

GPS データを用いた商業集積地来訪者の行動パターン抽出方法の検討

羽田野真由美・上山智士・秋山祐樹・Teerayut Horanont・柴崎亮介

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

Mayumi HADANO, Satoshi UHEYAMA, Yuki Akiyama, Teerayut HORANONT
and Ryosuke SHIBASAKI

Abstract: The diffusion of GPS devices enables people to log their location history. In this study, we propose a method for extracting activity patterns from location data obtained by GPS enabled mobile phone data. Especially, we focus on activity patterns of people who visit a commercial area in Japan. First we translated GPS data to a sequence of staying point clusters in time series, and then we applied time-based clustering method and obtained activity patterns. Finally, we obtained activity patterns by selecting feature patterns from those mined one. This study enables to analyze visitors' activity patterns at arbitrary commercial areas and will be expected to apply the service depending on the situation.

Keywords: 行動パターン (activity patterns), 商業地域 (commercial area), 空間データマイニング (spatial data mining), GPS

1. はじめに

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。そこで商業地域間での競争を勝ち抜き、他地域との差別化をはかる上でも、来評者の分析を行なうことは重要である。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が注目を浴びている。

商業地域来訪者の分析は、最近では羽藤ら (2002) が松山市で、梶井ら (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 (文京区, 2012) があるものの、これらのような現地調査は多大な労力がかかり、広域にわたって定期的に行うことは困難である。

一方で、近年 GPS 搭載型の携帯電話の普及によって、人の移動履歴に関する時空間データが蓄積されている。このようなデータを用いて、利用者の生活パターンの把握 (Ye *et al.*, 2009) や抽出方法の検討 (西野ほか, 2008) などを行った例がある。

そこで本研究では、GPS 搭載型の携帯電話データをもとにして、ある商業地域における来訪者の行動パターン抽出法を検討することを目的とする。ここでの行動パターンとは、自宅を出発してからどのような順番で滞在地を回るかという滞在地の遷移情報のことを指す。

羽田野真由美 〒277-8569 千葉県柏市柏の葉 5-1-5

東京大学 (柏キャンパス) 総合研究棟 435 号室

Phone: 04-7136-4307

E-mail: hadano@csis.u-tokyo.ac.jp

2. 使用データ

2.1 GPS データ

本研究ではゼンリンデータコム社の「混雑統計データ」を用いた。これは携帯電話の 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 GPS データとする)。なお、このデータの測位間隔は最短でも 5 分である。

2.2 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り、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データで観察出来るデータである。本研究では、2011 年の東京都のデータを用いた。

3. 手法

まず膨大な GPS データから商業集積地訪問者を抜き出し、利用者が一定時間以上留まる滞留点を抽出した。さらに複数の滞留点をクラスタリングアルゴリズムによって分類した。最後に各クラスタをそれぞれ分析して居住地・勤務地・商業地域のラベルをつけ、滞在地の遷移情報を得た。これらの工程の詳細を以下で述べる。

3.1 商業集積地訪問者の抽出

商業集積地来訪者の抽出は、GPS データによる位置座標が、ある商業集積ポリゴン内を一定時間 (15 分以上 4 時間未満) 留まるという条件によって行った。同一の商店街内での回遊時間は長くても 4 時間以内 (大岩ほか, 2005) であることから、それ以上の滞留は自宅や勤務地であると考え、上限の閾値を設けた。

3.2 滞留点の抽出

GPS データは測位間隔が一定ではないため、まず前処理として緯度・経度を 1 分間隔で線形補間した。その後、各位置座標から利用者の滞留点の

抽出を行なった。ここでの滞留とは、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、
滞留点はその代表点のことをいう。本研究では滞
留の条件として 10 分以上、半径 100m を設定し、
滞留点を滞留期間の点列からなる最大矩形の中
央点とした。なお、各滞留点は緯度・経度情報
の他、到着時間、出発時間という時間情報も保持
する。

3.3 クラスタリング

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留点は複数存在
する。そこで、空間的に近接した異なる滞留点
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。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-1 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。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 150m, $MinPts$ を 2 と設定して計算した。

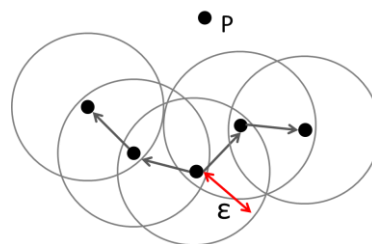


図-1 DBSCAN アルゴリズムの接続関係
(このとき点 P はノイズとして分類される)

3.4 居住地・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、滞留点の時間帯に応じて居住地・勤務地を、場所に応じて商業地域のラベルをつける。各クラスタの滞留点が一番多く時間帯 a (00:00-03:59) に属するクラスタを居住地とした。また、勤務地は平日の滞留点の中で一番多く時間帯 b (12:00-15:59) に属するクラスタとした。最後に、クラスタの滞留点のうち一点でも商業集積統計におけるポリゴンに含まれるものには商業地域のラベルをつける。

4. ケーススタディ

本手法を用いて実際に、ある商業地域における来訪者の行動パターンを抽出した。

4.1 商業集積地訪問者の抽出

訪問者を抽出する対象の商業集積ポリゴンは神保町駅南東に位置するもの (194 店舗を含む) を利用した。神保町駅付近は書店が集積していることで有名であるが、秋葉原をはじめとする周辺の再開発地区に押されて、近年衰退傾向だといわれている。このデータを用いた抽出によって東京都全域の 1 日分のデータである約 14 万人分の GPS データから 125 人が抽出できた。

4.2 滞留点の抽出

抽出された商業集積地訪問者の日頃の行動パターンを分析するために、1 か月分 (2011 年 6 月) の GPS データを利用した。また、GPS データ量が極端に少ない利用者を除外した (除外前 125 人, 除外後 86 人)。滞留点の抽出結果の例を図-2 に示す。全体では、1 日平均 14.4 個の滞留点が抽出された。



図 2 GPS データ (上)と抽出された滞留点 (下)

4.3 クラスタリング

ある利用者のクラスタリング結果の一部を図-3 にて示す。ここでは皇居を中心としてクラスタが A から D まで 4 つ確認できるが、クラスタ D はノイズとして検出されたものである。

DBSCAN によるクラスタリングの結果、1 か月間の平均クラスタ数は 11.8 となった。



図-3 ある利用者のクラスタリング結果 (一部)

4.4 滞在地遷移パターンの分析

各クラスタを居住地・勤務地・商業地域にラベル付けしたデータをもとに、ある一人の利用者の滞在地遷移情報を表-1 にまとめた。ここで滞在地の記号をそれぞれ H : 自宅, W : 勤務地, CA : 商業地域, A : その他を表し、その時間的な繋が

りを矢印で表した．つまり， $H \rightarrow W \rightarrow H$ という記号は，自宅から勤務地へ移動し，そのまま自宅へ帰るという1日の動きを示す．これをみると，利用者が商業地域で滞在するときのパターンとして勤務地から自宅へ帰る途中に寄るパターン ($H \rightarrow W \rightarrow CA \rightarrow H$, $H \rightarrow W \rightarrow CA \rightarrow H \rightarrow A \rightarrow H$) と，勤務地から商業地域へ移動して勤務地へと戻るパターン ($H \rightarrow W \rightarrow CA \rightarrow W \rightarrow H$) の2つが確認された．

表-1 ある利用者の一か月分の滞在地遷移
(H: 自宅, W: 勤務地, CA: 商業地域, A: その他)

滞在地遷移パターン	日数 (day)
$H \rightarrow W \rightarrow H$	13
$H \rightarrow A \rightarrow H$	4
H	2
$H \rightarrow W \rightarrow A \rightarrow H$	2
$H \rightarrow W \rightarrow A \rightarrow W \rightarrow H$	2
$H \rightarrow W \rightarrow CA \rightarrow H$	2
$H \rightarrow A \rightarrow A \rightarrow A \rightarrow H$	1
$H \rightarrow A \rightarrow A \rightarrow H$	1
$H \rightarrow A \rightarrow W \rightarrow H$	1
$H \rightarrow W \rightarrow CA \rightarrow H \rightarrow A \rightarrow H$	1
$H \rightarrow W \rightarrow CA \rightarrow W \rightarrow H$	1

5. おわりに

本研究では，GPS データをもとにして，商業地域における来訪者の行動パターンを抽出する手法を検討した．また実際にケーススタディとして利用者の1か月間分の滞在地遷移パターンを得た．今後は各パラメータの調整や，クラスタリングアルゴリズムの改良，クラスタリング結果の検証を行うことが課題である．

謝辞

本研究は株式会社ゼンリンとの共同研究によって実現した．GPS データは株式会社ゼンリンデータコムよりご提供頂いた．ここに記して謝意を表したい．

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).
- 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.
- 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.
- 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.
- Ester, M. and Kriegel, H.P. and 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*, 226-231.
- Y. Ye, Y. Zheng, Y. Chen, J. Feng, and X. Xie, 2009. Mining Individual Life Pattern Based on Location History, *Proc. Int'l. Conf. on Mobile Data Management Systems, Services and Middleware*, 1-10.