

デジタル電話帳データを用いた店舗・事業所の時系列データ構築と 分析手法の研究

水野弘規・秋山祐樹・柴崎亮介

Development and Analysis of Time-Series Geodata of Shops and Offices Using Digital Telephone Directory

Hiroki MIZUNO, Yuki AKIYAMA and Ryosuke SHIBASAKI

Abstract: This study develops a method of developing time-series dataset of shops and offices using digital telephone directory which allows us to monitor changes of shops and offices throughout Japan every year. First, shops and offices in digital telephone directories in different two years are integrated spatially based on their locations and names. Second, results of their time-series change are classified into continuation, change, emergence, and demolition to identify their names. Finally, time-series data of each shop and office in broad areas are realized to apply this method for mass telephone directories in multi years. This study developed time-series dataset of each shop and office throughout Tokyo prefecture between 2011 and 2012 using this method.

Keywords: 時系列データ(time-series geodata), デジタル電話帳(digital telephone directory), 都市(city), 店舗(shop), 事業所(office)

1. はじめに

都市を形作る建物は時間とともに立地や姿を変え都市の様子を変容させる。建物の機能はそこに入居する店舗・事業所によって決定される。その時系列的な変化を知ることは都市の過去を分析し、現在を評価し、未来を予測することにつながる。建物に入居する店舗・事業所は、人の活動パターン変化からも変化することから、時系列的な変化を明らかにすることは都市の様子を明らかにするという意味でも重要であると言える。

本研究では全国規模で毎年、店舗・事業所の情報が観察できるデジタル電話帳データを用いて、店舗・事業所の時系列的変化を明らかにする。す

で、関口ほか(2011)が青山などの地域で店舗の業種や位置などの時系列な変遷を明らかにすることに着手しているが、特定の地域に留まっており、変化を表す時間軸も5年ごとと、着手した地域は狭く時間軸のスパンも限られている。このことは全国的な店舗・事業所の時系列的な変化のデータベースとしては不十分である。

現在、全国規模の広域に渡って都市空間の店舗・事業所の時系列的な変遷を把握できるデータは殆ど整備されていない。本研究は全国規模で利用可能なデジタル電話帳データを用いて、広域な範囲を対象とした時系列変化データの構築を行う。なお、デジタル電話帳データは最短で2ヶ月おきに更新される為、より短いスパンでの店舗・事業所の時系列変化を追うこともできるが、今回は店舗・事業所の1年のスパンでの時系列的な変遷明らかにした。

水野弘規 〒277-8568 千葉県柏市柏の葉 5-1-5

東京大学大学院新領域創成科学研究科

Phone: 080-1167-8297

E-mail: h.mizuno@sis.u-tokyo.ac.jp

2. 手法の概略

2.1 概要

本研究では全国規模で毎年、店舗・事業所の情報が観察できるデジタル電話帳データを用いて、店舗・事業所の時系列データを構築し、その変化要因の分析を行う。本研究で用いる電話帳データには店舗・事業所毎の名称、緯度・経度、住所、業種等の情報が含まれている。

時系列データの構築として、デジタル電話帳データから異なる2時点の電話帳データを店舗・事業所毎にその名称と位置情報に基づいて結合し、またその同一性を判定することで、店舗毎の時系列変化を継続、入れ替わり、新規出現、消滅に分類し、店舗・事業所の立地や業種勢力図などの時系列的な変化を表すことを目的とする。

今回は2011年と2012年のデジタル電話帳データを用いて2時点の店舗・事業所の時系列的な変化を明らかにした。2時点の建物の同一性を判断する上で、3つの段階に分けて考えていった。第一段階として、経緯度が完全一致するもの、第二段階として経緯度が一致しないもの、第三段階として、結合先が見つからないものの三つに分けて建物の同一性を判断していく。なお、全体的な時系列的変化の分類手法としての流れを図1に示した。以下、段階毎に方法を述べる。

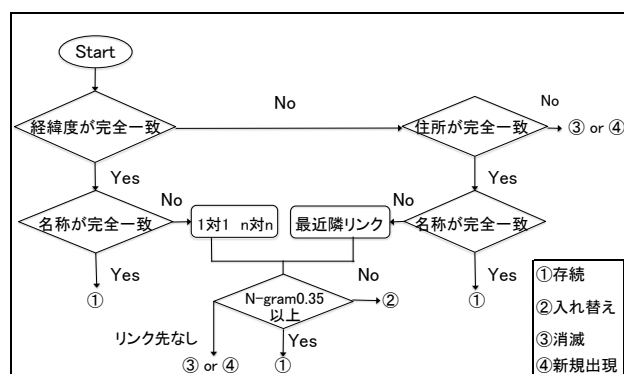


図1 同一性判断手法

2.2 経緯度が完全一致するデータ

第一段階として、2011年と2012年の二時点で経緯度が完全一致する建物を取り出していく。

完全に同一地点にあるデータ同士をリンクし、さらにその中で店舗・事業所の名称の同一性を判断していく。名称の同一性の判定は N-gram を用いて行った。N-gram とは自然言語処理の手法であり、異なる二つの文字列の類似性を明らかにできる。N-gram を用いた文字列の同一性判定においては、文字列の類似度を0から1の値で定量的に表現出来る。この際に同一と判定する値を閾値として設定する必要がある。秋山ほか(2011)はこの値を0.4付近としており、本研究で用いた電話帳データの場合もこの値を0.35とすることで同一性の判定を適切に行えることを確認した。

名称が完全一致する場合は、店舗・事業所が二時点間で場所・名前が同じということなので存続扱いとした。

名称が完全一致しない場合で、同一地点に新旧年ともに1件の場合、名称の同一性を N-gram を用いて判断する。名称の同一性を判断する上で前述した N-gram の閾値の0.35より高い場合は存続、低い場合は入れ替えとする。雑居ビルや商業施設のように同一地点に複数の店舗・事業所が存在する場合も、N-gram により類似度が0.35よりも高い結果を存続とし、低いものは入れ替えとした。

以上の処理を行い残ったもので、新旧年でリンクできない場合は、新しい年のみに店舗・事業所のデータがある場合は新規出現、古い年のみにデータがある場合は消滅扱いとした。

2.3 経緯度が一致しないデータ

第二段階として、経緯度が一致しないデータの空間結合を行う。

経緯度が一致しないが、店舗・事業所の住所と名称が完全一致するものは存続扱いとする。

また、デジタル電話帳データでは雑居ビルなど同じ建物の複数の建物が入っている場合、その建物の重心を入居している店舗・事業所群の経緯度としているが、経緯度が一致しないもの

で一致しない理由として店舗・事業所の増築・改築などで建物形状が変わり、建物の重心がずれてしまい、結果として経緯度も新旧年でずれてしまう場合がある。そのような場合を考慮に入れる為に、ある上限以下の距離の最近隣の店舗・事業所をリンクし名称から同一性を判定するため N-gram を用いた。その結果、存続・入れ替えを検出していく。ここでいう最近隣とは今回は 20m とした。20m とした理由は個人事業所・店舗だけでなく大規模な商業施設などの増築により建物の重心が大きくずれてしまう場合にも対応できるようにこの値をとった。

2.4 結合先が見つからないデータ

第三段階として、第一段階、第二段階の処理を行っても新旧の年で結合先が見つからないものは、新規出現か、消滅に分類していく。結合先が見つからないデータのうち、新しい年のデータしかないものは新規出現、古いデータしかないものは消滅として扱う。

3. 適用例

3.1 東京都八王子市

都市的な側面と郊外的な側面を併せ持つ東京都八王子市を対象として、前述した同一性判断手法を用いて店舗・事業所の時系列的变化を明らかにし、図2の様に表した。図2は店舗・事業所の4つの時系列的な変化の分類（存続、入れ替え、消滅、新規出現）を色ごとに分け、1件毎にポイントデータとして地図にプロットしたものであり、これによってどこの場所がどのような時系列的变化をしたのかが視覚的に分かる。今回は、平面的に時系列的变化を表したので、雑居ビルなどの一つの場所に複数の事業所・店舗がある場所では、複数の変化を表しきれていないことに留意したい。複数の変化が同一場所にある場所でも変化を表す平面的なブロックは重なり合い、一つの変化に見えてしまう。今後は立体的に時系列的变化の分布が可視化できるような表し

方についても考えていきたい。

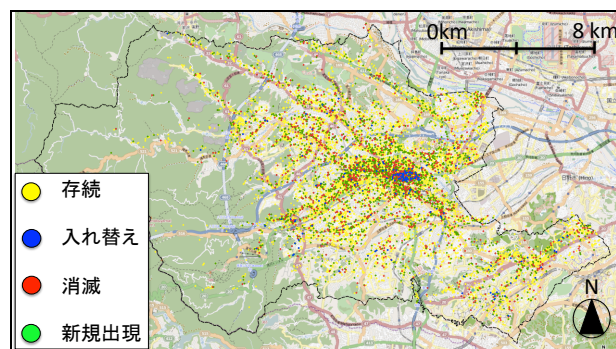


図2 八王子における時系列空間結合の例

3.2 東京都

八王子市で行ったものと同様に東京都全域についても同様の手法により変化を抽出し時系列データを整備し、八王子市の結果とともに表1にその結果を表す。

今後は今回行った結果の信頼性を目視したものと今回の結果と照らし合わせることで検証していき、N-gram の閾値や近くの建物のリンクに使用する最近隣の距離の値についても今回の設定が適切であるか検討していく。これらを行い十分な信頼性が得られたら、全国的な時系列的变化を明らかにし、店舗・事業所の全国版の時系列的变化を表すデータセットを構築して行きたい。

表1 時空間結合結果

	八王子市	東京都
新しい年のデータ数	15873	542483
古い年のデータ数	17443	615158
存続	14992	503552
入れ替え	194	12787
消滅	1457	42101
新規出現	3630	134047
合計件数	20273	692487

4. 時系列的空間データの活用例

今回構築してきた店舗・事業所の時系列变化データベースを用いた応用的な分析の例として、商業店舗の売上の時系列的な変化と周辺建物の時

系列的な変化の関係を考察できると考えられる。

商業施設の売上が変化する要因として、自社要因、他社要因（競合関係など）、市場要因（消費動向）などが考えられるが、今回のデータを使えばこれらの要因のうち他社要因を分析できると考えられる。競合となる店舗の閉店や開店、または大規模な事業所の新規出現、消滅による人口の流入、流出などが構築した店舗・事業所の時系列的变化のデータベースから認識、または予想できる。これらが分かれば、店舗の売上の時系列的な変化から売上の周期性を解析し、周期性に反した異常値を検出したときに、その店舗の周辺建物の変化との関係を分析することで売上に影響を与える要因を検出できるかもしれない。個別の商業店舗の売上に影響する個別の周辺建物変化を分析すれば、相性の良いあるいは悪い業種関係または距離関係などが分かり、企業の出展戦略にも有意義な情報を得ることができると言える。

5. おわりに

本研究ではデジタル電話帳データの名称と位置情報を用いて店舗・事業所の新旧の年について同一性を判断し、時空間結合を行った。今回は2011年と2012年の二時点間で店舗・事業所の時系列的な変化を明らかにしたが、この手法を用いてさらに長期的な時間軸、またより短いスパンで複数の時点間のデータを過去に遡って整備することにより、店舗・事業所の時系列的变化のデータベースを構築し、商業店舗売上の変化要因の検出や企業間の相性をはかる基本データとして分析していきたい。この店舗・事業所の時系列的变化のデータは、都市の店舗・事業所の変遷を表し、前述の通り企業の売上和店舗・事業所の時系列的な変化から出展戦略のデータに使うなど商業的にも利用でき幅広い分野の分析で利用可能な基本的なデータセットになり得ると言える。今後の展望として、構築した店舗・事業所の時系列的变化のデータセットを用いた分析の可能性を広げ

るべく、前述した商業施設の売上和周辺店舗・事業所の時系列的な変化を始めとする分析を行う。

謝辞

今回の店舗・事業所の時系列的变化データセットを構築するにあたり、デジタル電話帳（テレポイント Pack!）をご提供いただいた株式会社ゼンリンに感謝の意を示し、ここに記します。

参考文献

- [1]伊藤香織・曲渕英邦(2001)既存情報を活用した時空間データ作成手法-地図内・地図外情報の曖昧性を考慮した空間要素同定を用いて-。「地理情報システム学会講演論文集」, 10, 147 - 150.
- [2]関口達也・貞広幸雄・秋山祐樹(2011)住宅地滲出型商業集積の形成過程とその要因に関する研究-原宿地域・青山地域・代官山地域を事例とした時空間分析-。「日本都市計画論文集」 Vol. 47, No. 3, 2012年10月
- [3]秋山祐樹・柴崎亮介(2011)位置と名称情報を持つ店舗・事業所データの時空間結合手法の開発-都市地域分析への応用に向けて-GIS-理論と応用、Vol. 19, No. 2, pp. 1 - 11
- [4]Akiyama, Y., Shibasaki, R. (2009) Development of Detailed Spatio-temporal Urban Data through the Integration of Digital Maps and Yellow Page Data and Feasibility Study as Complementary Data for Existing Statistical Information. CUPUM2009, Reference No. 187 Session8.2-2.
- [5]Akiyama, Y., Shibasaki, R. (2012) A Method for Identifying Japanese Shop and Company Names by Spatiotemporal Cleaning of Eccentrically Located Frequently Appearing Words. Hindawi Publishing Corporation Advances in Artificial Intelligence Volume 2012, Article ID 562604, 18 pages doi:10.1155/2012/562604