

# ランダム・フォレストによる回遊行動データの移動滞在判別手法の検討

熊本大学工学部 学生会員 ○瀧澤憲駿

熊本大学大学院 学生会員 井村祥太郎  
熊本大学 正会員 円山琢也

## 1. はじめに

現在、交通分析において、ある地域の公共交通機関利用状況や回遊行動、道路利用実態などの把握のため、PP調査（プローブ・パーソン調査）が盛んに行われている。PP調査では、位置情報や加速度を用いた利用者の活動・交通行動の把握が行われ、政策提言や調査法の改良に役立っている。しかし、そこで得られるデータは膨大であり、その分析には多くの時間と労力が必要となる。また、分析において「どの地点で多くの人が滞在しているのか」等の把握が必要となるが、従来の研究では多くが手動でそれらを判別しており、分析の効率を上げることが課題とされていた。そこで、本研究ではその移動滞在判別を自動かつ高速で判別する手法を検討し、より効率の良いPP調査の研究に役立っていることを目的とする。

## 2. 移動滞在判別手法

### 2.1 移動滞在判別

移動滞在判別とは、GPSで得られた位置情報（緯度、経度）や加速度などをもとに、ある地点において調査対象がそこで滞在もしくは移動のどちらの状態なのかを判別することである。本研究では、「移動」「滞在」に加え、より詳細に回遊行動を把握するため、屋内での回遊行動を示す「店舗内回遊」の合計3項目にて判別を行い、回遊行動を分析する。

### 2.2 ランダム・フォレスト

本研究において移動および滞在の自動判別を実現するためにR言語で「ランダム・フォレスト（randomForest）」を実行する。ランダム・フォレストとは、集団学習により高精度の分類・予測を実現するアルゴリズムであり、Leo Breimanが2001年に提案した手法である。特徴として、説明変数が膨大でも効率的に作動する点、他の機械学習に比べ一般的に精度が高く、計算速度が速い点などが挙げられる。

本研究では、まず複数のデータを教師データとしてランダム・フォレストにて学習させ、それをもとに判別したいデータの移動滞在判別を行う。

## 3. 調査データによる分析

### 3.1 分析について

今回の研究に用いるデータとして、2014年12月に表1の内容で調査を行い、位置情報等のデータ取得を行った。

表1 追加調査概要

概要	内容
日時	2014年12月3日(水)14:20~16:20
場所	熊本市中心市街地
調査参加者	熊本大学学生8名
使用端末	調査参加者のスマホ (スマートフォン4台、Nexus 7(23台))
記録状態	各班2名 (記録1名、端末を持って移動1名)

調査時に際し、2名でグループを構成し、ひとりがスマートフォンおよびタブレットを所持し、もうひとりが記録用紙への経路や時刻、観測状態の記録を行った。観測状態とは、その時刻・位置においてどのような状態で移動もしくは滞在をしたのかを表すもので、表2に記述する8項目で観測状態を記録した。続いて、分析を行う際には、移動滞在判別のために新たに判別項目を定めた（表2参照）。屋内での回遊は「店舗内回遊」としてまとめ、その場に立ち止まる観測状態は「滞在」としてまとめ、計3項目に分類した。

表2 観測状態および分析時の判別項目名

観測状態	分析時の判別項目名
通常の移動	移動
店舗内回遊	店舗内回遊
階段	店舗内回遊
エスカレータ	店舗内回遊
エレベーター	店舗内回遊
信号停止	滞在
休憩 座る	滞在
休憩 立ち止まる	滞在

この調査で得られたデータを用いて移動滞在判別を行うが、データ数が膨大になるほど多くの計算時間がかかる。今回は高速で移動滞在判別を実現するため、追加調査で得られたデータから無作為に一部を抽出し、それらを複数のデータから採取して統合することで判別時間を短縮し、教師データを作成することとする。なお、教師データとは、移動滞在判別を行う際に、判別の判断基準となるデータのことであり、この教師データを基にして判別したいデータの移動滞在判別を行う。

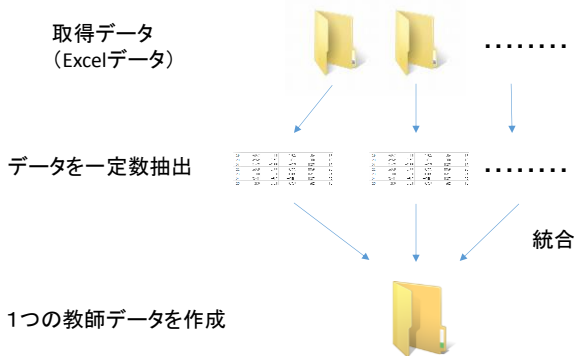


図1 教師データの作成

次に、教師データにおけるデータ数での精度の違いを把握するため、追加調査で得られた複数のデータから、それぞれ無作為に3,000行もしくは6,000行を抽出し、それを8つのデータから取り出した。つまり、ここでは教師データとしてデータ数が24,000行と48,000行の際の精度の違いを把握する。

### 3.2 移動滞在判別結果

以下表3および表4に判別結果を記述する。表3から順に、3,000行×8データ(計24,000データ、以降データAとする)、6,000行×8(計48,000データ、以降データB)、9,000行×8(計72,000データ、以降データC)の際の判別結果である。各表において、対角線上の赤字が正しく判別されたデータ数であり、それ以外の黒字は、列方向がそれぞれの項目に誤って判別されたデータ数を示す。

3項目を合わせた全体の判別率はいずれも約70%であった。「移動」の判別率が比較的悪く、「滞在」の判別率が比較的高かった。データ数が変化してもそれぞれの判別率に大きな差はなく、ある一定数のデータがあれば、教師データとして同様の精度が得られると考えられる。

表3 移動滞在判別結果(データA)

3,000行×8	移動	店舗内回遊	滞在	判別率(%)
移動	3191	993	600	53.3
店舗内回遊	1376	5927	4034	67.9
滞在	1412	1813	16272	77.8
合計	5979	8733	20906	71.3

表4 移動滞在判別結果(データB)

6,000行×8	移動	店舗内回遊	滞在	判別率(%)
移動	2848	1159	1113	47.6
店舗内回遊	1504	6122	3346	70.1
滞在	1627	1452	16447	78.7
合計	5979	8733	20906	70.6

表5 移動滞在判別結果(データC)

9,000行×8	移動	店舗内回遊	滞在	判別率(%)
移動	2628	1692	857	44.0
店舗内回遊	1379	5531	2516	63.3
滞在	1972	1510	17533	83.9
合計	5979	8733	20906	72.1

判別において、計算に要する判別時間は、データAでおおよそ30秒、データBでおおよそ1分弱、データCでおおよそ4分であった。データ数が増えるほど指数関数的に計算時間が増えると考えられる。いずれも手動と比較して、判別時間の短縮が期待される。

## 4. おわりに

今回用いた移動滞在判別手法では、約70%の精度で高速判別を実現することができた。教師データは、データ数を増やしてもほぼ同様の精度が得られることが分かった。また、計算時間において、手動による計算よりも大幅に時間短縮できることが確認された。

今後の展開として、移動滞在判別結果をGIS等の活用により可視化し、さらなる精度の検証を行う。また、平成25年度に実施された「くまもとまち歩き調査」での取得データで移動滞在判別等を行い、まちなかでの回遊特性の把握を行うことでより実践的な応用を行う。

## 【参考文献】

- 1) 井村祥太郎, 佐藤貴大, 円山琢也: スマホアプリ型回遊調査のデータ特性とSVMによる移動滞在判別の基礎分析, 土木計画学研究・講演集, Vol.50,2014