3.4 まとめ

この章では GPS データから商業地域への訪問を抽出し、その手法が実際に信頼できるものなのかを比較検証を行なうことで説明した。

抽出手法を以下の通りである。まず GPS データから距離や時間の閾値をつかって滞在点を抽出したのち、クラスタリングアルゴリズムを利用して、何度も訪れるような重要な場所を滞在クラスタとして抽出した。その後、自宅・勤務地・商業地域訪問の順にラベルをつけ、商業地域への訪問を抽出するという手続きをとった。

この手法で得た推定データと行動日記データを比較して手法の信頼性を検証した結果、複数の抽出条件のうち一番結果が良かった手法では 60.56% の一致率であった。また、複数の抽出条件を比較することによって、自宅および勤務地ラベルやクラスタリング時のノイズの中に買い物行動が含まれることによる誤差や、滞在時間の閾値の設定による誤差の影響を評価することができた。

第4章 商業地域訪問者に関する基礎データセットの作成

この章では、第3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結果について述べる。さらに、商業地域訪問者の特性を分析するための計算手法と、既存統計との比較結果も説明する。

4.1 使用データ

4.1.1 GPS データ

商業地域訪問者に関する基礎データセットの作成の際に使用するGPSデータは、できるだけ大規模で長期にわたったものが望ましい。そこで、本研究ではGPSデータとして、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、個人情報の秘匿処理を行った統計データである。今回はゼンリンデータコム社の協力のもと、個人が特定できない状態にした非集計データを用いた(以下大規模GPSデータとする)。GPSデータの保持する情報は、匿名化されたID、時刻、緯度、経度、高度、測位誤差だけであり、個人属性は一切含まれていない。なお、このデータの測位間隔は最短でも5分である。

本研究では対象期間を2010年8月1日から2011年7月31日までの1年間、対象場所を埼玉県、千葉県、東京都、神奈川県、南関東エリアとした。その結果、約22万人分(221,269人)のデータが得られた。

4.1.2 収入・支出に関する統計データ

収入に関する項目には、2008年の住宅土地統計調査の世帯の年間収入階級データ(市区町村第29表)を利用した。このデータは、市区町村別の世帯の年間収入を表4.1のように12階級にわけて集計したものである。本研究ではこのデータから市区町村別の平均世帯年間収入を計算して利用した⁸。

また、支出に関する項目には、2011年の家計調査(家計収支編総世帯 用途分類 1世帯あたり1ヶ月間の収入と支出)を利用した。このデータは市区町村別データが存在せず、都市階級ごと、または政令指定都市・県庁所在地がある市について集計されている。そこで、対象である1都3県の市区町村に関して、政令指定都市ではないものについては、各都市階級の支出データを使用した。また、用途分類を再分類し(参考資料2を参照のこと)、食

⁸ 平均を計算するときは、100万円未満の階級を50万円、1500万円以上の階級を1750万円として計算した。

料費 (外食除く), 外食費, 被服費の3項目で集計した。最後に集計した3項目のデータに12を乗じて年間の世帯支出データとした。図4.1にその結果を示す。

表 4.1 住宅土地統計調査における世帯の年間収入階級

世帯の年間収入階級 (万円)	
～100	600～700
100～200	700～800
200～300	800～900
300～400	900～1000
400～500	1000～1500
500～600	1500～

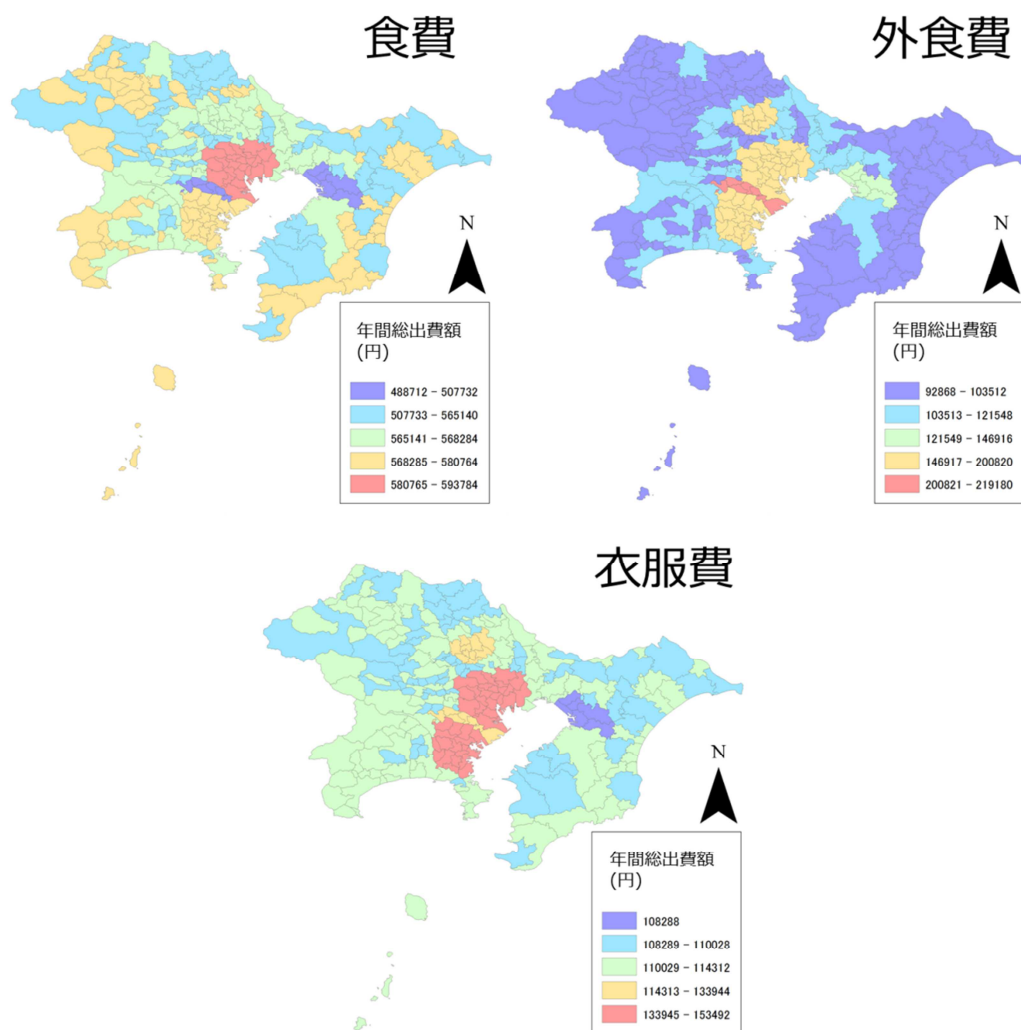


図 4.1 家計調査を基に計算した世帯の年間総出費額
(左上図：食費，右上図：外食費，下図：衣服費)

4.2 商業地域訪問者の基礎データセット作成手法

商業地域訪問者の基礎データセット作成のフローチャートを図 4.2 に示す。大規模・長期の GPS データを第 3 章で述べた提案手法によって商業地域訪問を抽出した。さらに GPS データを実数に拡大するために拡大係数を計算し、支出や収入に関する項目と併せて統合した。最後に、商業地域ごとに集計したデータを他統計データと比較し検証を行なった。以下にその詳細を述べる。

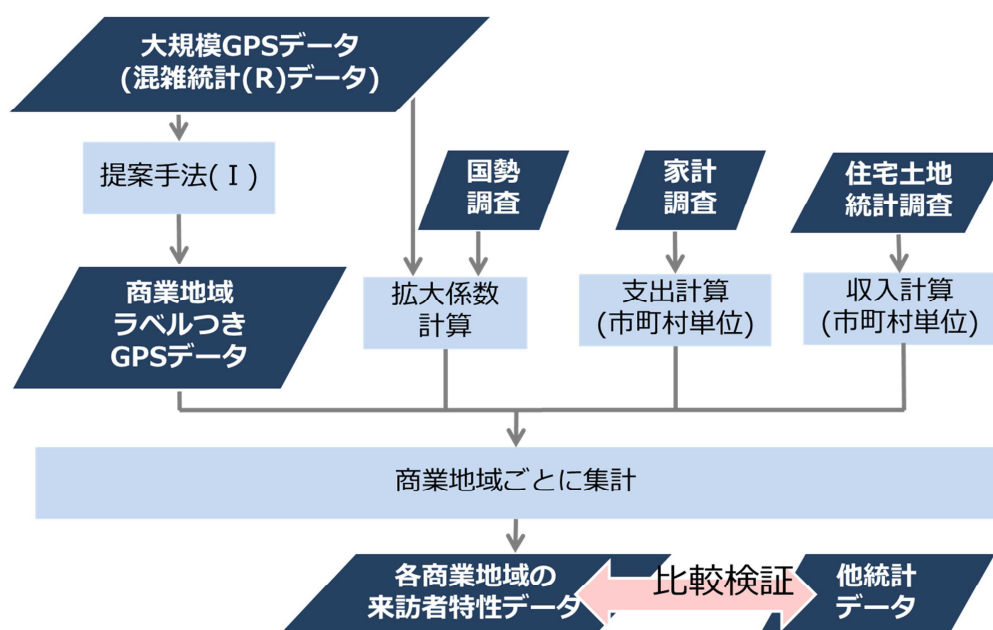


図 4.2 商業地域訪問者に関するデータセット作成のフローチャート

4.2.1 商業地域訪問の抽出

第 3 章で説明した GPS データから商業地域訪問を抽出する手法を、大規模 GPS データに応用した。第 3 章によると、商業地域訪問ラベルを付与する対象は、滞在クラスタにするよりも（手法 I）、滞在点にした方が（手法 II, III, IV, V）精度が良いことが分かっている。しかしながら、本章ではデータの対象人数が約 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。

基本的には第 3 章で述べた手法 I のやり方をそのまま踏襲したが、滞在点の抽出法だけ変更点がある。大規模 GPS データの測位間隔は不規則で最低 5 分という事情を考慮して、滞在の条件を 20 分以上、半径 150 m に変更した。

4.2.2 データセット整備のための諸項目の計算

商業地域訪問者に関するデータセットを整備する際、訪問頻度や訪問時間帯、推定出費額などマーケティングに有効な指標をあらかじめ計算しておけば、各商業地域での訪問者分析や戦略立案に使いやすくなる。そこで、本研究では商業地域訪問ラベルがついた項目について、平均滞在時間、訪問頻度、訪問時間帯割合、平均世帯年間収入、平均世帯支出（食費、外食費、衣服費、諸雑費）、推定年間出費額（食費、外食費、被服費、諸雑費）を計算した。以下にその計算法を説明する。

- 平均滞在時間 (分)

クラスタ i を構成する滞在点における滞在時間の平均。

- 訪問頻度 (%)

クラスタ i を構成する滞在点における滞在日の総数を、全クラスタにおける滞在日の総数で除した値。

- 訪問時間帯割合 (%)

1日を表4.2のように6分割してaからfまでの時間帯を設定したとき、クラスタ i を構成する滞在点において、訪問時がそれぞれの時間帯に属する割合。なお、訪問日が平日か休日かを区別して計算した。

- 訪問者の自宅からの距離 (m)

商業地域訪問クラスタの中心点と自宅クラスタの中心点との距離。

- 年間世帯収入 (百万円)

自宅クラスタの中心点が属する市区町村の年間世帯収入の値。

- 年間世帯支出額 (円)

自宅クラスタの中心点が属する市区町村の年間世帯支出の値。食費、外食費、衣服費の3項目が存在する。

- 推定年間出費額 (円)

商業地域訪問者に関するデータセットを作成する際、そこに訪れる人の出費額は重要な指標である。そこで、商業地域での滞在時間が長ければ長いほど、そこでの支出額も大きいと仮定して、年間出費額を推定した。具体的には家計調査の支出額を年間総訪問時間で比例配分することによって推定した。つまり平均滞在時間を t_i 、訪問頻度を f_i とするとある人が商業地域クラスタ i に訪問したときの年間総滞在時間 T_i は

$$T_i = 365 f_i t_i$$

とあらわせる。家計調査での年間世帯支出額を P 、平均世帯構成人数 (=2.46) を n とすると、ある人が商業地域クラス i で出費する推定年間出費額 p_i は以下のように計算できる。

$$p_i = \left(\frac{P}{n} \right) \frac{T_i}{\sum_i T_i}$$

なお、推定年間出費額も食費、外食費、衣服費の3項目で計算した。

表 4.2 時間帯とその項目名

項目名	時間帯
a	AM 02:00 – AM 06:00
b	AM 06:00 – AM 10:00
c	AM 10:00 – PM 02:00
d	PM 02:00 – PM 06:00
e	PM 06:00 – PM 10:00
f	PM 10:00 – AM 02:00

4.2.3 実数への拡大計算

この大規模 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。

ここで、サンプル調査のひとつであるパーソントリップ調査の例を用いて拡大処理の説明をする。パーソントリップ調査では、サンプルの属性情報を国勢調査の属性情報と比較して、拡大係数を設けている。これは、同じ属性情報を持つものは類似した行動をするという仮定に基づいた方法である。例えば、(年代、性別、職業)という3つの属性情報をもとに拡大係数を設ける際、(20代、男性、学生)を満たす人が国勢調査では N 人、サンプルでは n 人いたとする。このときのサンプル抽出率は n/N であり、拡大係数はこの逆数である N/n になる (図 4.3)。

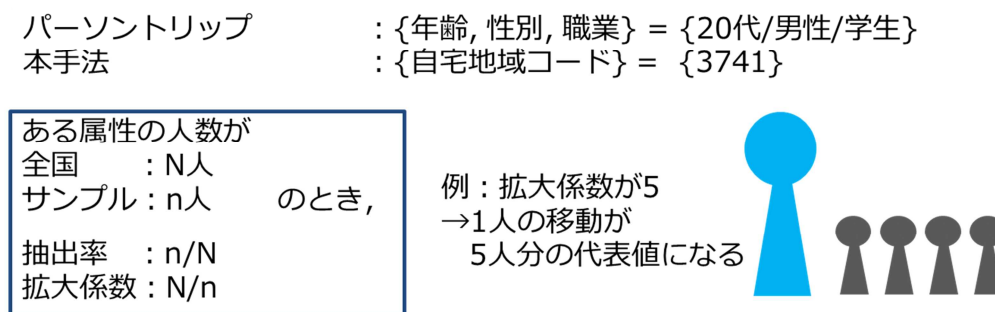


図 4.3 拡大係数の算出法および概念図

以上のような拡大処理を大規模 GPS データに適用する際、年齢・性別などの属性情報は得られないため、本研究では地域内人口をもとにして拡大係数を設けることとする。地域内人口の集計単位は、標準メッシュや行政区分などが考えられるため、複数の集計単位を用いて計算した拡大係数で得られた推定人口を既存統計と比較することによって、最適な集計単位を探索した。

最適な集計単位を探索する方法は以下の通りである。対象範囲は、一都三県（東京都、千葉県、埼玉県、神奈川県）とした。ある平日⁹の 4:00 における大規模 GPS データを地域ごとに集計し、GPS データの夜間人口とする。2005 年の国勢調査における人口を GPS データの夜間人口で除し、それぞれ拡大係数を算出した。なお、人口の集計単位は 3 次メッシュ、4 次メッシュ、町丁目の 3 通りを用いた。最後に拡大係数の再現性を評価するために、求めた拡大係数を GPS データに適用して得た推定人口を、2005 年の市町村別人口（夜間・昼間）と比較した。

拡大係数利用して得られた推定人口と既存統計の人口との比較結果を表 4.3 にまとめた。夜間人口での再現性はどの集計単位でも決定係数がおおよそ 1 であり夜間人口での再現性はかなり高い（表 4.3）。これは拡大係数を算出するのに使用したのが夜間人口であったため、十分に予想できる結果である。一方昼間人口での再現性は夜間人口と比較して下がり、決定係数はおおよそ 0.7 程度という結果になった。集計単位を 3 次メッシュとして計算した推定昼間人口と国勢調査の昼間人口の散布図（図 4.4）をみると、赤丸に示した部分の外れ値の影響が強いことが分かった。赤丸で示した部分の地域を参照すると、右から港区、千代田区、中央区、新宿区、中央区を示していた。一般に、国勢調査で得られる昼間人口は通勤・通学地の住所を元に作成されているため、外回りのような勤務地外の移動や買い物などの

⁹ 本研究では代表として 2011 年 9 月 20 日を選択した。

不規則な移動者は含まれない。そのため、大企業の本社や官公庁の建物、大規模な商業施設が密集するような地域では、統計調査による昼間人口は実際のものよりも過少である可能性が高く、相関関係を弱めたのだと考えられる。

本研究では、それぞれの集計単位での決定係数と RMSE を参考に、一番昼間人口の RMSE が小さい3次メッシュを利用して決定係数を計算した。

表 4.3 拡大係数を用いた推定人口の決定係数と RMSE

集計単位	夜間人口		昼間人口	
	R ²	RMSE	R ²	RMSE
3次メッシュ	0.999	6279.3	0.694	86138.3
4次メッシュ	0.997	10677.0	0.695	86248.4
町丁目	0.999	5318.8	0.692	86454.0

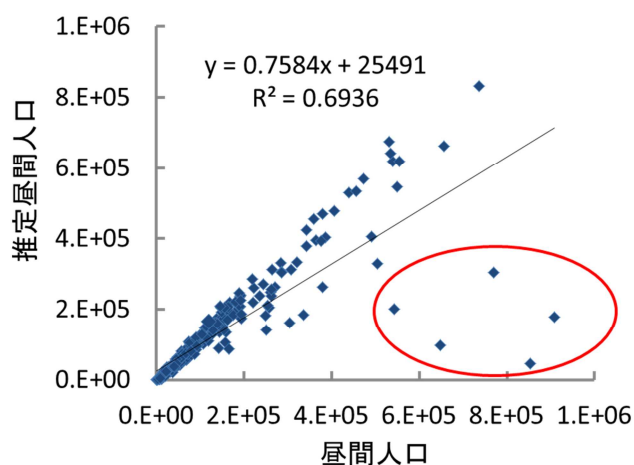


図 4.4 拡大係数を用いて計算した推定昼間人口と国勢調査の昼間人口の散布図
(赤丸内は、右から港区、千代田区、中央区、新宿区、中央区を示す)

4.3 結果

4.3.1 全体の結果

はじめに作成したデータセット全体の結果を述べる。滞在クラスタ数は全体で3,528,790, 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。これらの値をより詳しく考察するために頻度分布をみると, 滞在クラスタ数は6-10の階級をピークとして右に裾が長い山形であり, 25以下の対象者が全体の8割を超えていた(図4.5上図)。また, 商業地域訪問ラベルが付与された人数全体における頻度分布をみると(図4.5下図), 右肩下がりの分布をしており, 5以下の対象者が全体の85%程度であった。

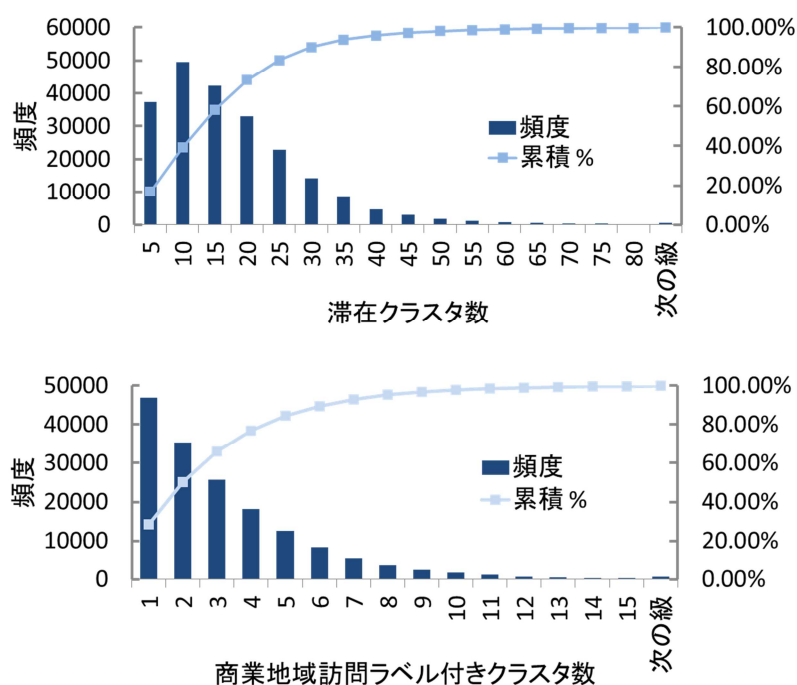


図 4.5 クラスタ数の頻度分布

(上図：滞在クラスタ数, 下図：商業地域訪問ラベル付きクラスタ数)

4.3.2 商業地域ごとの集計結果

次に、拡大係数を利用して商業地域ごとに各項目を集計した結果を以下で述べる。なお、以下の項目（訪問頻度、訪問時間帯割合、訪問者の自宅からの距離）では訪問者の平均値を取っているため、サンプル数が極端に少ないと信用性が減ってしまう。そのため、拡大前の訪問者が4人以上のものを抜き出して計算した。

・ 年間ユニーク訪問者数

都心部では渋谷や新宿、池袋などの主要都市で大きい値が得られた（図 4.6）。さらに横浜や千葉、大宮などの地域でも鉄道駅を中心に数多く訪れている様子が分かる（図 4.7）。また、年間総訪問者数の上位 20 位までを表 4.4 にまとめた。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。訪問者の多い商業地域はほとんどが鉄道駅周辺の商業地域であるという特徴が挙げられる。また店舗数に着目すると、17 位の有楽町駅周辺以外の商業地域では地域内店舗数が 100 店を超え、大規模な集積が起る地域が上位にランクインしたことが分かる。

次に 1 店舗当たりの集客力を把握するために一店舗当たりのユニーク訪問者数を算出し、上位 20 位までを表 4.5 にまとめた。その結果、1 位になったのは秋葉原駅西ダイビル周辺の商業地域であり、一店舗あたり年間訪問者数は約 1 万 4 千人であった。さきほどの訪問者数の上位 20 位までの商業地域と比べて、店舗数は 100 店以下のものがほとんどを占めるという結果になった。また、2 位の二子玉川高島屋をはじめ 6 位の川口駅そごう、9 位のゲートシティ大崎、11 位のららぽーと横浜など、大型デパートや大型商業施設のランクインが目立つ結果となった。

・ 訪問時間帯割合 (%)

訪問時間帯割合を時系列で見ると（図 4.7, 図 4.8）、明け方は割合が低く、通勤時間帯から帰宅時間帯は高く、深夜にまた低くなるという生活パターンが見て取れる。また、平日と休日を比較すると、特に時間帯 b (AM06:00-AM10:00) と時間帯 e (PM06:00-PM10:00) の差が大きいことが分かる。

・ 訪問頻度 (%)

都心部の主要都市では 10%台の地域が多くみられた（図 4.10）。訪問頻度が 10%台というのはつまり、10 日から 19 日に 1 回程度の訪問頻度といえる。また、訪問頻度が高い、あるいは低い地域は周辺部に位置することが多く、これは訪問人数が少ない地域と一致する。

サンプル数が小さいため過剰な値が出てしまった可能性も考えられる。

- ・ 訪問者の自宅からの距離 (m)

訪問者の自宅からの距離の空間分布をみると (図 4.11), 明らかに都心部で大きく, 周辺部で小さいという特徴がみられる。都心部の主要都市は商圈が広く, 遠くの地域からも集客している様子が分かる。

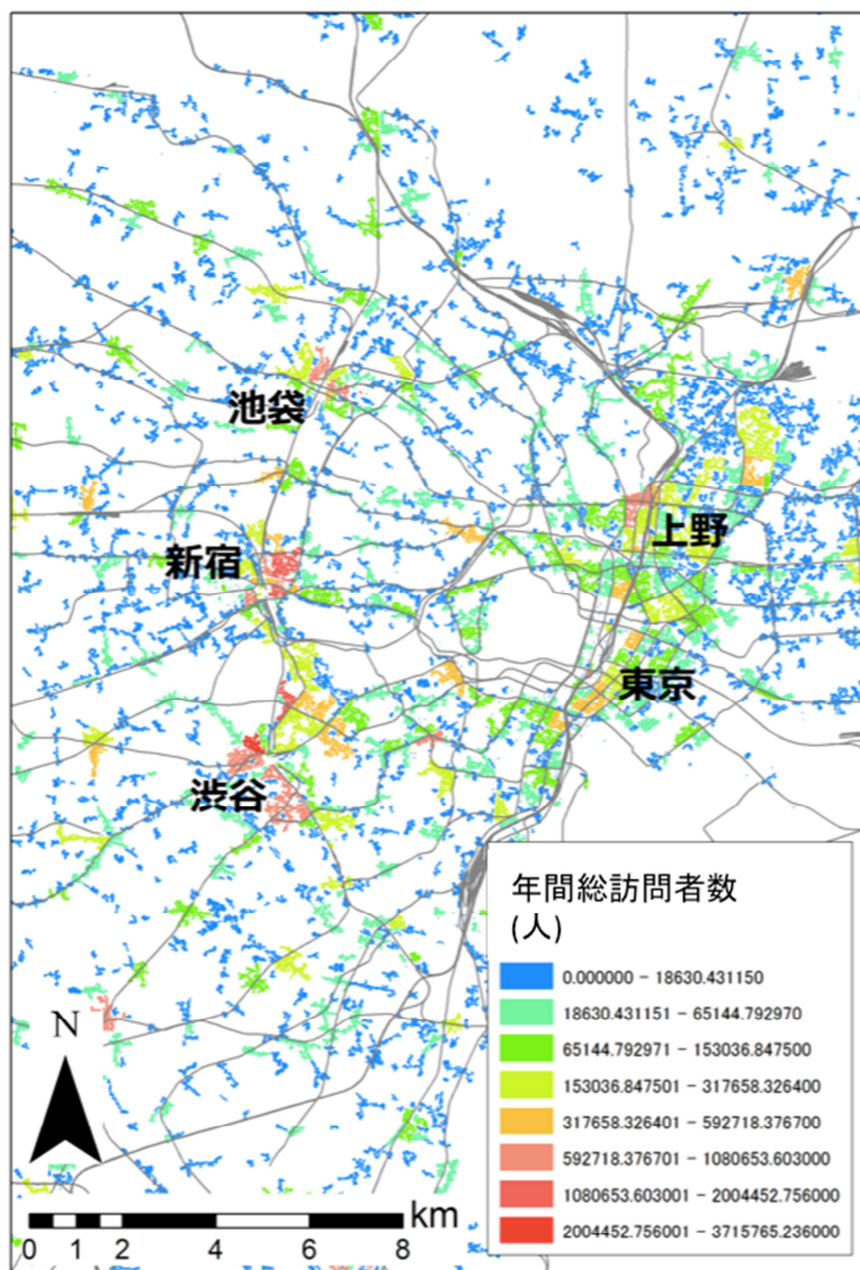


図 4.6 東京都心部における年間ユニーク訪問者数

第4章 商業地域訪問者に関する基礎データセットの作成

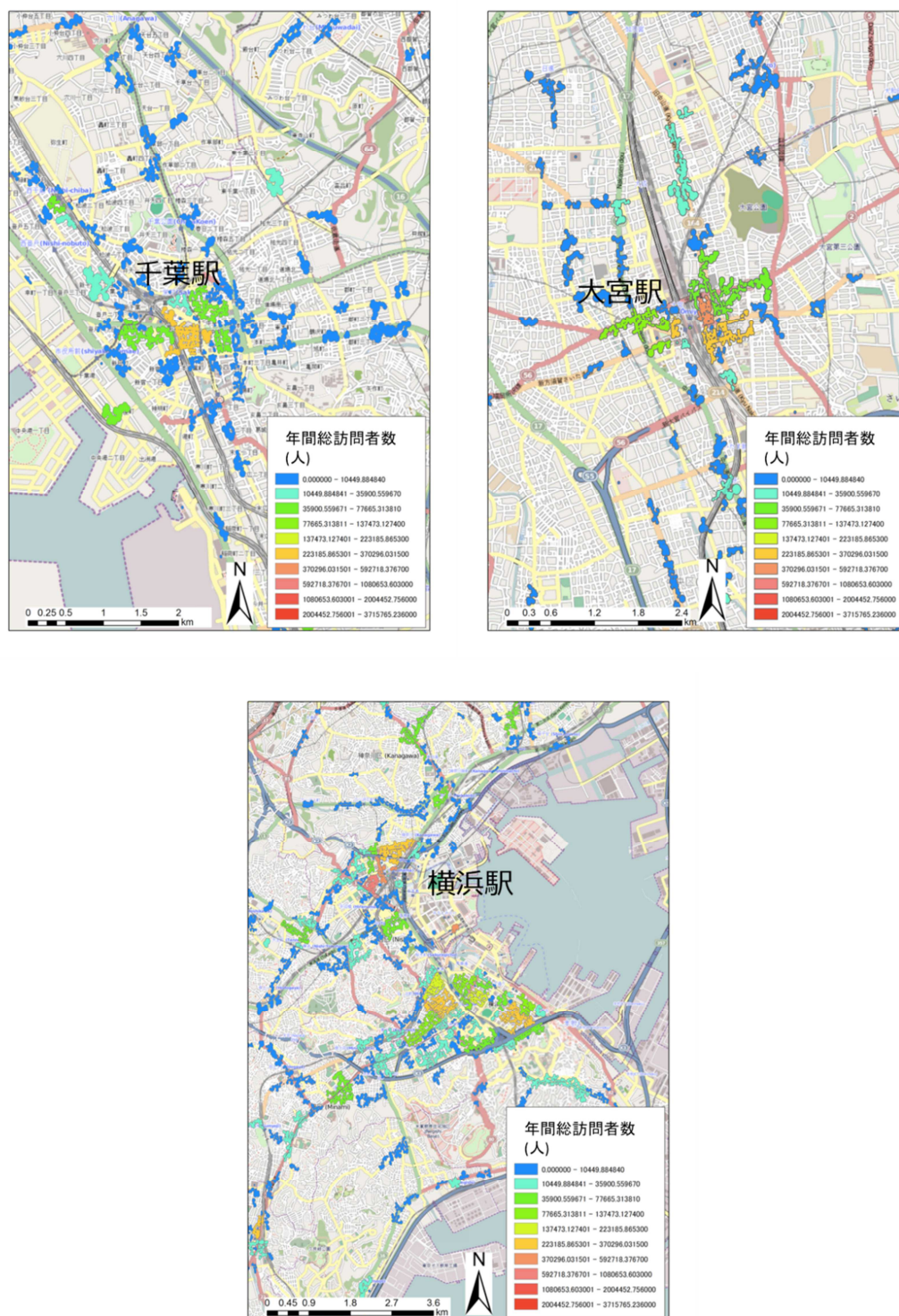


図 4.7 各主要都市における年間総訪問者数
(左上図：千葉駅周辺，右上図：大宮駅周辺，下図：横浜駅周辺)

表 4.4 各商業地域の年間総訪問者数
(上位 20 位を抜粋)

順位	商業地域	年間総訪問者数		店舗数	店舗当たり 総訪問者数
		拡大前	拡大後		
1	渋谷駅西 井の頭通り沿	14,998	3,715,765	770	4,826
2	町田駅東	7,238	2,004,453	1,130	1,774
3	新宿駅東	7,097	1,746,055	484	3,608
4	原宿駅東	6,200	1,526,740	506	3,017
5	新宿駅東 靖国通り北1	5,502	1,335,862	1,415	944
6	吉祥寺駅北	5,073	1,327,894	768	1,729
7	上野-湯島駅間	4,567	1,080,654	1,437	752
8	渋谷駅西 道玄坂南	4,285	1,059,860	813	1,304
9	池袋駅東1	3,961	957,879	292	3,280
10	横浜駅西2	3,292	936,339	428	2,188
11	新宿駅 東靖国通り南1	3,523	855,084	189	4,524
12	池袋駅西	3,491	842,394	730	1,154
13	新宿駅西	3,431	842,100	435	1,936
14	新宿駅 東靖国通り北2	3,352	839,001	289	2,903
15	自由が丘駅周辺	3,273	834,906	1,459	572
16	代官山駅周辺	3,434	828,899	1,539	539
17	有楽町駅周辺	3,278	783,818	77	10,179
18	渋谷駅西道玄坂北	3,159	770,747	722	1,068
19	柏駅東	3,029	758,631	684	1,109
20	池袋サンシャインシティ周辺	3,104	758,051	159	4,768

表 4.5 各商業地域の1店舗当たり年間総訪問者数
(上位20位を抜粋)

順位	商業地域	年間総訪問者数		店舗数	店舗当たり 総訪問者数
		拡大前	拡大後		
1	秋葉原駅西ダイビル周辺	2,397	592,718	42	14,112
2	二子玉川高島屋	635	165,517	15	11,034
3	有楽町駅周辺	3,278	783,818	77	10,179
4	京王堀之内北	446	120,748	13	9,288
5	綱島駅周辺	505	133,134	15	8,876
6	川口駅そごう	600	142,293	17	8,370
7	秋葉原駅東2	434	101,231	13	7,787
8	上北台駅北	264	75,909	10	7,591
9	ゲートシティ大崎	306	73,762	10	7,376
10	橋本駅北	376	94,837	13	7,295
11	ららぽーと横浜	1,793	518,979	84	6,178
12	みなとみらいクイーンズスクエア横浜	1,680	469,969	77	6,103
13	恵比寿ガーデンプレイス	622	155,411	26	5,977
14	東府中駅北	239	64,339	11	5,849
15	亀戸駅西	338	80,898	14	5,778
16	千葉駅そごう	229	56,436	10	5,644
17	池袋駅ルミネ	201	52,071	10	5,207
18	秋葉原駅東1	1,040	247,396	48	5,154
19	後楽園駅周辺	503	121,639	25	4,866
20	渋谷駅西井の頭通り沿	14,998	3,715,765	770	4,826

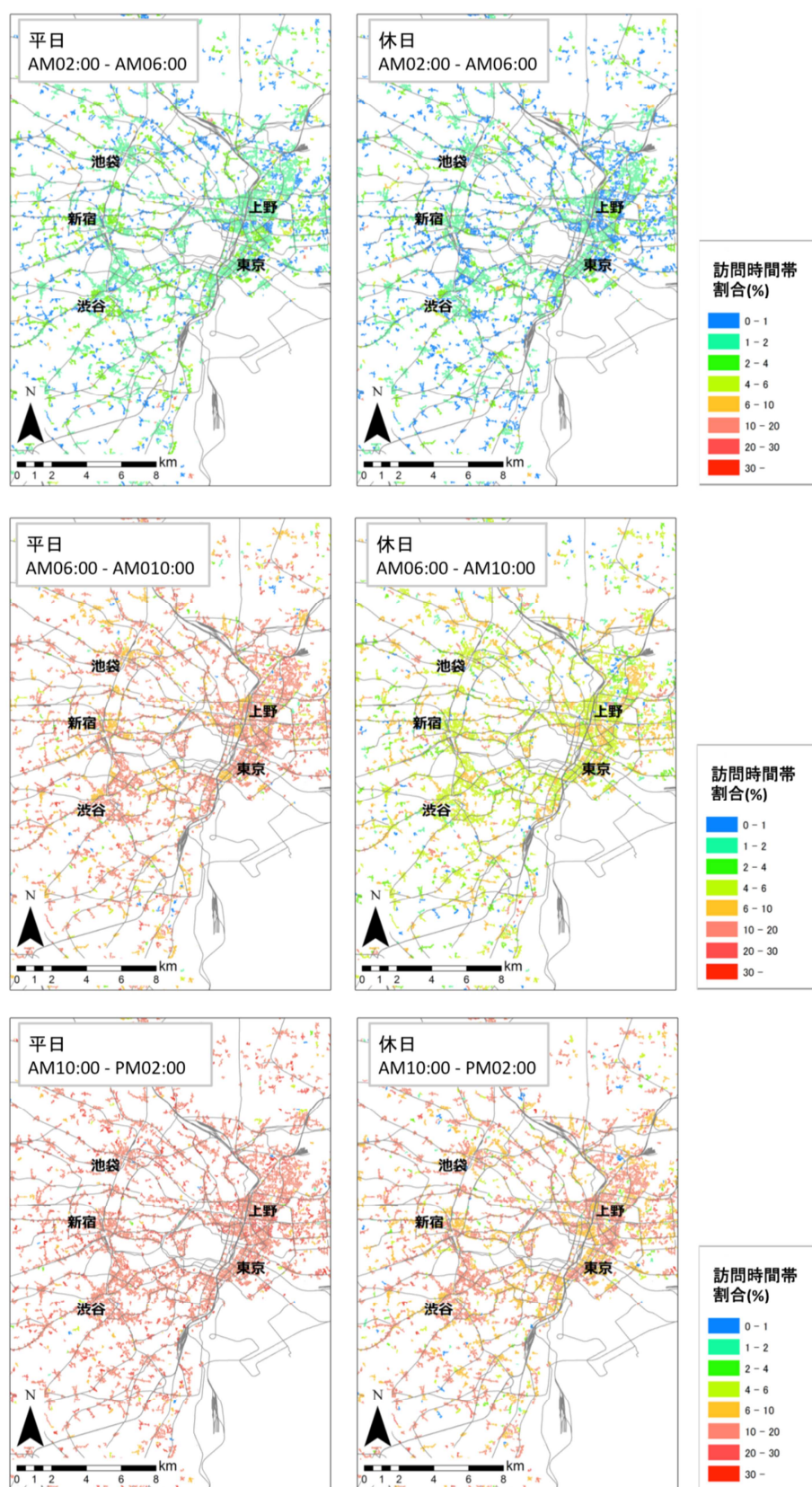


図 4.8 各時間帯の訪問割合 (AM 2:00 - PM 2:00)

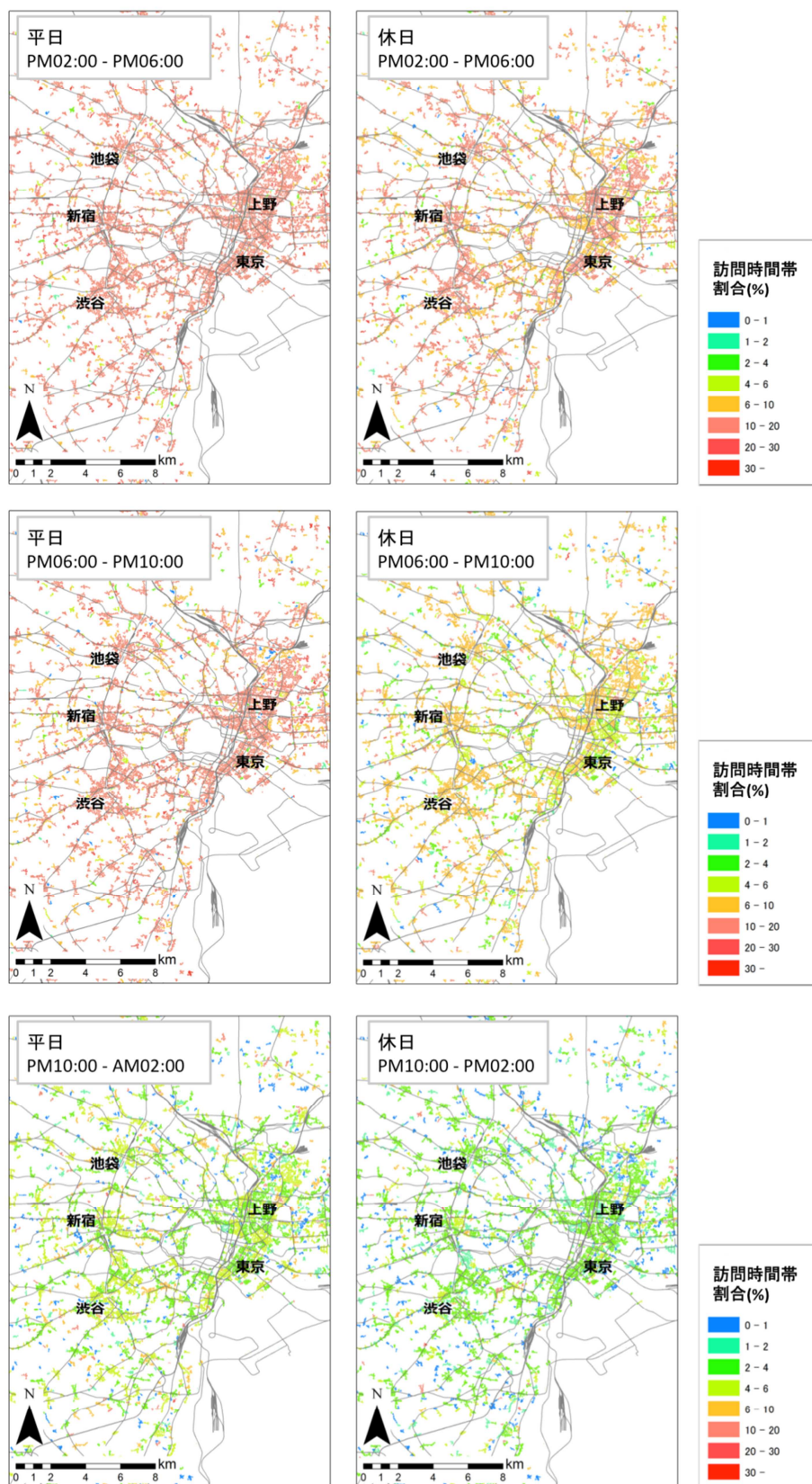


図 4.9 各時間帯の訪問割合 (PM 2:00 - AM 2:00)

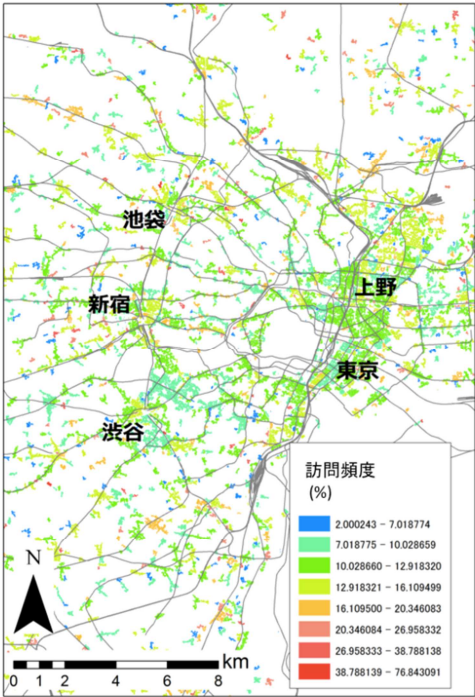


図 4.10 東京都心部における訪問頻度

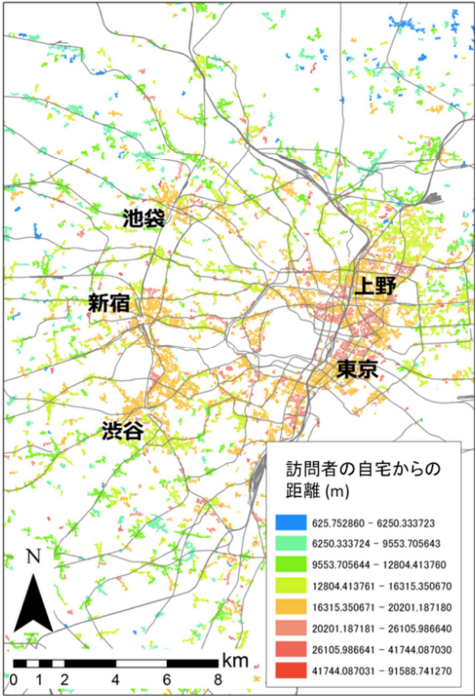


図 4.11 東京都心部における訪問者の自宅からの距離

4.3.3 商業統計との比較

本手法で計算された推定年間出費額は拡大係数を利用すれば、対象地域での総出費額となるので、もし精度よく計算できていれば対象項目の年間販売額と相関関係にあるはずである。そこで、本手法で計算した推定年間出費額が信頼できるかを検証するために、既存統計の年間販売額と比較した。

既存統計には経済産業省が実施する商業統計データを用いた。商業統計とは、卸売り・小売業に属する全国の事業所を対象に、従業員数や年間販売額、売り場面積などを調査したものである。なお、本研究で利用した商業集積統計と商業統計の都道府県ごとの商業地域数には高い相関があることが確認されている(Akiyama *et al.*, 2011)。

検証方法は以下の通りである。まず、本研究で得られた商業地域ごとの推定年間出費額のデータを3次メッシュ集計し、2004年の商業統計(3次メッシュ 産業別)と比較した。比較項目は商業統計の項目を再分類して食料品と衣服品の2項目を設定し(商業統計の項目再分類については参考資料3をみよ)、それぞれ食費と衣服費の推定出費額と比較した。なお対象範囲は東京都全域とした。

検証結果を図4.12に示す。食料品、衣服ともに推定年間出費額と区部で高い値を示すという空間的な分布は類似している。また、この結果を散布図にしたものを図4.13に示す。食費の推定年間出費額と商業統計での年間販売額は一定の相関関係があり決定係数は0.48程度であった。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散布図(図4.13)をみると、y軸付近で推定出費額の過大評価が目立ち、決定係数も0.22程度であった。つまり、実際はほとんど衣服が販売されていないのにも関わらず、そこで出費があったと推定してしまったということになる。その原因を以下で考察する。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、今回のような非常に簡便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。

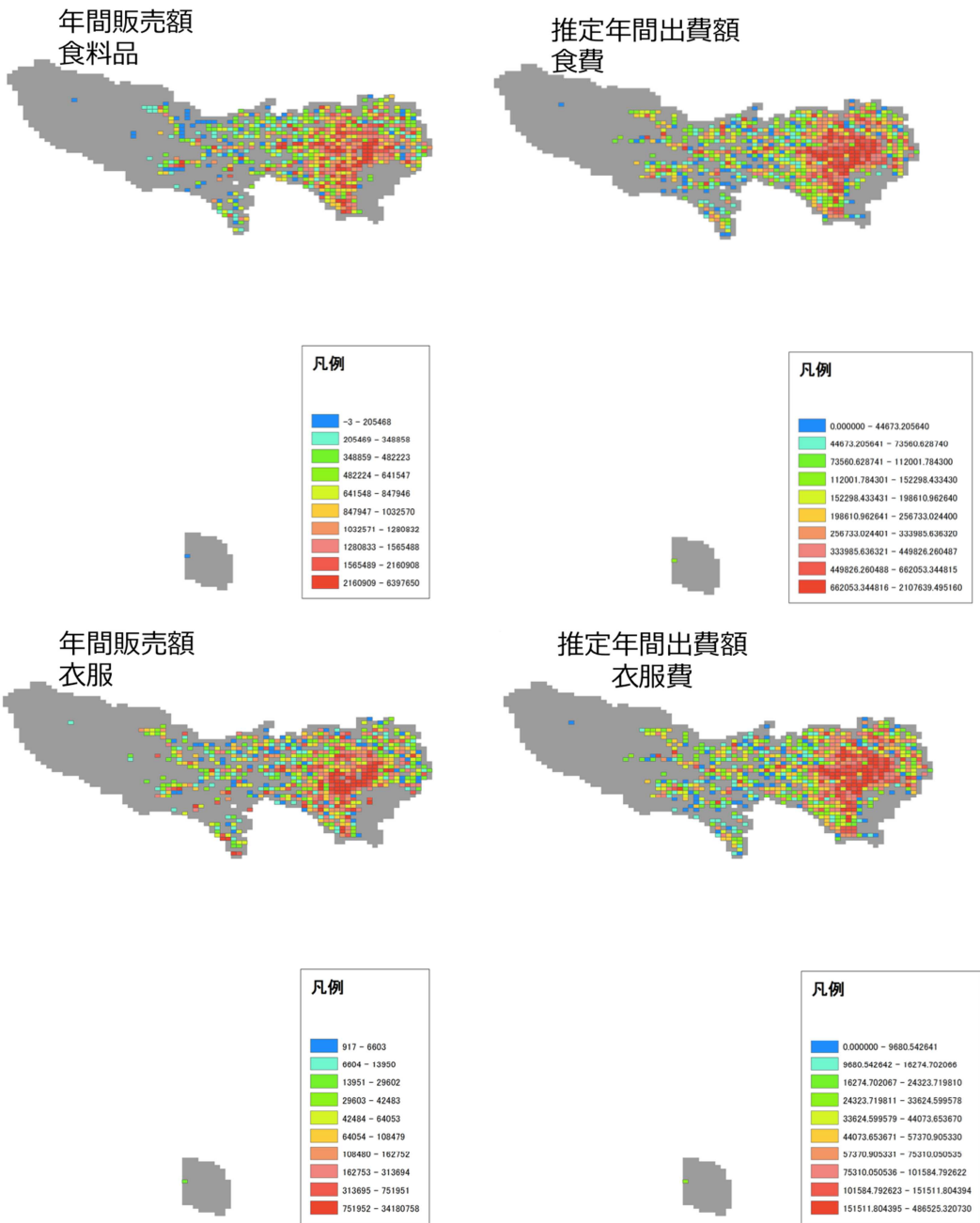


図 4.12 推定年間出費額と年間販売額の空間分布

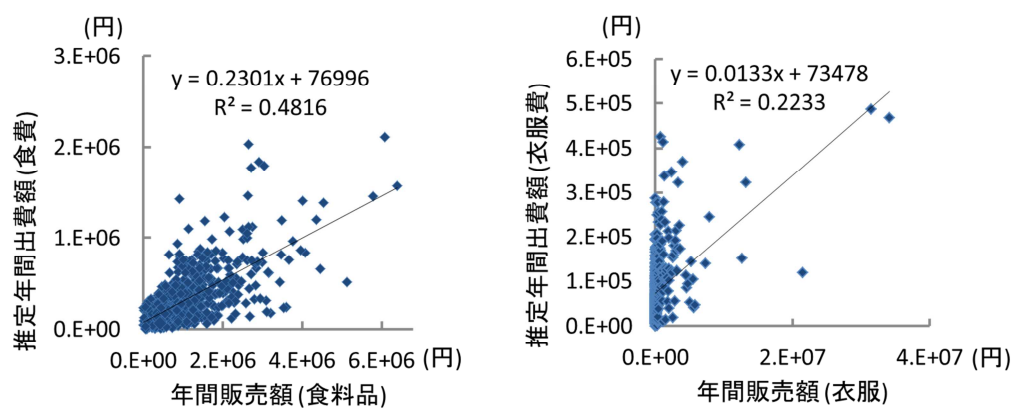


図 4.13 推定年間出費額と年間販売額の散布図
(左図：食料品，右図：衣服)

4.4 まとめ

本章では、第3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。また、GPSデータから得られた推定自宅位置と他統計データを組み合わせることによって、サンプルデータを実数に変換し、ある項目の年間出費額を推定する手法を提案した。

基礎データセットの作成手法は、第3章で述べた手法Iをそのまま踏襲したが、滞在点の抽出法だけ変更点がある。第4章では、大規模GPSデータの測位間隔は不規則で最低5分ということ considering、滞在の条件を20分以上、半径150mに変更した。さらに、商業地域訪問者の分析をより詳細に行えるよう、平均滞在時間、訪問頻度、訪問時間帯割合、平均世帯年間収入、平均世帯支出、推定年間出費額の項目も合わせて計算した。なお、推定年間出費額は、家計調査で得られる支出額を商業地域訪問の総滞在時間で比例配分することによって推定した。

大規模GPSデータを実数に変換するための拡大処理は、大規模GPSデータの夜間人口における対国勢調査抽出率を3次メッシュ単位で計算し、その値の逆数を各メッシュの拡大係数に設定することによって行なった。この手法を利用して拡大した推定人口と国勢調査の人口との決定係数は夜間人口で0.999、昼間人口で0.649であり、一定の相関が得られた。計算の結果、データセット全体の滞在クラスタ数は全体で3,528,790、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。

商業地域ごとに各項目を集計した結果、年間総訪問者数は主要都市で大きい値を示し、1番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。また、訪問者が多い商業地域には、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。1店舗当たりの集客力を把握するために1店舗当たりの総訪問者数を算出した結果、大型デパートや大型商業施設で高い値を示した。訪問者数以外の訪問時間帯や訪問頻度、自宅からの距離などの項目でも、商業地域ごとの訪問者特性を把握することが出来た。

本手法で計算した推定年間出費額が信頼できるかを検証するために、既存統計の年間販売額と比較した結果おおまかな空間分布パターンは類似した結果が得られた。また、両者の相関を確認したところ、食費の項目では決定係数は0.48と一定の相関が得られたものの、衣服費の項目では0.22程度と低い値が得られた。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、過大評価となった地域が多くなったと考えられる。

第5章 結論

5.1 本研究の成果

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。また、GPS データから得られた推定自宅位置と他統計データを組み合わせることによって、サンプルデータを実数に変換し、ある項目の年間出費額を推定する手法を提案した。

過去において GPS データから商業地域への訪問を抽出する研究例は少数あるものの、アルゴリズムが複雑で、大規模データへの応用が難しいという問題があった。本研究では、商業地域への訪問を滞在箇所と商業地域との位置関係、および滞在時間から判別するというシンプルなアルゴリズムを使用した。それでも一定の精度（最大で 60.56%の一致率）が得られることが分かった。

また、従来の商業地域訪問者の分析で使用されたデータセットは、対象エリアが小さく小規模商業地域を網羅しないという課題があったが、本研究によって小規模な商業地域を含む南関東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。データセット作成の際に利用したのは約 22 万人分の 1 年間にわたる GPS データであり、ここまでの大規模・長期にわたる GPS データを使って商業地域訪問分析を行なった例は過去にない。

さらに、本研究ではサンプルデータを実数に変換し、個人の年間出費額を推定する手法を提案し、実際に商業統計と比較した。その結果、食費の項目に関しては一定の精度が得られた。なお、GPS データから個人の年間出費額を推定する研究例は今までになく、マーケティングや商業地域における戦略策定の際の参考資料として役立つだろう。

5.2 本研究の課題

5.2.1 商業地域訪問の抽出手法における課題

商業地域の位置情報を得るために商業集積統計を使用した結果、大型ショッピングセンターやデパートなど、複数のテナントが一つの大型施設に入居するような商業地域の訪問が上手く抽出できない傾向がみられた。

図 5.1 は関東における郊外型大型ショッピングセンターであるイオンレイクタウンと商業集積ポリゴンの位置を示している。このように、商業集積統計は店舗の住所をジオコーディングした後に集積ポリゴンを作成しているため、複数テナントが一つの大型施設に入居するような場合、集積ポリゴンが現実の建物よりも小さくなるという特性がある。

本研究ではこのポリゴンを商業地域訪問の際の内外判定に使用するため、このような大型施設での訪問が過小になっている可能性が高い。実際に図 5.1 でみられる 3 つの商業地域における年間のべ総訪問者数を計算すると約 107 万人 (1,071,729 人) であったが、この値はイオンレイクタウンの計画年間来場者数の 2500 万人と比較してもかなり過小になっている。

この課題に対しては、大型施設の商業地域ポリゴンを実際の建物ポリゴンに変更することなどが対策として考えられる。



図 5.1 イオンレイクタウンおよび商業集積統計データの位置関係
(黄色点は商業集積データのポリゴンを示す)

また、商業地域ラベルを付与する際の手法についても課題はある。本研究では商業地域訪問ラベルを付与する際、鉄道利用時の待ち時間を考慮して、滞在時間の下限しきい値を 10 分以内と設定した。しかしながら、行動日記データによると 10 分以内の買い物は全体の約 4 割を占めるため、商業地域への訪問抽出が過小になっている可能性が高い。解決策としては、鉄道駅の位置情報を参考に鉄道待ち時間を先に抽出したあとに、商業地域訪問の抽出を行なうことなどが考えられる。

本研究では大規模 GPS データでの計算に手法Ⅰを適用したが、第 3 章における検証の結果、手法Ⅰの精度はあまり高くない。計算時間の都合上、今回は手法Ⅱを適用しなかったが、計算時間短縮の工夫を行うことによって手法Ⅱを適用する方向性は十分にありうる。

5.2.2 基礎データセット作成手法における課題

本手法で計算した推定年間出費額を既存統計の年間販売額と比較した結果、おおまかな空間分布パターンは類似していたが、特に衣服費の項目では決定係数が 0.22 程度と低い値であった。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単純なもので、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、過大評価となった地域が多くなったと考えられる。解決策としては、訪問先の商業地域の構成業種をもとにして配分に重み付けをすることなどが考えられる。

5.3 本研究の展望

5.3.1 大規模 GPS データの属性の偏りを考慮に入れた拡大処理

本研究では、大規模 GPS データを実数に変換するための拡大処理として大規模 GPS データの夜間人口における対国勢調査抽出率をもとにして、拡大係数を設定した。しかしながら、大規模 GPS データの基となるサービス利用者は、その属性に偏りがあることがわかっている (参考資料 1)。そこで、GPS データから利用者属性をある程度推定することが出来れば、属性の偏りを考慮に入れた拡大処理方法が開発できると考えられる。

5.3.2 他地域への応用

本研究は、南関東エリアだけを対象地域としたが、本研究で使用した商業集積統計データおよび大規模 GPS データは日本全国分存在する。本研究を他地域に応用することで、小規模な商業地域を含む商業地域訪問者の基礎データセットを全国分作成することができる。これにより商業地域訪問者の地域特性なども明らかにできるだろう。また、ある特定の商業地域だけに焦点を絞って、その商業地域訪問者の特性をさらに細かく分析するという発展もありうる。

5.3.3 買い物行動モデルの作成

本研究では GPS データと家計調査のデータを統合して、ある項目の個人の年間消費額を推定した。この手法をさらに発展させれば、個人の移動データから買い物行動モデルを作成することが出来ると考えられる。買い物行動モデルが作成できれば、次に行なう買い物の予測や買い物行動支援アプリケーションの開発などの応用が期待できる。

参考文献

- 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）.
- 秋山祐樹・上山智士・Horanont Teerayut・仙石裕明・柴崎亮介（2012）：大規模移動データを用いた商業地域における来訪者の特性分析，第 21 回地理情報システム学会講演論文集（CD-ROM, F-7-3）.
- 石垣司・竹中毅・本村陽一（2011）：日常購買行動に関する大規模データの融合による顧客行動予測システム 実サービス支援のためのカテゴリマイニング技術，人工知能学会論文誌，26（6），670-681.
- 大岩優佳理・山田哲也・三阪朋彦・兼田敏之（2005）：回遊行動からみた商店街複合地区の動態分析，日本建築学会技術報告集，22，469-474.
- 小沢佳奈（2006）：顧客 ID 付 POS データによる顧客属性別レコメンデーション提案：あるレコード店のケーススタディ，大阪府立大學経済研究，2006，52（3），99-118.
- 大藪勇輝・寺田雅之・山口高康・岩澤俊弥・萩原淳一郎・小泉大輔（2012）：モバイル空間統計の信頼性評価，NTT DOCOMO テクニカル・ジャーナル，20（3），17-23.
- 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.
- 木下斉（2009）：まちづくりの「経営力」養成講座，学陽書房.
- 京阪神都市圏交通計画協議会（2000）：人の動きからみる京阪神都市圏のいま，
http://www.kkr.mlit.go.jp/plan/pt/research_pt/h12/pt_h12.pdf（最終アクセス日：2013 年 1 月 26 日）.
- 島崎康信・関本義秀・柴崎亮介・秋山祐樹（2009）：人の流れによる時間帯別人口と店舗数との相関関係についての研究-パーソントリップ調査の時空間内挿データと国勢調査データとの比較分析-，都市計画学会都市計画論文集，44（3），781-786.
- 鈴木英之・関本義秀（2012）：新設されたショッピングセンターにおける収益の予実乖離要因-人の流れデータを用いて-，地理情報システム学会講演論文集（CD-ROM, F-6-5）.
- 関庸一（2003）：ID 付き POS データからの顧客行動パタンの抽出，オペレーションズ・リサーチ，48(2)，75-82.
- 高松瑞代・田口東・服部優奈・太田雅文・末松孝司（2013）：PASMO データを用いた鉄道

- 利用者の購買行動分析, オペレーションズ・リサーチ, 58, 37-46.
- 寺田雅之・永田智大・小林基成 (2012): モバイル空間統計における人口推計技術, NTT DOCOMO テクニカル・ジャーナル, 20 (3), 11-16.
- 千葉県柏市 (2012): 柏市商業実態調査,
http://www.city.kashiwa.lg.jp/soshiki/090700/p011578_d/fil/Kashiwacity_syougyouchous.pdf
 (最終アクセス日: 2013 年 1 月 14 日).
- 中小企業庁 (2010): 平成 21 年度商店街実態調査報告書,
<http://www.chusho.meti.go.jp/shogyo/shogyo/2010/download/100331SJCS.pdf> (最終アクセス日: 2013 年 1 月 7 日).
- 東京都文京区 (2012): 文京区商店街活性化検討委員会報告書,
<http://www.city.bunkyo.lg.jp/var/rev0/0042/1898/syoutengaikasseika-kentouinnkaihoukokusyo.pdf> (最終アクセス日: 2012 年 8 月 30 日).
- 西野正彬・瀬古俊一・青木政勝・山田智広・武藤伸洋・阿部匡伸 (2008): 滞在地遷移情報からの行動パターン抽出方式の検討, 情報処理学会研究報告, 57-64.
- 羽藤英二・中西雅一・寺谷寛紀・柏谷増男 (2002): 都市内回遊行動評価のための空間データマイニング, 土木計画学研究・講演集, 26.
- 藤田朗・半明照三・山田雅夫・大内浩・三宅理一 (2003): GPS 携帯電話を用いた回遊行動の調査分析 小田原市中心市街地を事例として, 日本建築学会学術講演梗概集, 855-856.
- 梶井昌邦・斎藤参郎 (2005): 決定木分析による都市型アミューズメント施設の来訪者特性評価, 地域学研究, 35, 199-214.
- 松隈久昭 (2006): 中心部商店街の消費者満足度分析, 大分大学経済論集, 57 (6), 1-19.
- 宮城県: 仙台都市圏パーソントリップ調査/鉄道待ち時間調査概要-宮城県公式ウェブサイト,
<http://www.pref.miyagi.lg.jp/site/pt/pt-sakutei-h14s-h14soukou05-01.html>, (最終アクセス日: 2013 年 1 月 26 日).
- 安岡寛道・森田哲明 (2012): ID-POS 分析による戦略的マーケティング会員 ID をキーに付加価値の高いサービスを提供ー, IT ソリューションフロンティア, 29 (4), 12-15.
- 山田直治・磯田佳徳・南正輝・森川博之 (2010): GPS 搭載携帯電話を用いた移動経路履歴に基づく訪問地・経由地予測システム, 情報処理学会研究報告, 110 (130), 1234-12241.
- 渡辺亮・北村裕人・星野直人・関庸一 (2005): 買い回りタイプによる顧客購買行動の理解, オペレーションズ・リサーチ, 50 (9), 644-653.
- Akiyama, Y., Sengoku, H., Hiroyuki, T. and Shibasaki, R., 2011. Development of Commercial Accumulation Polygon Data Through Japan Based on the Digital Classified Telephone Directory, CUPUM2011, F-TC-3(1).

- Ashbrook, D. and Starner, T., 2003. Using GPS to learn significant locations and predict movement across multiple users. *Personal and Ubiquitous Computing*, 7(5), 275-286.
- Ester, M., Kriegel, H.P., Sander, J., and Xu, X., 1996. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, *Proceedings of the 2nd International Conference on Knowledge Discovery and Data mining '96*, 226-231.
- Jiang, S., Ferreira, J., and Gonzalez, M.C., 2012. Discovering Urban Spatial-Temporal Structure from Human Activity Patterns, *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*, 95-102.
- Liao, L., Patterson D.J., Fox, D., and Kautz, H., 2005. Building Personal Maps from GPS Data, In *proceedings of IJCAI MOO05*, 249-265.
- 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.
- Schüssler, N., and Axhausen K.W., 2008. Identifying trips and activities and their characteristic from GPS raw data without further information, *8th International Conference on Survey Methods in Transport*.
- Ye, Y., Zheng, Y., Chen, Y., Feng, J. and Xie, X., 2009. Mining Individual Life Pattern Based on Location History, *MDM '09 Proceedings of the 2009 Tenth International Conference on Mobile Data Management: Systems, Services and Middleware*, 1- 10.
- Zheng, Y., Zhang, L., Xie, X. and Ma, W. 2009. Mining Interesting Locations and Travel Sequences from GPS Trajectories, *WWW '09 Proceedings of the 18th International Conference on World Wide Web*, 791-800.

謝辞

多くの方々のご指導ご協力があり、本修士論文を仕上げることができました。皆様へ感謝の気持ちを込めて、ここにお礼の言葉を述べさせていただきます。

指導教員である柴崎亮介教授には、研究の指導はもちろんのこと、研究を進める上での心構えを折に触れて教えて頂きました。また、海外を含む学会への参加や研究会の参加、勉強会の呼びかけなど、二年に渡る研究生活において、充実した環境を提供して頂きました。心より感謝いたします。

副指導教員としてご指導いただいた浅見泰司教授には、研究を深化させるべき方向性について、またその方法について指導して頂きました。いつもとは違う視点からの意見を頂きまして、大変勉強になりました。ここに厚くお礼申し上げます。

本研究は株式会社ゼンリンのデータセットを利用しており、田中様には研究過程で様々な助言をいただき、大いにお世話になりました。さらに、大規模 GPS データは株式会社ゼンリンデータコムからご提供いただきました。貴重なデータを提供して頂き大変感謝しております。

柴崎研究室特任研究員の金杉洋氏からは GPS データと行動日記データをご提供頂きました。おかげで本研究の手法を比較検証することができ、信頼性を評価することができました。また、おなじく特任研究員の Teerayut Horanont 氏には、大規模 GPS データの整形やダウンロード環境などのシステムを提供して頂きました。また大規模 GPS データのハンドリングや研究テーマについても相談に乗っていただきました。誠に感謝しております。同研究室の博士課程 2 年の Apichon Witayangkurn 氏には大規模 GPS データを利用したデータ処理の際に、Hadoop を利用した環境を提供していただきました。自分一人の力ではこのような大量なデータを扱うことは出来ませんでした。本当にありがとうございました。

同研究室特任研究員の秋山祐樹氏と博士課程 2 年の仙石裕明氏には、研究テーマを決めるところから何度も相談に乗っていただきました。親身になって相談にのっていただいたことで、安心して自分のやりたい研究を見つけることが出来、このように修士論文を完成させることができました。

ここに名前を挙げた方々以外にも、研究室内外を問わず、多くの方々のお力添えによって論文を完成させることができました。この場をお借りして心よりお礼申し上げます。

参考資料

参考資料 1 混雑統計 (R) データ利用者属性

混雑統計 (R) データは、あるサービス利用者の承諾を得て入手した GPS データを基にしているため、利用者の属性の偏りの影響を受ける。その程度を把握するために日経リサーチを通じて 5 万人を対象に 2012 年に行なったアンケート結果を用いて、性別人口、年代別人口、地域人口の 3 つの項目について利用者属性の割合を調べた。また、利用者属性の割合を 2010 年の国勢調査の結果と比較した。

はじめに利用者の性別人口は表 1 のようになった。ここでの対象サービス補正利用者数とは、アンケートが行われた集団にそもそも偏りがあることを考慮して、その影響を除いた数値である。対象サービス利用者数、アンケート被験者数割合、属性 i の国勢調査の割合、をそれぞれ n_i , r_i , R_i とすると、対象サービス補正利用者数 n_i^* は、

$$n_i^* = n_i \frac{R_i}{r_i}$$

と計算した。アンケートの被験者のうち対象サービスを利用している人の割合は、全体で 2.71%、男性間で 3.11%、また女性間で 2.28% という結果になった。また、対象サービス補正利用者と国勢調査の男女比をそれぞれ比較すると、対象サービス利用者は男性利用者が多い傾向にあることがわかった。アンケート被験者の男女比がそもそも男性に偏っている (51.42%) ことに対する影響を除外しても、男性に偏りがちであるという結果が得られた。

次に年代別人口の比率は図 1 のようになった。アンケート被験者は最小でも 15 歳、最大でも 79 歳であることから、対象サービスの利用者の年代別割合において 0 歳代と、80 代以上の階級が 0 となっている。また、対象サービス利用者の年代別割合は、国勢調査のものと比較して、20 代、40 代～70 代で過大、10 代で過小となった。

最後に、地域人口の比率は図 2 のようになった。対象サービス利用者の地域別割合は国勢調査のものと比較して、特に関東地方で大きかった。反対に北海道や近畿地方で特に小さかった。

今までみたように、大規模 GPS データの基となるサービスを利用している人の属性は、国勢調査と比較して、男性・20 代および 40 代～70 代・関東地方に集中している可能性があり注意が必要である。さらに携帯電話の普及率の低い 0 歳代、80 歳以上の年代の行動をほとんど把握できない可能性があり、必ずしもすべての人の行動を把握できる訳ではないことが分かった。

しかしながら、本研究では関東地方の一都三県（埼玉県、千葉県、東京都、神奈川県）だけを対象にするため、地方間の偏りは修正しない。また、GPS データから性別や年代を推定することは非常に難しく現時点では他の研究例もないため、その偏りを補正することは本研究の対象外とする。

表 1 対象サービス利用者の性别人口および男女比

	国勢調査人口	アンケート 被験者数	対象サービス 利用者数	対象サービス 補正利用者数
男性	62,327,737 (48.67)	25,710 (51.42)	801 (59.07)	758.19 (56.39)
女性	65,729,615 (51.33)	24,290 (48.58)	555 (40.93)	586.40 (43.61)
合計	128,057,352 (100)	50,000 (100)	1,356 (100)	1,345 (100)

括弧内は割合 (%)

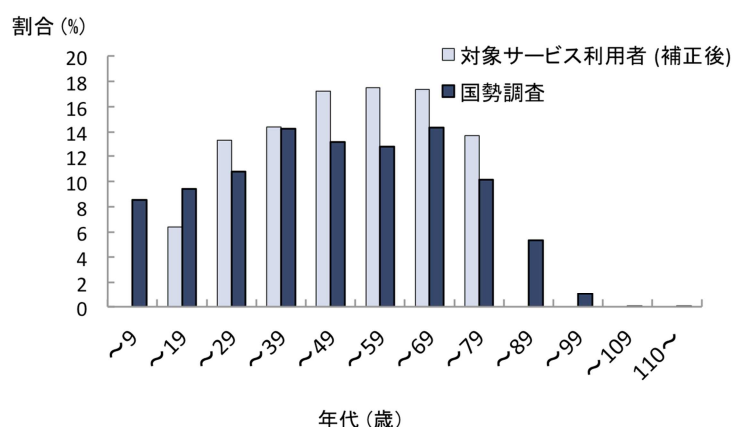


図 1 対象サービス利用者の年代別人口の割合

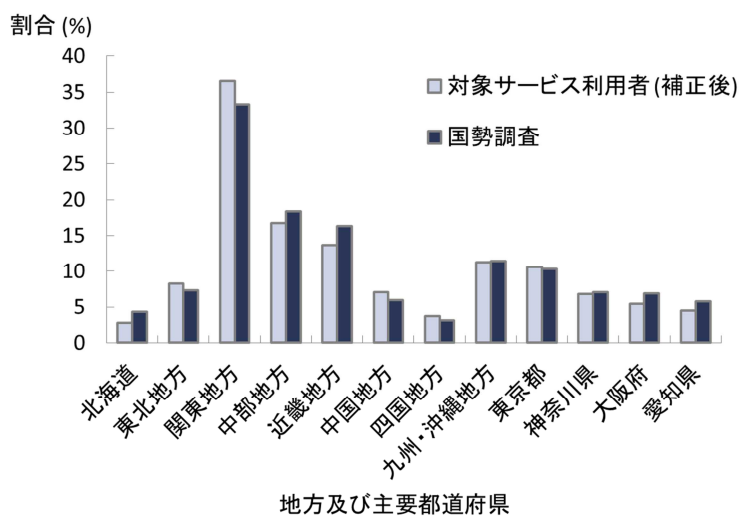


図 2 対象サービスの地域人口の割合

参考資料 2 家計調査における項目再分類の詳細

2011 年の家計調査における用途分類を以下のように再分類し，食料費，外食費，衣服費の 3 項目を設定した．

食料費：外食・賄い費を除く食料の項目（穀物，魚介類，肉類，乳卵類，野菜・海草，果物，油脂・調味料，菓子類，調理食品，飲料，酒類）の合計値

外食費：一般外食の項目

衣服費：被服及び履き物の項目

参考資料 3 商業統計における項目再分類の詳細

平成 16 年商業統計メッシュデータ（1 km メッシュ産業別表）から，以下の項目を足し合わせて食料，衣服の 2 項目に再分類した．

食料： 飲食料品，各種食料品，酒，食肉，鮮魚，野菜・果物，菓子・パン，米穀類，その他の飲食料品

衣服： 織物・衣服・身回品，呉服・服地・寝具，男子服，婦人・子供服，靴・履物，他織物衣服身回品