

平成 21 年度国土政策関係研究支援事業 最終成果報告書

## 研究題名

デジタル地図と電話帳データの統合による  
日本全土における都市の店舗・事業所変化の  
時空間解析

(所属) 東京大学生産技術研究所助教

代表研究者名…熊谷潤

<共同研究者>

(所属) 東京大学大学院新領域創成科学研究科

共同研究者名…秋山祐樹

# 目 次

I. 研究目的・意義	1
II. 研究手法	3
III. 成果内容	4
○ 要旨	4
○ キーワード	8
○ 本編	9
1. 時系列データセット開発	9
1-1. ソースデータ	9
1-2. データの時系列化	10
(1) 時系列化とは	
(2) 時系列化の課題	
1-3. システム開発	11
(1) 開発環境	
(2) 空間統合技術	
(3) 統合後データの同一性判定	
①□テキストの同一性判定	
②□雑音語の除去	
③□ローカル頻出語の除去	
(4) 処理の高速化	
(5) 処理精度の検証	
1-4. 作成データの例	20
1-5. 既存統計情報との突き合わせ検証	21
(1) 事業所・企業統計調査との突き合わせ検証	
①□事業所・企業統計調査とは	
②□データ突き合わせの方法	
③□結果	
(2) 商業統計メッシュデータとの突き合わせ検証	
①□商業統計メッシュデータとは	
②□データ突き合わせの方法	
③□結果	

2. 商店街・商業集積地域の実態とその変化動向の分析	28
2-1. 商業集積地域分布データ作成	28
(1) 本研究における商業集積地域の定義	
(2) 商業集積地域を構成する店舗の業種決定	
(3) 商業集積地域の特定	
① □一般的なバッファリング手法の問題点	
② □本研究のバッファリング手法	
③ □商店街名鑑との突き合わせによる本手法の妥当性の検証	
2-2. 南関東地域における商業集積地域の実態とその変化動向の分析	35
(1) 商業集積地域の分布	
① □規模の変化	
② □新設集積と消滅集積の分布	
(2) 集積ごとのテナントの回転率	
(3) 生鮮三品の充実度	
(4) チェーン店進出状況	
(5) まとめ	
3. 結論	47
(1) 結論	
(2) 国土政策への活用	
○ 資料編	50

## I. 研究目的・意義

都市における事業所、店舗などの立地分布の変化は、都市の活力や魅力に大きな影響を与える。その動向を出来るだけ詳細かつ定量的に明らかにすることは都市空間の持続再生に不可欠である。しかしながら一般的な統計情報である事業所統計、商業統計等では市区町村単位などで集計されており、その集計単位は商店街レベルの詳細な変化を把握することは困難なスケールである。メッシュ統計などでは空間的な詳細さはかなり改善されるものの、商店街のスケールからすれば必ずしも十分でないうえに、業態別店舗総数の時系列変化しか分からないため、個店単位での商店の入替・新規出店、チェーン店の進出など商店街活性化方策に資するより直接的な情報はなかなか得られない。また商店街の盛衰に関しては商店会などを対象にした全国アンケート調査（例：商店街実態調査報告書（中小企業庁））が存在するが、その回収率は良好であるとは言えず、網羅性で大きな問題があるだけでなく、店舗の減少状況や入替の概況をアンケートにより調査しているだけであり裏付けにも乏しい。一方、現地調査や住宅地図などの時系列判読を組み合わせることで、個別店舗の時系列変化を追うことが出来るが、この手法によるデータの整備には多大な労力が必要であり、ごく限られた広さの地域しかカバーできない。

そこで本研究では、全国的に整備されている住宅地図や電話帳を空間的位置と店舗名称に基づいて統合および時系列化し、両データに含まれる経緯度、住所、建物名称、店舗名称などを用いて個別店舗の時間的変遷を自動的に抽出し、上記の課題に応える詳細な時空間データを構築する手法を開発する。また上記の手法を実現出来るデータ処理のためのシステム開発を行う。

本手法が開発されることで個々の店舗・事業所の存続、転換（入替）、新規出現、消滅をほぼ毎年全国規模で把握出来る。全国規模でこうした技術開発を行っている例は皆無であり、本研究の新規性は非常に高いといえる。

また本技術で作成された時系列データは既存の統計情報の補完データとして活用可能である。例えば日本全国の事業所・企業の動向をモニタリング出来る既存のデータセットとして統計局が作成する事業所・企業統計調査がある。このデータは官民において頻繁に用いられているが、事業所・企業統計は更新頻度が2～3年毎、かつ時系列変化は5年毎に固定されている。集計単位も公開されているものは最小で大字であるという制約がある。しかし本技術で作成したデータは店舗、事業所を一件一件単位でモニタリング出来る上に、ユーザーは任意の時点、時点間、集計単位でデータ作成が可能になる。本研究では事業所・企業統計調査との突き合わせ検証を行い、本データの信頼性、利点、限界を明らかにする。

以上より本研究にて開発される技術は、日本全土の都市空間の時系列変化の状況を、極めて細かい空間単位をもって把握出来るデータセットの作成を可能にする。これは

我が国の国土政策の推進にも大いに貢献出来ると言えよう。

さらに本研究では日本各地の商店街・商業集積地域の実態とその変化の動向を上記の手法で開発したデータを用いて把握する試みを行う。現在日本各地の商店街は大型店の展開による競争環境の変化や消費者ニーズの変容、店舗施設の老朽化、後継者不足等により衰退の一途を辿っているケースが少なくない。その結果、郊外の住宅地や地方都市など過疎地域の商店街の中には、空き店舗の増加により商業施設の連続性や業種構成が崩壊され、従来は店舗の集積によって顧客に提供していた利便性や商業地としての魅力を失い、衰退がより一層加速するという悪循環に陥っている。一方で新規店舗の出店・誘致が活発な商店街も存在する。まちの活力を維持するためにも、商店街の実態を把握する必要がある。しかし地方都市をはじめ人口集中地域外にある商店街では、商店街振興組合が成立しないような場合や情報が入手しにくい状態であり、現状把握が困難である。商業集積の実態把握の取り組みは既存研究にも数多く見られるが、それらの多くは対象地域が限定的な定静的ケーススタディであることが多く、労力・時間を要することから汎用性のある手法とは言えないものが多く、上記した広域把握のための手法としては適用出来なかった。

以上の課題を鑑み、本研究では主に電話帳データを用いて個店ベースにおける商店街・商業集積地域の抽出手法を提案する。スタディエリアとして南関東地域全域（東京都・神奈川県・千葉県・埼玉県）を用いる。また抽出した商業集積地域ごとのテナントの回転率、チェーン店の進出状況、生鮮三品店舗の有無、スーパーマーケット等の大規模店舗立地との関連性、消滅および新規出現した商業集積の分布を明らかにする。最終的に南関東地域における商業集積地域の現状を考察する。

本研究の意義を今一度確認しておく。本研究の意義は以下の2点である。

第一に既存の空間データ（住宅地図や電話帳データ）を個店単位で時系列化する手法を実現し、新しいデータセットを開発することである。同時に既存の統計情報との突き合わせ検証も行うことで、本手法にて実現される新しいデータセットの有用性も明らかにする。既存の時系列データでは捉えられなかった詳細さを持ちつつも、既存の統計データに近い性能があることを明らかにする。

第二に電話帳データを用いて国土スケールの商業集積地域の分布を明らかに出来ることである。勿論これまでに商業集積地域・商店街についての研究は数多く成されてきている。それらの多くが様々な調査手法や文献調査を通して詳細な知見を得られているが、こうした研究の多くに共通している課題はその手法の汎用性の低さである。個々の商業集積地域・商店街の分析については既存研究に見られる手法を用いる場合の方が詳細な結果を得られることが期待出来るが、本手法の意義は個々の商業集積地域や商店街の極めて詳細な分析結果を得る事では無い。商業集積地域・商店街の分布を国土スケールで把握出来ることに大きな意味がある。現時点で商業集積地域や商店街の分布とその変化の動向を国土スケールで把握出来るデータは存在しない。

## II. 研究手法

### 1. 時系列データセット開発

前述したデータセットの実現には建物名称、階数、部屋番号までを考慮した3次元空間統合の手法と、店舗・事業所名称に着目した同一性判定手法が必要である。住宅地図の入居者名称や電話帳のテナント名称は同一のものでも時点が異なる場合に表記が変化するケースも少なくない。そのため表記揺れの影響を吸収出来るテキストの同一性判定技術、特に同一性判定にあたり雑音となる頻出語・地名・駅名等を除去するためのライブラリの作成、それを用いた名称データのクリーニング技術が重要であり、本研究では開発と処理精度の検証をあわせて行う。

また開発されたデータセットの信頼性をチェックするため、既存の時系列データとの突き合わせ検証を行う。既存の時系列データとして事業所・企業統計調査の時系列変化情報を用いる。両データの集計単位を合わせ、その結果を突き合わせることでデータセットの信頼性を明らかにする。

### 2. 商店街・商業集積地の実態とその変化動向の分析

上記の手法にて開発されたデータを用いて商店街・商業集積地の実態把握とその変化の動向を観察出来る手法を具体的に検討する。まず電話帳データを用いて各種商店街や商業集積地（観光地における店舗群や郊外型のロードサイドショップ群を含む）の分布とその広がりを明らかにする。作成したデータを商店街名鑑データとの突き合わせを行うことで、その分布や規模の妥当性を検証する。

本研究では上記の手法で開発した商業集積地分布データに電話帳データや住宅地図の時系列化データを合わせて利用することで、南関東地域全域（東京都・神奈川県・千葉県・埼玉県）の商店街・商業集積地のテナントの回転率、チェーン店の進出状況、生鮮三品店舗の有無、スーパーマーケット等の大規模店舗立地との関連性、消滅および新規出現した商業集積の分布を明らかにする。最終的に南関東地域における商業集積地の現状を考察する。

### Ⅲ. 成果内容

#### ○要旨

都市空間の詳細なモニタリングを行うために、既存の空間情報の時空間統合処理を自動で行うシステムを開発した。本システムは住宅地図と電話帳の統合および新旧データの時空間統合が可能である。

本システムの特徴は一般的な GIS ソフトウェアでは扱えない 3 次元的な位置情報を認識し空間結合を行える点である。ソースデータが建物の入居情報(入居する建物名、階、部屋番号)を持つ場合、階情報や部屋情報を認識し、部屋単位やフロア単位での空間結合が可能になる。同一地点に複数のデータがプロットされてしまう高層ビルや雑居ビルのテナントデータのようなデータは一般的な GIS ソフトウェアでは適切な結合先を見つけることが出来ないが、我々のシステムはこの点をクリアし、人が目視で統合作業を行ったような正確な結果を得ることが出来る。

さらに統合後データの同一性を判定することで、統合したデータ同士の変化情報を取得することが出来る。新旧 2 時点のデータを投入した場合には、名称変化情報から 2 時点間でのテナントの存続、入替、新規出現、消滅の情報を取得出来る。このような名称の変化情報を認識させるのは容易ではない。たとえ同一のデータであっても、時点間やデータ間でその表記に揺れが見られるためである。そこで本システムではこうした表記の揺れを吸収し、人間の判断に近いファジーな結果が得られる手法を開発した。まず **n-gram** と呼ばれる自然言語処理の手法を用いて、2 つの文字列の類似度を定量化する。コンピュータは通常は文字列の一致不一致の判断を完全一致か否か、という単純なルールによって行っている。しかし表記揺れの問題でそのような単純な方法では適切な結果が得られない。そのため文字列同士の類似度を定量化し、ファジーな判断を行えるようにした。

しかしこれだけでは不十分である。店舗や事業所の名称内には頻出する文字列(株式会社・喫茶店・銀行など)や地名、駅名が含まれており、それらの影響で **n-gram** だけに依存した判別方法では適切な結果が得られない場合がある。そこで南関東全域の住宅地図の入居者名称と電話帳の店舗・事業所名称を形態素解析し頻出語辞書を作るとともに地名や駅名の辞書も作成し、**n-gram** による名称類似度判定の前に頻出語、地名、駅名を名称情報内から削除することで判定精度を大幅に向上させた。

さらに上記の辞書でも洗い出せない時間的、あるいは空間的に偏りが見られる頻出語が存在する。本研究ではこうした頻出語を「ローカル頻出語」と呼ぶ。ローカル頻出語の除去は入力データを経緯度 1/1000 度のグリッドに分配し、ローカル頻出語削除対象のデータ自身が含まれるグリッドとその周囲、合計 9 グリッド内に含まれる全データの名称情報を **n-gram** を使って **9-gram** から **4-gram** まで作成し、さらにその中から複数回出現するものを候補として、それらを使ってクリーニングする。こうすることでローカル頻出語のクリーニングを実現する。

上記のシステムを用いることで店舗・事業所の新旧 2 時点間の変化の様子を業種も含めて個店単位で観察することが出来る時系列データセットが実現する。なお本システムは日本全土のデータ処理にも対応している。

作成した時系列データセットの信頼性を評価するために、既存の統計情報である事業所・企業統計調査との突き合わせ検証を行った。我々の時系列データは 2000 年と 2005 年の 2 時点の電話帳データの時系列データである。一方事業所・企業統計調査は 2001 年と 2006 年の 2 時点の時系列データである。事業所・企業統計調査は様々なスケールの集計単位と業種分類体系が存在する。本研究で我々が使用したのは東京 23 区を区ごとにそのテナントの時系列変化を集計したものである。業種ごとに存続、新設、廃業の件数を得ることが出来る。一方我々のデータは電話帳の業種情報が利用出来る。電話帳側の業種情報を事業所・企業統計側の業種に置換するとともに、時系列変化情報も事業所・企業統計側の集計方法に合わせることで、東京 23 区を区ごと、業種ごとに突き合わせて検証することが出来る。

検証の結果、我々のデータは既存の統計データと比較した場合に多少の誤差や限界はあるものの、多くの業種でその時系列変化の割合や件数が類似した結果となっており、事業所・企業統計調査の代替に耐えうると判断出来た。また我々のデータの強みはユーザーが任意のエリア、年次間のデータ、集計単位を設定して時系列データを作成可能な点である。事業所統計は 2～3 年おきにしか更新されない上に、時点間も 5 年に固定されている。この制約を取り払うことが出来たことは極めて有意義である。しかも最小集計単位はテナントスケールである。これは事業所・企業統計調査を他の統計データや空間データと統合して利用したい際の障害であった集計単位の整合という課題を解決出来る。

続いて電話帳データと商業統計メッシュデータの突き合わせ検証も行った。本研究では電話帳データを用いて商店街・商業集積地域の抽出と分析を行うため、電話帳データは既存の統計情報と比較した場合、同程度の性能を発揮出来るのか、と言う点を検証しておく必要がある。そこで商業統計メッシュデータとメッシュ単位での突き合わせを行い、その掲載件数の当てはまりの良さを確認する。電話帳データは 2005 年、商業統計メッシュデータは 2004 年のものを使用した。検証対象地域は東京都全域とした。

検証の結果、小売店総計、買回品総計、最寄品総計等の複数業種の総計では当てはまりが良く、商業統計と類似した性能を持つことが分かった。一方業種を絞ると当てはまりは悪くなるが、業種によってはまずまず類似した結果が得られる。業種の突き合わせ方法の再検討で、より良好な結果が得られる可能性もある。電話帳データからある特定の業種だけの件数を正確に取得することは業種によっては難しいが、複数業種にくくり込んで使えば実用的であることが分かった。本研究の商業集積地域の抽出は個別の業種データの立地はあまり重要ではなく、商店街を構成すると見られる複数



業種のデータ集積からその集積の程度を明らかにして行くものである。そのため電話帳データは商業集積地域や商店街の分布抽出を行うに適切なデータと判断出来る。

続いて電話帳データを用いて商業集積地域を自動抽出する手法を提案する。本研究で扱う「商業集積地域」とは、鉄道駅周辺に構成される駅前商店街、都市の繁華街等の中心部に構成される都心型商店街、鉄道が整備される以前に起源を持つ宿場町型や門前町型の商店街、観光地に構成される商店群など一般的な商店街全般を指す。また近年登場した主要道路沿いのロードサイド型店舗群や大規模ショッピングセンターとその入居テナント群も含む。本研究の目的達成のためにはこうした「商業集積地域」の分布を特定する手法が必要である。

そこで本研究では電話帳データが保有する全件に付加された 1751 種類（2005 年の場合）の業種情報を利用し、商業集積地域を構成する店舗・事業所の抽出を行う。業種情報を利用するためには商業集積地域を構成する業種を決定する必要がある。商業集積地域を構成する業種の決定にはそれらの地域を構成する店舗の実態情報を集めるべきである。そこで東京 23 区内に存在する 28 箇所の商店街、商店会のホームページから店舗情報（店舗名称・住所）を入手し、アドレスマッチングにより経緯度情報を与えた。それらを前述した時空間データの統合システムを用いて、電話帳データと統合した。我々が保有する電話帳データは 2005 年 12 月時点のものであるため統合出来ないデータもあったが、ホームページから収集した 3476 件中、1932 件が統合出来た。統合に成功したデータの電話帳の業種より、商業集積地域を構成する業種 401 種類を決定した。各種食料品・衣料品・日用品販売店、美容・衛生、各種飲食店、スポーツ・趣味・娯楽関連施設、医療、不動産、金融、宿泊施設等が含まれる。

電話帳データはアドレスマッチングによって得られたポイントデータである。このデータから商業集積地域の範囲を意味するポリゴンを作成し、それらが近隣関係にあるポイント同士を同じ領域に含む集積範囲を意味していれば、それを 1 つの商業集積地域と見なせる。そこで本研究ではポイントデータからバッファリングを行い、重複するバッファポリゴンを統合して 1 つのポリゴンとして連担させていく。バッファリング手法は本研究独自の手法を用いる。本手法は地域特性を考慮出来る、そして商業集積地域から遠いデータには連担が起こらない可変的なバッファリング距離を設定出来るものである。最後に収録件数が少ないポリゴンを取り除く。商業統計調査表によると 20 店舗以上の店舗集積が見られる地域を「商店街」と定義している。ただし郊外のロードサイド型集積や衰退傾向にある地方の小規模な商店街では 20 件もの集積の構成は期待出来ない。そこで本研究では 10 件を閾値として用いることにする。すなわち作成されたポリゴンに含まれる店舗数が 10 件未満の場合、そこは商業集積地域と見なさない。以上の手法によって南関東地域から東京都 4022 箇所、神奈川県 1879 箇所、千葉県 1134 箇所、埼玉県 1387 箇所の商業集積を得た。

商業集積地域を抽出する手法を提案したがこの手法の妥当性を明らかにしておく必

要がある。そこで本研究では2004年の東京都の商店街名鑑（全国商店街振興組合連合会発行）を用いて、東京都全域の商店会が組織されている明らかに商店街が存在すると考えられる地点を収集した。商店街名鑑には商店会の代表組織や代表者の住所と商店会に加盟している店舗の件数が収録されている。この店舗件数を近似的にその商店街の店舗数とする。また住所情報があるためアドレスマッチングを行うことで経緯度情報に変換出来る。すなわち商店会に加盟している店舗数情報を保有するポイントデータを作ることが出来る。一方我々のデータはポリゴンデータである。そこでこれらのポリゴン内に商店街名鑑のポイントデータが重複して存在した場合に、それを我々のモデルが空間的に一致したと見なす。2004年の商店街名鑑ポイントデータと2005年の商業集積ポリゴンを統合した結果、商店街名鑑に登録されている2184件のポイントデータのうち、1617件（74.04%）が商業集積ポリゴンに内包された。なお一つのポリゴンに複数の商店会が含まれた場合にはその総和をとって比較している。我々の手法で作成した商業集積ポリゴンに含まれるデータ件数と商店街名鑑の収録件数との相関をとると相関係数は $R=0.6643$ とかなり強い正の相関が見られる。我々の商業集積地域の抽出手法が空間的な位置という意味でも、またその規模を捉える意味でも適した手法であることが分かる。

以上の手法で得られた商業集積分布データと電話帳や住宅地図の時系列データを組み合わせて南関東地域全域における商業集積地域の分析を行った。南関東地域における商業集積地域の現状と変化を以下の点に注目して分析した。

- ① 集積の分布とその盛衰と規模の変化
- ② 集積ごとのテナントの回転率
- ③ 生鮮三品の充実度
- ④ チェーン店進出状況

集積規模は特に東京都心部でその規模、件数ともに増加傾向にあった。東京都心部では以前は複数の別々の集積であったものが統合して巨大な集積に成長するケースが多く見られた。東京都心部以外でも地方の中心都市や県庁所在地都市の中心部では集積規模はそれほど大きくなってはいないが、件数が増加しているケースが多く、集積内の店舗・事業所密度が高まっていることがうかがえる。一方で東京都市部や東京都以外の県の中心都市以外では規模も件数も減少傾向であった。

テナントの回転率では特徴的な分布が見られた。東京都心や地方中心都市の中心部では非常に高い回転率となる。そしてその近傍の周辺地域では回転率が急激に小さくなり、さらにその外では都心部と近傍周辺地域の間くらいの値をとるという構造が見られる。大都市周辺部にドーナツ状の回転率が低い地域が出現するという構造である。大都市中心部での回転率の高さはそれらの地域の不動産価値の高さや商業活動における競争率の高さに起因すると思われる。一方前述したドーナツ状地域の外側でもまずまずの回転率の高さが見られるのは、チェーン店の進出によるものが大きいのではな

いかと予想される。

生鮮三品業種の充実度では生鮮三品業種が全て揃っている集積を抽出した。その結果、東京都心部周辺や東京都市部に多くの該当集積が見られた。また東京都と神奈川県では全集積の約 10%が該当するのに対し、千葉県や埼玉県では 6%前後と低い値を示す事も分かった。千葉県や埼玉県では東京都や神奈川県よりも郊外化が進展しているためだと考えられる。また 2000 年は三品業種が揃っていたが 2005 年にはそれが崩壊した集積の分布を見ると、大都市の中心部にそうした集積が集まっていることが分かった。生活のための商店街という機能を捨てて、特定業種が偏在した専門店街や遠方からの集客力が大きい買回品中心の業種構造、あるいは観光客向け店舗群という機能にシフトしたものと考えられる。

チェーン店の進出状況は東京都心や地方中心都市の中心部ほどチェーン店率は低く、郊外に向かうに連れてその値が大きくなることが分かった。とりわけ千葉県や埼玉県ではチェーン店率の高い集積が多くなっており、郊外化が進んでいることがよく分かる結果となった。

以上をまとめると東京都心部や地方中心都市の中心部では生鮮三品の充実度が低い集積が多く、生活のための商店街では無くなりつつあることが分かる。大都市の周辺部（上述のドーナツ状地域）では集積規模が安定しており、生鮮三品の充実度も高く、テナントの回転率も低い。これらの地域では日々の生活に密着した商店街が機能していると言えよう。郊外では全体的に規模も件数も縮小傾向である。また東京都市部以外では生鮮三品業種も不足しておりチェーン店が増加し続けている。明らかな郊外化傾向が確認出来る。

## ○キーワード

都市モニタリング、時空間データセット、電子地図、電話帳データ、商業集積地域、商店街、統計データ

## ○本編

### 1. 時系列データセット開発

本章ではデータセット作成の流れおよびその課題を紹介する。まずソースデータについて説明する。続いてデータセット作成のためのシステム開発の流れ、およびその課題を説明する。

#### 1-1. ソースデータ

我々が欲しているような詳細な空間情報はどのようにして得ればよいであろうか。幸いにもわが国にはその要求に答えられる精細なソースデータが存在する。住宅地図と電話帳データである。いずれのデータもほぼ日本全土をカバー出来る個店スケールのミクロな情報を保有している。

住宅地図（株式会社ゼンリン『Zmap-TOWNⅡ』）の場合、2007年4月1日現在で全国1827市町村のうち1817市町村をカバーしている。住宅地図は地図情報を保有しており、建物に入居する入居者データを属性情報として保有している。本研究ではこの入居者データを用いる。東京都の場合、個人宅を除いても約150万件のデータを持つ。データの収集は調査員による現地調査によって行われており、基本的には現地にて目視により確認出来る全ての入居者情報が収録されている。

電話帳データ（NTT情報開発株式会社『タウンページデータベース』）は全国約1000万件の店舗・事業所等の情報を保有する。また本研究ではタウンページデータベース以外の電話帳データも利用している（株式会社ゼンリン『テレポイントデータ』）。電話帳データは事業主の掲載申請によって収録されるため、網羅性は住宅地図よりも劣る（概ね住宅地図入居者データ件数の半分弱）。ただし業種情報を持つ点が優れている。タウンページデータベースはテキストデータであり、位置情報は住所しか保有していない。そこで住所に基づいてアドレスマッチングを行うことで、経緯度情報を得る。

いずれのデータも店舗・事業所等の名称、住所、入居する建物情報（建物名・階・部屋番号等）、経緯度情報を持たせることが出来る。また住宅地図の入居者データは上記の必要な情報だけを集めたテキストデータとして出力しておくことで、電話帳データとデータの種類を揃えておく。

本研究では商店街の変化の動向を把握すると言う意味で、近年商店街のあり方に大きな影響を与えた大規模小売店舗立地法（2000年6月）が施行された直後の2000年12月頃のデータとその5年後の2005年12月頃のデータを用いた。大規模小売店舗立地法は従来の大規模小売店舗法と異なり、その店舗の規模についての審査が無く、あくまでも地域社会との共存に主眼が置かれている。このことは結果的にスーパーマーケットや大規模ショッピングセンターの経営者にとって、従来よりも出店が容易な環境を作り出した。その結果、日本各地の商店街のシャッター街化の進行がさらに深刻になっていると言われている。そこで本研究ではこの時期の変化に着目していく。

## 1-2. データの時系列化

### (1) 時系列化とは

本研究における「時系列化」とは住宅地図や電話帳の異なる時点間のデータ同士を個店単位で統合するとともに、それらが保有する店舗・事業所名称の同一性を判定し、その変化情報を取得するものである。図1にデータの時系列化のイメージを示す。

### (2) 時系列化の課題

以上のようなデータ統合処理を行うに当たって、大きく分けて以下の2つの課題が立ち上がる。まず建物情報を処理出来る三次元的空間統合技術の実現である。一般的なGISシステムでは三次元的な空間統合は行えない。その上電話帳データの場合、建物情報はテキストデータで格納されている。このテキスト内から建物名、階、部屋番号を認識させる必要がある。その上で空間的な位置情報、階、部屋番号までを含めて適切な統合先を検索出来る手法を開発しなければならない。

もう一つの課題が統合後のデータの同一性判定である。両データの同一性判定はその店舗や事業所の名称に基づいて行われる。すなわちテキスト情報の同一性を判定する技術が不可欠である。



注: 上記の店舗・施設・建物等の名称、住所、経緯度等は全て架空のものである。

図1 新旧2時点のデータの時系列化の出力結果のイメージ

以上2つの処理を自動で行うシステムの開発が必要である。本研究で扱うデータは従来の商店街に関する研究に見られる小規模なエリアを対象としたものではない。対象地域は南関東全域(将来的には日本全土)である。例えば東京都全域の電話帳データの収録件数は2005年の場合で690189件、2000年は776998件である。このような膨大なデータの統合とその同一性判定を手作業で行うのは余りにも非現実的である。

### 1-3. システム開発

本システムは名称情報（店舗・事業所名称等）と位置情報（住所，経緯度，建物情報）のみを使って異なる時点間のデータの時系列化を実現する。二時点間の個店単位の時系列変化情報（存続・入替・新規出現・消滅）の取得が可能となる。なお本手法はそのまま異なる2種類のデータ統合にも利用できる。

#### (1) 開発環境

本システムは市販されている一般的なパソコンを用いて開発した（Intel Core2 Quad Q6700 2.66GHz・メモリ 3.23GB・OS: Windows XP SP3）。言語はRuby1.9.1およびRuby1.8.6である。一般的なパソコンと環境で開発を行った理由は、本システムが特別な環境下のパソコンで無くとも動かすことが出来るよう考慮したためである。

#### (2) 空間統合技術

ソースデータが保有する位置情報を用いて、最も類似した位置情報を持つデータ同士の統合を行う。一般的なGISソフトウェアの空間結合では空間データの二次元的な空間情報しか参照出来ない。即ち経緯度情報しか参照することが出来ない。また同一地点に複数のデータが存在する場合（高層ビルの入居テナントデータ等）統合先の候補が複数出現してしまうため統合先の特定制が出来ない。そこで本システムでは住所や建物情報（入居する階や部屋番号）も参照出来る三次元空間統合を実現することでこれらの課題を解決した。図2に本システムの空間統合の流れを示す。

電話帳データの場合、建物情報は1つの文字列で表記される。即ち建物名称、階、部屋番号は分かち書きされること無く1つの文字列で表記される。この文字列の中から建物名、階、部屋番号を認識させなければ三次元的な位置情報を取得することは出来ない。そこで2005年の南関東地方全域の住宅地図の入居者情報と電話帳データが持つ建物情報を用いて、建物情報の分割パターンライブラリを作成した。このライブラリ

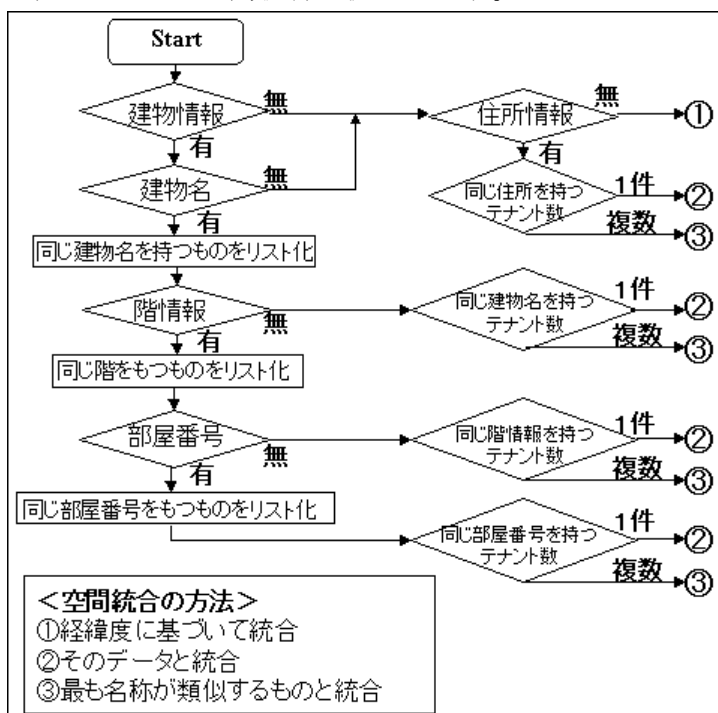


図2 空間統合のアルゴリズム

を用いることで建物情報を建物名称、階、部屋番号に分割して認識させることが出来る。図3は建物情報分割処理の例である。本研究で開発した位置情報の統合技術は経緯度だけでなく、住所や建物情報といった非定量的な位置情報の処理も可能な点が優れている。これは一般的なGISソフトウェアが成しえない空間統合技術である。

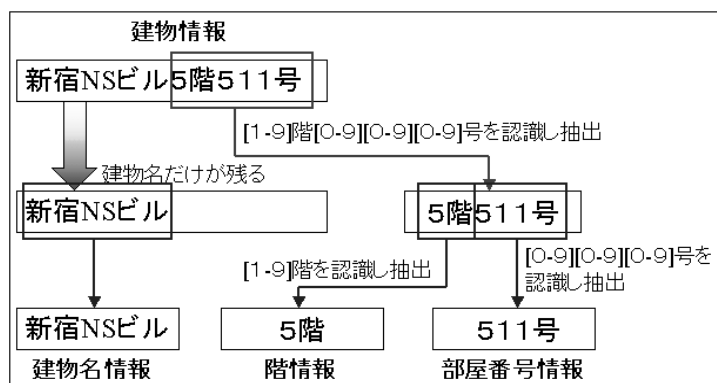


図3 建物情報分割処理の例

ただしこれだけでは不十分である。建物情報は常に建物名、階、部屋番号を完全に保有しているとは限らない。また建物情報を持っていないデータも少なくない。建物情報がこのように欠損した場合、複数の候補が出てきてしまう。そこでそのような場合には次項(3)で紹介する店舗・事業所名称の類似度を用いて、全候補の中から最も名称が類似したデータと統合を行う。

### (3) 統合後データの同一性判定

#### ① テキストの同一性定量化

統合後データの同一性判定は双方の店舗・事業所等の名称を用いて行われる。統合後データの同一性判定を行うことで二時点間の時系列変化情報を取得出来る。図4に時系列処理で得られる結果を示す。データ統合に利用する場合には存続を統合成功、入替を名称不一致、新規出現を統合元のみ、消滅を統合先のみとして利用できる。

新しいデータ	古いデータ	統合先有無	名称同一性	結果
A	A	有り	一致	存続
B	C	有り	不一致	入替
D		無し(統合元のみ)	---	新規出現
	E	無し(統合先のみ)	---	消滅

図4 時系列処理で得られる結果

本手法の実現には名称情報の同一性判定、即ち2つの文字列の同一性の判定技術が必要になる。店舗や事業所の名称はデータの時点間によってその表記に揺れが生じる。表1に表記揺れの例を示す。表1は2005年と2000年の東京都千代田区の電話帳データの一部を手作業で統合した結果、名称情報が完全一致しなかったものの一部である。同様の検証を東京都新宿区西新宿全域で行ったところ、名称が一致すると判断出来た

4161 件のうち、724 件（17.40%）は完全一致ではなかった。即ち完全一致か不一致かといった単純なルールでは適切な結果を得ることが出来ないことが分かる。そのためテキストの類似度を定量化する方法が求められる。

そこで本技術では n-gram を用いてテキストの類似度の定量化を行った。n-gram とは自然言語処理の一手法であり、異なる 2 つの文字列や文章の類似性を明らかに出来る手法である。近年では文学や言語学等の分野でも注目され始めている手法である。

本研究では隣り合う 2 文字を次々と取り出して同一性を比較する手法である bi-gram を用いる。図 5 に bi-gram による処理の例を示す。文字列 i と文字列 j の類似度は式 1 によって定義される。

$$S_{ij}^{(n)} = \frac{n_{ij}^{(n)} + n_{ji}^{(n)}}{m_i^{(n)} + m_j^{(n)}} \quad (1)$$

$S_{ij}^{(n)}$  : 文字列 i と文字列 j の類似度

$m_i^{(n)}$  : 文字列 i から取り出した n 文字の文字群の数

$m_j^{(n)}$  : 文字列 j から取り出した n 文字の文字群の数

$n_{ij}^{(n)}, n_{ji}^{(n)}$  :  $m_i^{(n)}$  と同じ  $m_i^{(n)}$  の個数

表 1 表記揺れの例

電話帳2005年・名称情報	電話帳2000年・名称情報
室町クラフト	室町工芸
本間経理事務所	本間会計事務所
有限会社マインドプランニング	マインドプランニング
村役場秋葉原店	村役場
株式会社イーエスピーテクニカルハウス	株式会社イーエスピー お茶の水店
リナスサンドウィッチカフェアテネフランセ店	リナスサンドイッチカフェ
はるだんじ本店	海鮮丼屋はるだんじ
台南担仔麵ドーム店	台南担仔麵 水道橋ドーム店
プロント 大手町アネックス店	プロント大手町店
株式会社水戸忠交易	株式会社水戸忠
クリツア	クリツア
有限会社カセコ商店	安全帽のカセコ
コンピューターのおっと UNIX本舗	コンピューターのおっと1号店
ホビーランドぼち秋葉原2号店	ソフトランドぼち2号店
イタリアン・トマトカフェジュニア秋葉原店	イタリアン・トマトカフェジュニア・秋葉原店
フリーク	ビデオBOXフリーク
秋葉原ラジオ会館共栄会	ラジオ会館共栄会
むさし坊 市ヶ谷番町店	むさし坊 番町店
株式会社ヤスキ東京支社	株式会社ヤスキ 本社
旭化成株式会社 旭リサーチセンター	株式会社旭リサーチセンター
ジョージエンセン インペリアル店	株式会社ジョージエンセンジャパン
ゴールドファイルインペリアルプラザ店	ゴールドファイル
小林法律事務所	小林・平湯法律事務所
双日プライベートエクイティ株式会社	ニチメンプライベートエクイティ株式会社



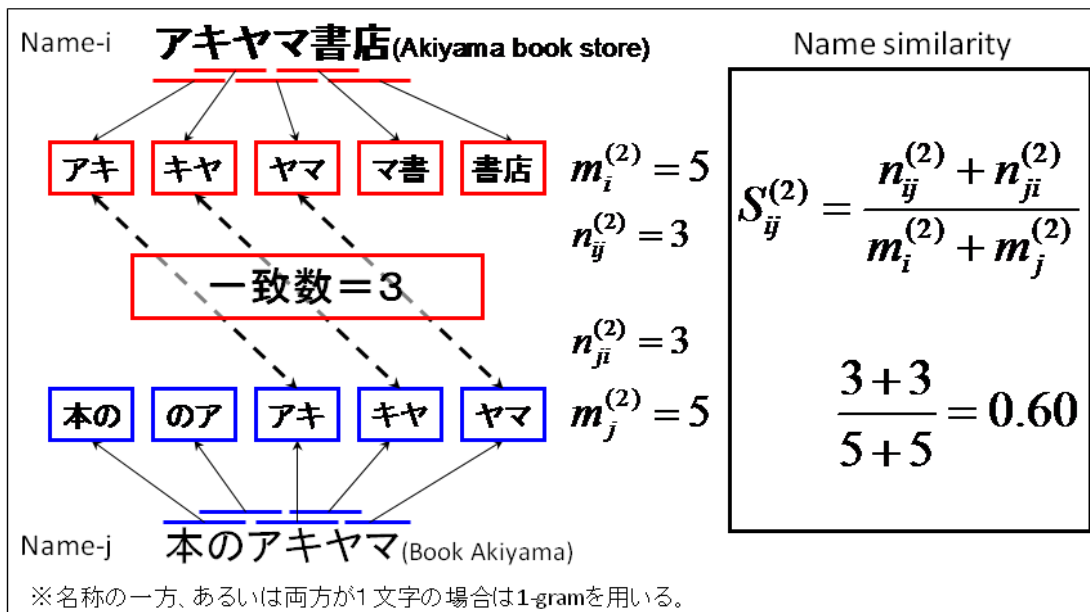


図4 bi-gram

$S_{ij}^{(n)}$  の閾値はサンプルデータを取り出し、手作業にて同一性判定を行った結果と bi-gram での処理結果を比較して適性値を決定した。図5に東京23区から抽出した3000件のサンプルの bi-gram の値の分布を示す。他のいくつかの地域でも同様の分析を行った結果、 $S_{ij}^{(n)}$  の閾値は0.4が適切であった。即ち  $S_{ij}^{(n)}$  が0.4以上の場合にそのテキスト同士を同一のものと見なす。

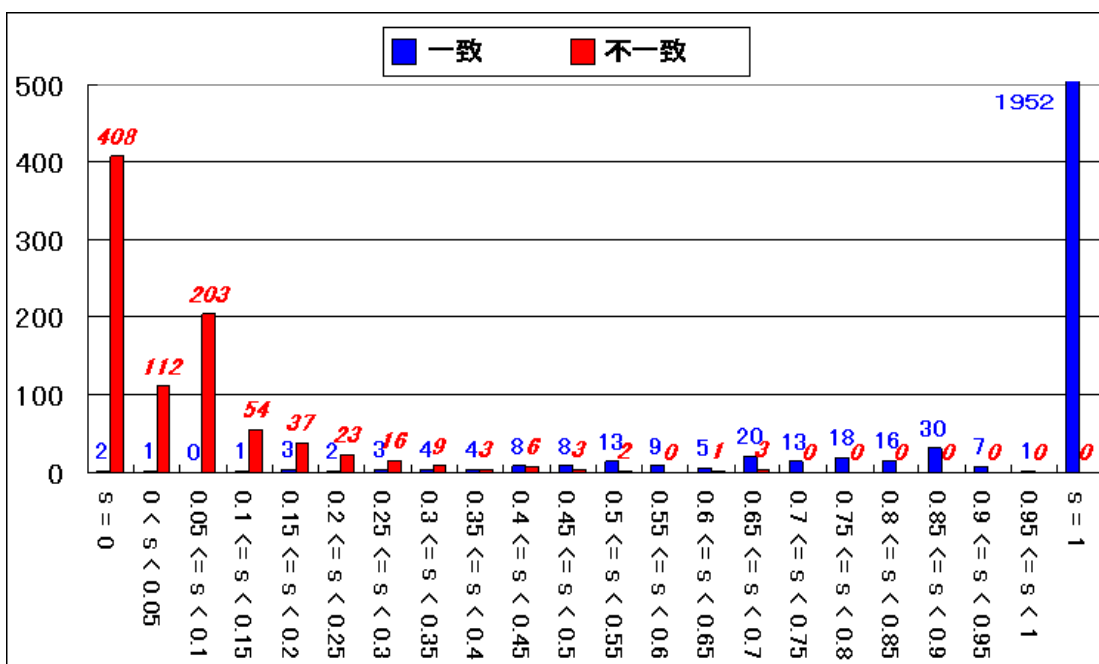


図5 東京23区からとり出した3000件の名称一致の場合と不一致の場合の bi-gram 分布

## ②雑音語の除去

店舗や事業所等の名称には頻出語や地名、駅名が含まれているケースが多い。表 2 に電話帳データに見られる頻出語の例を示す。これらの影響で n-gram のみに頼る名称の同一性判定では適切な結果が得られない。このように名称の同一性判定に影響を与える文字列を本研究では「雑音語」と呼ぶ。

そこで頻出語、地名、駅名のライブラリを作成し、各種頻出語を除去することでこの問題に対応する。頻出語は 2000 年と 2005 年の東京 23 区全域の住宅地図の入居者名称情報と電話帳データの名称情報を形態素解析することで作成した。形態素解析には形態素解析システム『茶筌』を用いた。地名は『日本行政区画便覧データファイル(日本加除出版株式会社)』から、駅名は現在および過去の時刻表から駅名を収集し作成した。地名辞書には明治期以降の全国の丁目スケールまでの全地名、駅名辞書には 1945 年以降の全国全ての鉄道駅名を収録している。図 6 に頻出語除去処理の例を示す。

表 2 電話帳データに見られる雑音語の例

名称情報	雑音語	種類
グレイスジャパン株式会社	株式会社	頻出語
株式会社ゼンリン	株式会社	頻出語
サーティワンアイスクリーム 麻布店	麻布店	地名
ストロベリーコーンズ 白金店	白金店	地名
ハーゲンダッツショップ 青山店	青山店	地名
ホブソンス・アクアシティお台場店	お台場店	地名
マリオジェラテリアウイング高輪店	高輪店	地名
ビックユニフォーム赤坂店	赤坂店	地名
てもみん 新橋駅前店	新橋駅前店	駅名
まんが喫茶ゲラゲラ赤坂店	まんが喫茶	頻出語
	赤坂店	地名
株式会社ホブソンスジャパン 西麻布店	株式会社	頻出語
	西麻布店	地名
アコム株式会社 新橋駅前支店	株式会社	頻出語
	新橋駅前支店	駅名
株式会社アイシービーアイワイジャパン白金台店	株式会社	頻出語
	白金台店	地名

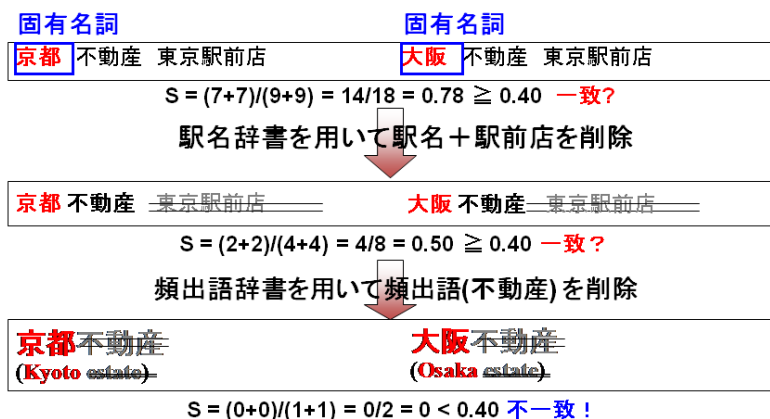


図 6 頻出語除去処理の例

### ③ローカル頻出語の除去

前述した雑音語の除去だけでは不十分な場合がある。時空間的に偏りがある頻出語が存在するためである。本研究ではこうした頻出語を「ローカル頻出語」と呼ぶ。例えば「新宿西口」や「八重洲口」は地名ではなく、また「八重洲口」は駅名でもない。しかしこれらの文字を含む店舗や事業所等は新宿駅西部や東京駅東部（八重洲口側）で数多く見られる。これらが「空間的に偏りがある」頻出語の例である。図7に2005年の電話帳データにおける「新宿西口店」が名称情報に含まれるデータの分布を示す。

「六本木ヒルズ」はどうだろうか。これは空間だけでなく時間的にも偏りが見られる頻出語の例である。六本木ヒルズの開業は2003年であり、2003年以前の店舗や事業所名に「六本木ヒルズ」という文字列が含まれることは殆ど無い。例えば東京都の電話帳データの場合、「六本木ヒルズ」をテナント名称に含むデータは2005年の場合で105件、2009年の場合は141件存在するが、2000年のデータには0件である。このようなローカル頻出語を除去する技術が必要である。

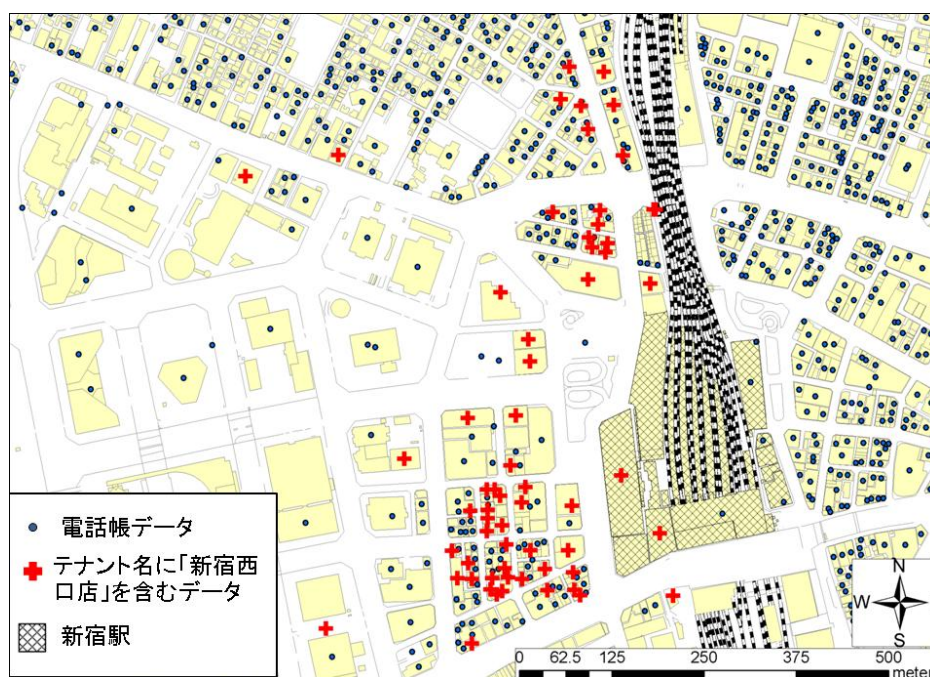


図7 名称情報が「新宿西口店」を含むデータの分布（2005年・電話帳データ）

ローカル頻出語除去の難しさは前述した時空間的な偏りである。時空間的に極めて局所的に出現するために、前項で紹介したような頻出語の辞書を事前で作ることが難しい。しかしながらローカル頻出語を無視してしまうと、統合精度は低下してしまう。

このような問題の解決には「局所的な」頻出語辞書を作ってやれば良い。そこで本研究ではソースデータの経緯度を用いて経緯度 1/1000 度四方のグリッドを作成し、グリッドごとに頻出する文字列を検索することでこの課題を解決した。頻出する文字列

の検索には n-gram を用いる。グリッドごとにそのグリッドと隣接するグリッドを含めた全 9 グリッドに含まれる全データの名称情報を用いて 9-gram から 4-gram を作成する。周囲のグリッドを含む理由はグリッドの境界付近に位置するデータにも適切な処理を行うためである。除去対象となる文字列は 9-gram から 4-gram で検出された文字列のうち複数回の出現が見られ、かつ語尾に「店」・「支社」・「営業所」・「東口」・「西口」・「南口」・「北口」を含むものとする。また除去の際には可能な限り長い文字列を確実に除去するために文字数の多い頻出語から順に除去していく。すなわち 9-gram で得られ、かつ上記の条件を満たす文字列から除去し、最後に 4-gram で得られたものを除去する。図 8 は新宿駅西側のあるグリッドに見られるローカル頻出語とその除去処理の例である。一部で雑音語が残る場合や余分な除去を行ってしまう例も見られるが、多くは適切に処理されていることが分かる。なお雑音語が部分的に残ったり、余分な削除を行ったりした場合でも、名称の類似度判定に n-gram を用いるため、その影響をほぼ無視出来る。

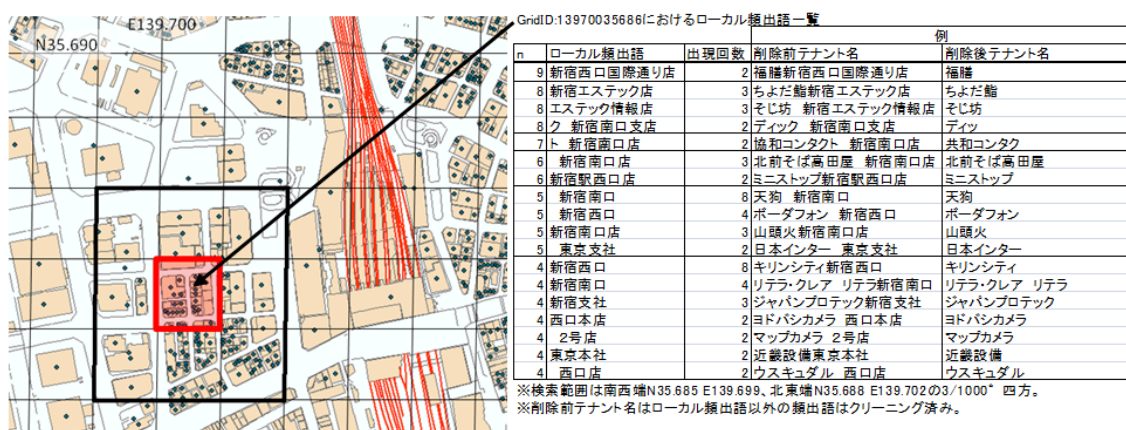


図 8 新宿駅西側におけるあるグリッドでのローカル頻出語除去の成果

#### (4) 処理の高速化

空間検索を行う際に入力データが膨大な場合、問い合わせを行うデータも同様に膨大な数となる。そこで経緯度に基づき、あらかじめ入力データにその経緯度に基づいた ID を付加し、経緯度 1/1000 度グリッドに分配する。空間検索時には問い合わせ元のグリッドと、それに隣接するグリッドのみを検索範囲とすることで処理の大幅な高速化を実現した。図 9 に本研究の空間インデックスの手法と検索範囲決定のルールを示す。また表 3 に空間インデックス有無による処理時間の違いを示す。本手法を採用したことで、ある一都道府県全域という大規模データの一括処理も現実的な処理時間にて行うことが可能になった。例えば表 3 に示したコンピュータを使って 2005 年と 2000 年の東京 23 区全域の住宅地図の入居者データ (2005 年: 2502990 件、2000 年: 件) を時系列化した場合の計算時間は 65874.19 秒 (約 18 時間 18 分) である。

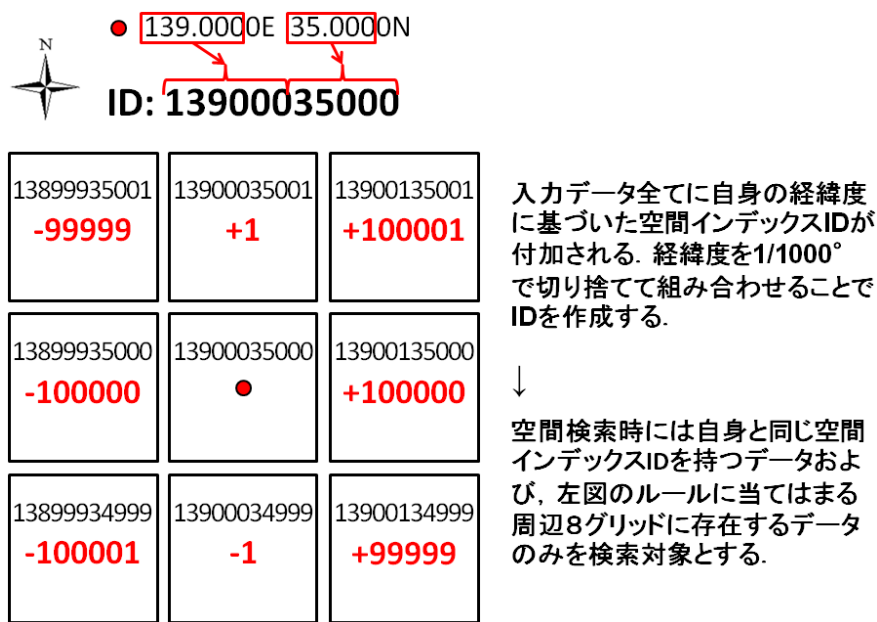


図9 本研究で採用した空間インデックス

表3 空間インデックス有無による処理時間の違い

入力データ	入力データ(新)	タウンページデータベース(2005年)世田谷区全域:28805件
	入力データ(古)	タウンページデータベース(2000年)世田谷区全域:32224件
処理方法	時系列化	
PCスペック	OS	Windows XP Professional Service Pack 3
	CPU	Intel® Core™ 2 Quad CPU Q6700 2.66GHz
	メモリ	3.23GB
計算時間	空間インデックス有り(T1)	891.89秒
	空間インデックス無し(T2)	207552.84秒
	T2/T1	232.71

### (5) 処理精度の検証

南関東地域の幾つかの地域を対象にして、システムにて作成した結果と手作業にて作成した結果を比較する。検証には住宅地図の入居者データと電話帳データを用いる。サンプルの抽出地域を表4に示す。都市部のサンプルとして繁華街と駅前商店街の地域を用いた。また地方都市とデータの分布が粗な地域として離島の集落を用いた。

まず2000年と2005年の南関東全域の住宅地図同士、電話帳同士を本システムによって時系列化する。その後対象地域の全件を地域ごとに手作業で作成した結果と比較する。

表5～表12にシステムの時系列化処理の処理精度を示す。全体としては住宅地図同士の時系列化が95.17%(827/869)、電話帳同士の時系列化が95.62%(480/502)の精度で処理出来ることが分かった。データが高密度かつ三次元的に分布する都市部ほど若干の精度低下が見られるが、一致しなかったデータの中には目視による統合作業においても統合先を判断するのが困難なデータ(候補が複数あるため)や、アドレスマ

表 4 処理精度検証のためのサンプル地域

サンプル地域		件数(2005年)		件数(2000年)	
住所	地域特性	住宅地図	電話帳	住宅地図	電話帳
東京都新宿区歌舞伎町 1丁目17・23番地	大都市の繁華街	191	88	210	94
東京都世田谷区経堂 農大通り商店街	駅前商店街	276	130	247	141
千葉県鴨川市 天津漁港周辺	港町	121	82	136	94
東京都八丈島八丈町 中之郷・櫻立地区	離島の集落	140	105	144	121
合計		728	405	737	450

表 5 (左) 処理精度 (電話帳時系列化・新宿区歌舞伎町の場合)

表 6 (右) 処理精度 (住宅地図時系列化・新宿区歌舞伎町の場合)

	システム					
	総件数	存続	入替	新規	消滅	正答率(%)
手作業	50	49	1	0	0	98.00
入替	23	0	23	0	0	100.00
新規	15	0	2	13	0	86.67
消滅	21	1	1	0	19	90.48
合計	109	うち一致		104	95.41	

	システム					
	総件数	存続	入替	新規	消滅	正答率(%)
手作業	87	79	0	7	0	90.80
入替	74	3	67	4	0	90.54
新規	30	1	1	28	0	93.33
消滅	49	4	1	0	44	89.80
合計	240	うち一致		227	91.16	

表 7 (左) 処理精度 (電話帳時系列化・世田谷区経堂の場合)

表 8 (右) 処理精度 (住宅地図時系列化・世田谷区経堂の場合)

	システム					
	総件数	存続	入替	新規	消滅	正答率(%)
手作業	90	80	2	8	0	88.89
入替	20	1	17	2	0	85.00
新規	20	0	1	19	0	95.00
消滅	31	1	1	0	29	93.55
合計	161	うち一致		145	90.06	

	システム					
	総件数	存続	入替	新規	消滅	正答率(%)
手作業	154	141	2	11	0	91.56
入替	51	0	48	3	0	94.12
新規	71	3	1	67	0	94.37
消滅	42	3	1	0	38	90.48
合計	318	うち一致		294	92.45	

表 9 (左) 処理精度 (電話帳時系列化・天津小湊の場合)

表 10 (右) 処理精度 (住宅地図時系列化・天津小湊の場合)

	システム					
	総件数	存続	入替	新規	消滅	正答率(%)
手作業	70	70	0	0	0	100.00
入替	1	0	1	0	0	100.00
新規	11	0	0	11	0	100.00
消滅	23	0	0	0	23	100.00
合計	105	うち一致		105	100.00	

	システム					
	総件数	存続	入替	新規	消滅	正答率(%)
手作業	116	116	0	0	0	100.00
入替	1	1	0	0	0	0.00
新規	4	0	0	4	0	100.00
消滅	19	0	0	0	19	100.00
合計	140	うち一致		139	99.29	

表 11 (左) 処理精度 (電話帳時系列化・八丈島の場合)

表 12 (右) 処理精度 (住宅地図時系列化・八丈島の場合)

	システム					
	総件数	存続	入替	新規	消滅	正答率(%)
手作業	97	96	0	1	0	98.97
入替	2	0	2	0	0	100.00
新規	6	0	0	6	0	100.00
消滅	22	0	0	0	22	100.00
合計	127	うち一致		126	99.21	

	システム					
	総件数	存続	入替	新規	消滅	正答率(%)
手作業	110	109	0	1	0	99.09
入替	3	0	2	1	0	66.67
新規	27	0	0	27	0	100.00
消滅	31	1	1	0	29	93.55
合計	171	うち一致		167	97.66	

ッチング精度が低いデータが含まれていた。このようなソースデータ側に問題がある理由で適切な統合が期待出来ないデータを除けば、処理精度はさらに向上する。一方地方部では上記した統合時の難しさが殆ど見られないためほぼ 100%に近い処理精度となっている。なお何れの結果もその総件数が表 4 の総件数と一致しない理由は、消滅が出力されるためである(例えば消滅を除いた件数の和は 2005 年の総件数に一致する)。

手作業の場合でもヒューマンエラーが発生する上に、その作業量、作業時間が膨大なものになることを考慮すると本システムの処理精度は十分に実用に堪え得るものであると言える。

#### 1-4 作成データの例

本システムにて作成したデータの例を示す。

図 10 は新宿駅周辺の 2000 年、2005 年間のテナント交代の 3D マップである。電話帳を時系列化したデータを住宅地図に統合することで作成している。テナント交代の様子がフロア単位で観察出来る(実際のデータは部屋単位である)。

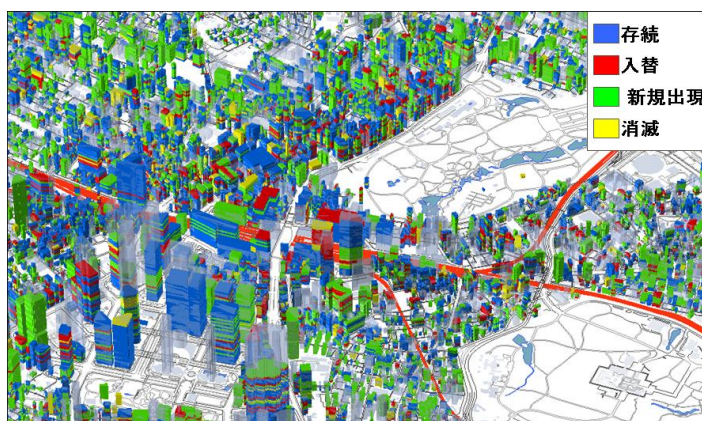


図 10 テナント交代 3D マップ

図 11 は南関東地域全域におけるテナントの存続率を 3 次メッシュ単位で集計した結果である。2000 年と 2005 年の住宅地図を時系列化したデータを元に作成した。

これ以降の商業集積地域の分析で用いる「時系列データ」とはここで紹介した電話帳や住宅地図を時系列化したものである。何れのデータも部屋単位のマイクロなデータであるため、図 10 のようなデータの詳細な分布状況の観察から、図 11 のようなマクロスケールデータの作成まで、幅広いスケールのデータ作成が可能である。

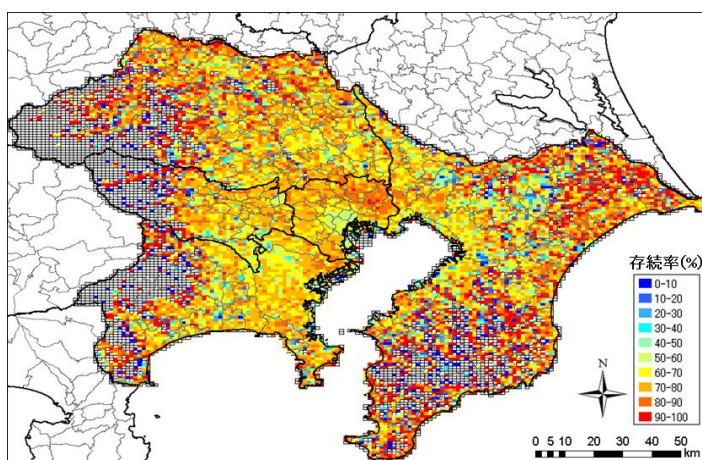


図 11 テナント存続率グリッドマップ

## 1-5 既存統計情報との突き合わせ検証

本研究では主に電話帳の時系列データを用いて商業集積地域の分析を行っていく。しかしその前に確認しておくべきことがある。電話帳データを従来の商店街・商業集積地域の分析に多用されてきた統計情報の代替データとして利用することに問題は無いのか、という点である。そこで従来の研究で多用されてきた事業所・企業統計調査や商業統計メッシュデータ（株式会社 JPS）との突き合わせ検証を行う。事業所・企業統計との突き合わせは時系列データの信頼性の検証であり、商業統計との突き合わせはある一時点の静的な現況データとしての信頼性検証である。まず電話帳の時系列データと事業所・企業統計を突き合わせることで時系列データの信頼性を明らかにする。また次節にて電話帳データを用いて商業集積地域の分布の特定を行っていく。電話帳データから商店街や商業集積地域を構成する店舗を抽出することで商業集積地域を特定していくが、集積の特定には投入データ作成時期の現況統計データとの当てはまりが良いことを示しておかなければ、そのデータを持って商業集積地域を説明することは心許ない。そのため現況統計データとの突き合わせとして商業統計メッシュデータを採用した。

### (1) 事業所・企業統計調査との突き合わせ検証

本項では事業所・企業統計調査と電話帳の時系列データの突き合わせ検証を通して、我々のデータが既存の統計情報の代替データとして使用するに耐えうる性能が有るか検証する。

#### ①事業所・企業統計調査とは

事業所・企業統計調査は統計局が発行する日本全国の事業所や企業の実態とその時系列変化についての情報である。事業所・企業統計は 2~3 年おきに調査が実施され、事業主への調査票配布・回収という形で集計したものである。日本全国約 600 万件のデータが収集される。現在我が国で最も大規模な事業所や企業の実態調査、時系列変遷情報である。行政や民間において産業・経済施策の立案、評価、将来展望等の基礎資料として幅広く利用されている地域統計情報である。

#### ②データ突き合わせの方法

事業所企業統計調査は様々なスケールの集計単位と業種分類体系が存在する。今回我々が使用したのは東京 23 区を区ごとにそのテナントの時系列変化を集計したものである。時系列変化は 2001~2006 年の 2 時点間の変化である。いずれの区も 17 種類の業種分類（大分類）を持つ。業種ごとに存続、新設、廃業の件数を得ることが出来る。表 13 に 17 種類の業種と東京 23 区における時系列変化ごとの収録件数を示す。

一方我々のデータは電話帳の業種情報が利用出来る。電話帳は約 1800 種類の業種



分類情報を持つ。ただしこれらの業種の多くは事業所統計と完全には一致しない。そこで電話帳の業種を全て事業所・企業統計側の業種（大分類）に置換した。事業所・企業統計の業種分類は都道府県単位での集計では小分類と呼ばれる 515 種類にも及ぶ詳細な業種分類を持つ。そこでこの 515 種類の小分類に電話帳の約 1800 種類の業種を対応させ、さらに上位クラスの 17 種類の業種に置換した（図 12）。

また我々のデータの時系列変化情報は事業所統計と異なる。我々のデータが事業所・企業統計と異なる点、そして強みでもある点はテナントの入替が分かることである。そこで図 13 に示すようにデータを分配し直した。以上の処理を東京 23 区全ての区において行った。なお我々のデータは 2000～2005 年の時系列化の結果であるため、事業所企業統計とは新旧年次ともに約 1 年の誤差がある。

表 13 事業所統計全 17 種類の業種と時系列変化の件数（東京 23 区全域の場合）

業種	存続	新設	廃業
農業	116	50	43
林業	8	9	0
漁業	1	2	2
鉱業	32	31	29
建設業	23325	7286	10137
製造業	45972	8528	19749
電気・ガス・熱供給・水道業	124	64	56
情報通信業	8436	11224	9208
運輸業	13318	4820	7288
卸売・小売業	102794	39266	55098
金融・保険業	4565	3528	5062
不動産業	32499	9500	10245
飲食店、宿泊業	55599	26037	31874
医療、福祉	17391	9395	5646
教育、学習支援業	7500	4041	3399
複合サービス事業	821	1224	308
サービス業(他に分類されないもの)	72904	38789	41862

事業所・企業統計調査			電話帳データ		
大分類	中分類	小分類	小分類	業種コード	
不動産業 ←	不動産取引業	建物売買業、土地売買業	住宅販売	3064000	
			不動産取引	6665000	
		不動産代理業・仲介業	分譲住宅	6699000	
			別荘分譲・管理	6701000	
		不動産賃貸業	賃展示場	賃展示場	1557000
				貸ビル	1561000
	住宅展示場		住宅展示場	3063000	
			貸家	1553000	
	貸家業、貸間業		アパート・マンション	0127000	
			学生アパート・マンション	0129000	
	不動産賃貸業・管理業	貸別荘	ウィークリーマンション	0502000	
			貸別荘	1571000	
貸間		貸間	1575000		
		ログハウス	8716000		
駐車場業		駐車場	4637000		
		月極駐車場	4639000		
不動産管理業	駐輪場	4644000			
	ビル管理	6313000			
		マンション管理	7338000		

図 12 事業所・企業統計の業種「不動産」の業種対応表と業種置換

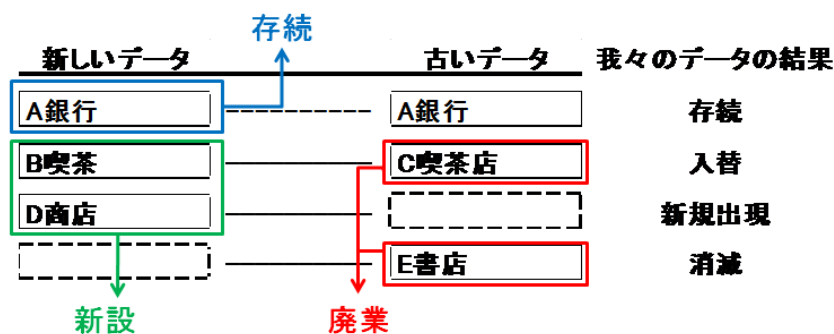


図 13 我々のデータの時系列変化の結果を事業所・企業統計の結果に再分配する方法

### ③結果

図 14 に東京 23 区全域の業種ごとの事業所統計および我々のデータの時系列変化の積み上げグラフを、図 15 に 100%積み上げグラフを示す。農林漁鉱業は件数が極めて少ないため、比較対象としては不適であるが、それ以外の業種については（電気ガス熱供給水道業と複合サービス業を除く）多くの業種でその件数、時系列変化の割合ともに類似した結果が得られている。

同様の集計を東京 23 区全ての区においても実施した。その結果を表 14 に示す。EOS は総件数の誤差 (Error of sum total) であり、我々のデータの件数 / 事業所統計の件数で算出される。ER は誤差率 (Error ratio) であり、事業所統計の結果 (%) に対する、我々のデータの結果 (%) の誤差を意味しており以下の式にて算出される。

$$ER_i = \frac{R_i^O - R_i^C}{R_i^C} \quad (2)$$

$R_i^O$  : 我々のデータの結果 (%)

$R_i^C$  : 事業所統計の結果 (%)

特に卸売小売業や飲食宿泊業、医療福祉では収録件数、時系列変化の割合ともに事業所統計と類似する区が多い。また建設業は時系列変化の割合は多くの区で類似するが、総件数の当てはまりは余り良くない。これは我々のデータが一時的に設置される現場事務所のデータも含むために起こる現象と考えられる。実際に電話帳データには数多くの現場事務所のデータが登録されている。

製造業やサービス業、教育も建設業と類似した現象が見られる。製造業とサービス業では我々のデータ側の業種分類が業態では無く製品名に基づいて分類されている。そのため一部の業種で製造あるいはサービス提供と小売業を混同しているためにおこる現象である。教育は我々のデータが個人経営の小規模な塾まで含むために、統計側よりも件数が多く出てしまうためである。

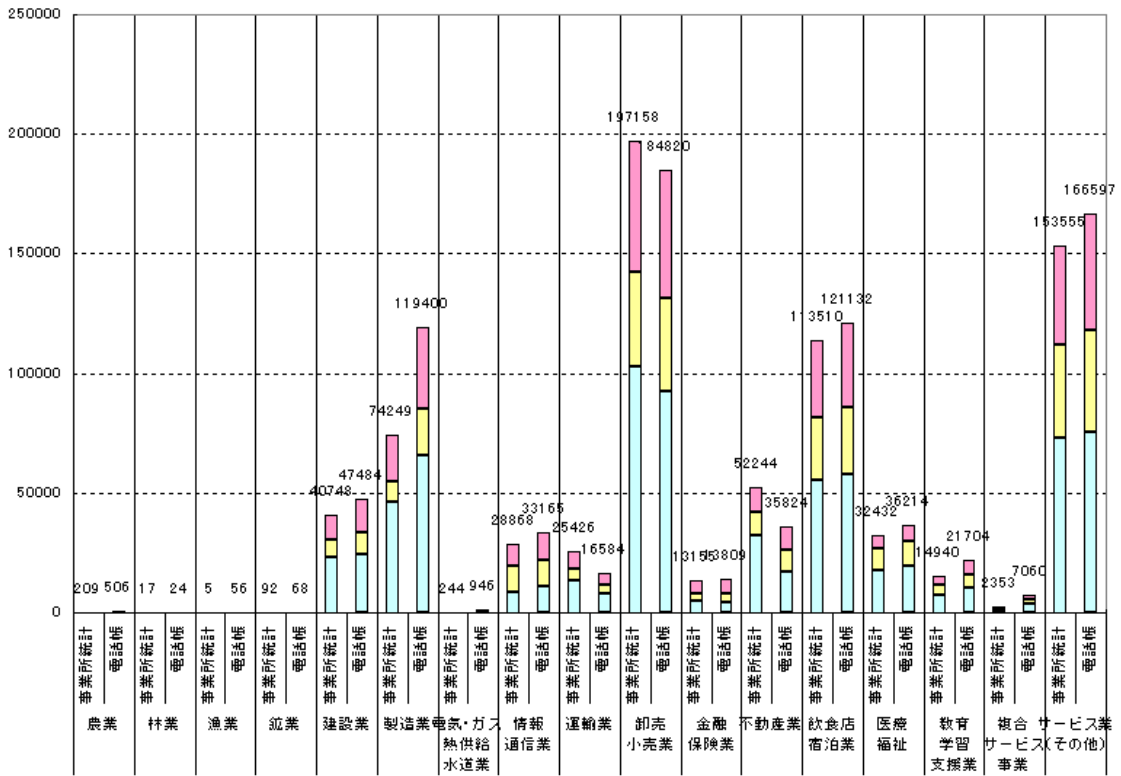


図 14 東京 23 区の集計結果 (積み上げグラフ)

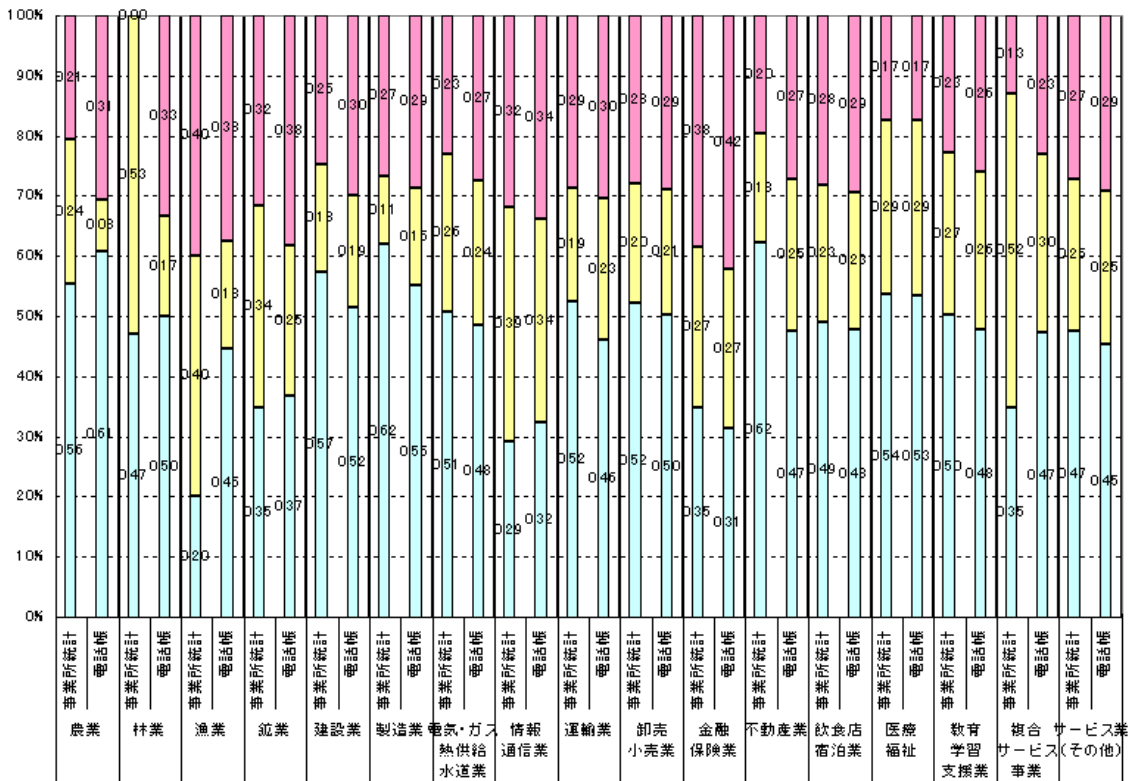


図 15 東京 23 区における集計結果 (100%積み上げグラフ)

表 14 時系列変化ごとの総件数の誤差、割合の誤差率一覧

業種 (サンプル数)	存続		新設		廃業	
	EOS±20%	ER±20%	EOS±20%	ER±20%	EOS±20%	ER±20%
農業(19)	1	7	6	3	4	0
林業(5)	1	2	1	0	0	0
漁業(1)	0	0	0	0	0	0
鉱業(7)	1	2	1	0	0	1
建設業(23)	22	23	9	16	5	14
製造業(23)	4	22	1	9	2	23
電気・ガス・水道・熱供給業(22)	1	4	0	4	1	4
情報通信業(23)	6	20	8	18	4	20
交通業(23)	15	21	8	8	9	19
卸売小売業(23)	20	23	22	21	19	23
金融保険業(23)	22	19	7	14	7	19
不動産業(23)	1	8	11	4	16	2
飲食宿泊業(23)	23	23	17	20	18	21
医療福祉業(23)	19	23	17	20	14	22
教育業(23)	5	23	8	17	4	19
複合サービス業(23)	1	5	7	4	0	3
その他のサービス業(23)	21	23	10	19	7	17

EOS: 総件数の誤差

ER: 誤差率 = (我々のデータの結果) - (事業所統計の結果) / (事業所統計の結果)

農林漁鉱業は東京 23 区における事業所数が非常に少ないため比較対象として利用出来なかった。事業所・企業統計では農林漁業の登録件数が 0 件の区も存在する状況である。このような状況は東京 23 区に限ったものではなく、平成 16 年の全国の農林水産鉱業の総件数を都道府県数で除した値（各件数の後の括弧内の件数は全国の総件数）は農業 301 件（14137 件）、林業 36 件（1675 件）、水産業 25 件（2706 件）、鉱業 70 件（3287 件）であり、このような現象は東京 23 区に限られるものではない。なお我々のデータは事業所・企業統計調査に登録されない個人の農家、林業家、漁師を含むため事業所・企業統計よりもかなり多くの件数が検出される。そういう意味では事業所・企業統計よりも実情を反映した結果と言える。

それ以外の業種についてはあまり良好な結果ではなかった。業種対応表の再検討、我々のデータの時期を事業所統計に合わせての再集計等で改善される可能性がある。

なおやむを得ない理由で良好な結果が得られていない業種も存在する。不動産業の不一致については我々のデータは不動産物件とそのオーナー（大家）を別々に収録している場合があるために、一つの物件に対して 2 重でカウントしていることが影響していると考えられる。また複合サービス業には「郵便局受託業」が含まれる。郵便局受託を受けている事業者は同時に他の業種を兼業している場合が多く、事業所・企業統計側には郵便局受託業と届け出ていても、電話帳側には異なる業種で申告している可能性が高い（そもそも電話帳には「郵便局受託業」という業種は無い）。

以上のように我々のデータは既存の統計データと比較した場合に多少の誤差や限界

はある。しかし多くの業種でその時系列変化の割合や件数が類似した結果となっており、事業所・企業統計調査の代替に耐えうると判断出来る。また我々のデータの強みはユーザーが任意のエリア、年次間のデータ、集計単位を設定して時系列データを作成可能な点である。事業所統計は2～3年おきにしか更新されない上に、時点間も5年に固定されている。この制約を取り払うことが出来たことは極めて有意義である。しかも最小集計単位はテナントスケールである。これは事業所・企業統計調査を他の統計データや空間データと統合して利用したい際の障害であった集計単位の整合という課題を解決出来る。

## (2) 商業統計メッシュデータとの突き合わせ検証

次節で商店街・商業集積地域の抽出と分析のために電話帳データを利用していくが、電話帳データは既存の商業系統計情報と比較した場合、同程度の性能を発揮出来るのか、と言う点を検証しておく必要がある。

そこで商業統計メッシュデータとメッシュ単位での突き合わせを行い、その掲載件数の当てはまりの良さを確認する。電話帳データは2005年、商業統計メッシュデータは2004年のものを使用した。検証対象地域は東京都全域とする。

### ① 商業統計メッシュデータとは

「商業統計」は、全国のすべての卸売業・小売業を対象にした調査統計データである。商業統計調査規則に基づき、商業者の分布と活動状況を販売額の面から明らかにすることを目的とした経済産業省によって実施される統計で、日本標準産業分類の大分類I-卸売・小売業に属する全国の事業所(飲食店を除く)を対象に調査をしている。商業統計メッシュは、メッシュ地図と併用し色分けや、グラフ表示により、エリア特性を視覚的にとらえることができるものである。(以上株式会社JPSホームページより引用：<http://www.jps-net.com/database/statistics/16stm.html>)。メッシュサイズは3次メッシュデータと2分の1地域メッシュ(4次メッシュ)データを持つが、本研究では出来るだけ小地域単位で、商業集積地域を構成すると予想される業種についての状況を把握したいため、4次メッシュデータを採用する。本研究では以下の小売業関連の業種分類と業態分類を利用し、分類ごとの収録件数を電話帳側の収録件数と突き合わせることで、電話帳データの信頼性を明らかにする。

<業種分類>

- ・買回品業種 ・最寄品業種
- ・各種商品小売業(スーパーマーケット・ホームセンター・デパート等)
- ・その他(各種燃料店・自動車二輪車販売店)

<業態分類>

- ・コンビニエンスストア ・スーパーマーケット ・百貨店 ・ドラッグストア

## ②データ突き合わせの方法

まず電話帳データに4次メッシュコードを付加する。電話帳データは経緯度情報を持っているため、経緯度から4次メッシュコードを作成することが可能である。続いて電話帳の業種を商業統計側の業種情報に合うように集計し直す。前項の事業所・企業統計調査に見られた業種置換と同様の手法を用いる。まず前述した業種分類と業態分類に電話帳側の業種を対応付けたマッチングテーブルを作成する。そしてそのマッチングのルールに従い、電話帳データの業種情報を上記の商業統計側の業種・業態に置き換える。最後にグリッドごと、業種ごとにその収録件数を比較し、その当てはまりを検証する。

## ③結果

電話帳データと商業統計メッシュデータの突き合わせ検証の結果を資料1にまとめた。グラフ上のポイントの1つ1つが1グリッド内に存在する電話帳データと商業統計データの収録件数を意味する。

まず業種分類別に見ていく。収録業種全てを含んだ小売店総計、買回品総計、最寄品総計ではいずれも非常に当てはまりが良いことが分かる。一方各種商品小売業の場合では良好な結果は得られない。これは電話帳データの場合、スーパーマーケットやデパート内に入居しているテナントの一部の業種が「スーパーマーケット」「百貨店」に分類されるケースがあるため検出件数が多くなってしまっていることが影響している。その他の業種については良好な結果は得られなかった。

次に業態分類別に見ていく。コンビニエンスストアでは両データの掲載件数間にある程度の相関が見られる。また掲載件数の誤差が±1件以内に収まっているメッシュは全体の83.94%である。このような誤差が生じる理由は、実態はコンビニエンスストアであるが酒屋や雑貨屋を名乗っているケース、あるいはその逆のケースの店舗があるためである。スーパーマーケットはコンビニエンスストアほどはつきりとした相関は見られないものの、±1件以内に収まっているメッシュが全体の85.16%を占める。件数のズレの原因はコンビニエンスストアの場合と同様で、電話帳データでは自己申告であるためか雑貨店等の比較的小規模な店舗でもスーパーマーケットを名乗っていることあるためである。田舎では見栄を張って「スーパー」と名乗っているケースがあるためだと考えられる。例えば八丈島全域を見た場合、商業統計ではスーパーマーケットの件数は0件であるが、電話帳では7件登録されている。百貨店は商業統計側の件数と比べて電話帳側の件数がかなり多くなっている。これは百貨店に入居している店舗の業種が「デパート」になっている場合があることが影響している。またデパートの窓口ごとに電話帳に登録がされている場合もあることも影響している。ドラッグストアについては良好な結果が得られなかった。電話帳の業種分類には「ドラッグストア」が無く、「薬局」「薬店」となっている。これらの業種には商業統計側で指す

量販店的な性質が強いドラッグストアだけではなく、小規模な薬店や調剤薬局も含まれてしまう。しかしながら業種情報だけをキーにして電話帳側のデータからドラッグストアを抽出することは困難である。

以上より小売店総計、買回品総計、最寄品総計等の複数業種の総計では当てはまりが良く、商業統計と類似した性能を持つことが分かる。一方業種を絞ると当てはまりは悪くなるが、業種によってはますます類似した結果が得られる。デパートに見られた入居テナントが名乗る業種の問題や、スーパーマーケットに見られる商業統計側が判断した業種と電話帳に申告された業種のずれといった課題も明らかになった。これは業種の突き合わせ方法の再検討で、より良好な結果が得られる可能性もある。

電話帳データからある特定の業種だけの件数を正確に取得することは業種によっては難しいが、小売店総計、買回品総計、最寄品総計等の複数業種にくくり込んで使えば実用的であることが分かる。次節で行う商業集積地域の抽出ではある特定の業種データの立地が正確であることはあまり重要ではなく、「小売店総計」の構成に近いデータからその集積の程度を明らかにして行くものである。そのため電話帳データは商業集積地域や商店街の分布抽出を行うに適切なデータと判断出来る。

## 2. 商店街・商業集積地域の実態とその変化動向の分析

本節では第一節で作成した時系列データを用いて商店街・商業集積地域の実態とその変化動向の分析を行う。まず電話帳データを用いて商業集積地域を自動抽出する手法を提案する。その後、商業集積地域分布データと時系列データを統合的に利用して商業集積地域の分析を行う。

### 2-1 商業集積地域分布データ作成

#### (1) 本研究における商業集積地域の定義

本研究で扱う「商業集積地域」とは、鉄道駅周辺に構成される駅前商店街、都市の中心部の繁華街等に構成される都心型商店街、鉄道が整備される以前に起源を持つ宿場町型や門前町型の商店街、観光地に構成される商店群など一般的な商店街全般を指す。また近年登場した主要道路沿いのロードサイド型店舗群や大規模ショッピングセンターとその入居テナント群も含む。本研究ではこうした「商業集積地域」の分布を特定する手法が必要である。

そこで本研究ではデジタル電話帳が保有する全件に付加された 1751 種類（2005 年の場合）の業種情報を利用し、商業集積地域を構成する店舗・事業所の洗い出しを行う。業種情報を利用するためには商業集積地域を構成する業種を決定する必要がある。

#### (2) 商業集積地域を構成する店舗の業種決定

商業集積地域を構成する業種の決定にはそれらの地域を構成する店舗の実態情報

を集めるべきである。そこで東京 23 区内に存在する 28 箇所の商店街、商店会のホームページから店舗情報（店舗名称・住所）を入手し、アドレスマッチングにより経緯度情報を与えた。それらを前述した時空間データの統合システムを用いて、デジタル電話帳と統合した。我々が保有する電話帳データは 2005 年 12 月時点のものであるため統合出来ないデータもあったが、ホームページから収集した 3476 件中、1932 件が統合出来た。統合に成功したデータの電話帳の業種より、商業集積地域を構成する業種 401 種類を決定した。各種食料品・衣料品・日用品販売店、美容・衛生、各種飲食店、スポーツ・趣味・娯楽関連施設、医療、不動産、金融、宿泊施設等が含まれる。

### （3）商業集積地域の特定

電話帳データはアドレスマッチングによって得られたポイントデータである。このデータから商業集積地域の範囲を意味するポリゴンを作成し、それらが近隣関係にあるポイント同士を同じ領域に含む集積範囲を意味していれば、それを 1 つの商業集積地域と見なせる。そこで本研究ではポイントデータからバッファリングを行い、重複するバッファポリゴンを統合して 1 つのポリゴンとして連担させていく。このように近隣関係にあるデータの集積からポリゴンを作成するという手法は既存研究でも見られる手法である。

#### ①一般的なバッファリング手法の問題点

バッファリングを行うに当たってバッファリング距離を決定する必要がある。一般的な手法としてバッファリング距離を固定して、全ポイントから同一半径のポリゴンを作成するというものがある。しかし商業集積地域はその地域特性や機能によって店舗間の距離も変化するため、ある固定値によるバッファリングでは適切な結果は得られなかった。図 16 では小田急線経堂駅南部に形成されている農大通り商店街の抽出を試みている。他地域では適切な結果が得られた 50m でバッファリングを行ったが、その抽出地域は現実の商店街の広がり

に比べて余りにも広域である。固定値でのバッファリングでは適切な結果が得られないことが分かる。近隣関係にあるポイントとの連担を考えるのであれば、最近隣ポイントまでの距離をバッファリング距離にすることも考えられる。しかしこの手法の場合、商業集積地域から遠く離れたデータもその近隣データまでの距離をバッファリング距離とし

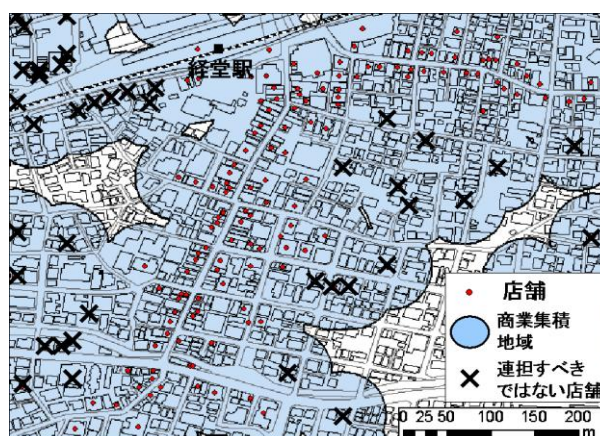


図 16 固定値（50m）でのバッファリング結果



て採用してしまうため、明らかに連担が起らないデータ同士の連担が発生してしまう（図 17）。ある一定の距離以上のバッファポリゴンを用いなければその問題を回避出来るが、ある一定距離を設定することは前述した固定値によるバッファリングで適切な結果が得られなかったことから、地域によって商業集積地域の平均的な店舗間距離の変化が見られることが明らかであり、その影響を考慮すると適切ではないと考えられる。

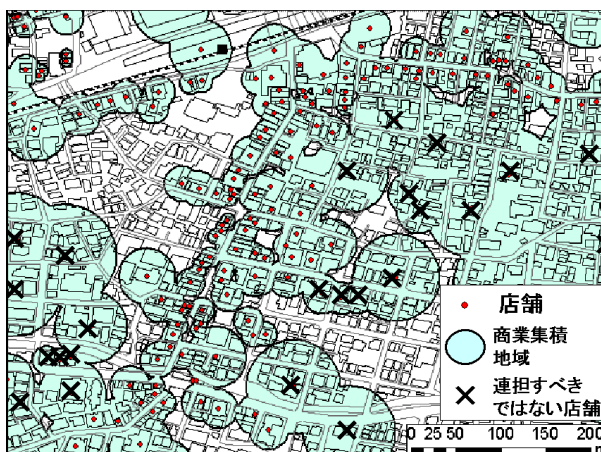


図 17 最近隣データまでの距離を用いた  
バッファリング結果

## ②本研究のバッファリング手法

バッファリング距離をある一定の値に固定した場合は地域特性によって変化する店舗間距離に対応出来ず、最近隣距離では商業集積地域から離れたデータとも連担が起ってしまうことが分かった。

以上の問題点を解決するためには地域特性を考慮出来る、そして商業集積地域から遠いデータには連担が起らない可変的なバッファリング距離を設定すべきである。ここで言う「地域特性」とはある商業集積地域の平均的な店舗間距離のことであり、図 16 のような狭小路沿いの駅前商店街ではその値は小さくなるが、郊外の主要道路沿いの店舗群等では大きくなることが考えられる。

そこで本研究ではバッファリング距離を式 3 で定義する。ポイント  $N$  の最近隣データまでの距離が  $dmin_N$ 、ポイント  $N$  の  $R$  メートル圏内に存在するデータの個数が  $n$ 、それらのデータの最近隣データまでの距離が  $dmin_n$  である時、ポイント  $N$  のバッファリング距離  $D_N$  を

$$D_N = \frac{\sum_{k=1}^n dmin_k + dmin_N}{n+1} \quad (3)$$

とする。

$R$  の値を大きくすることでより広域の地域特性を取得出来るが、広域にし過ぎると、自身が含まれる集積以外の範囲に存在するデータの影響を受ける。あるいは集積に含まれないようなデータが近隣の集積内に含まれるデータの影響を受ける。一方  $R$  の値を小さくし過ぎた場合には、バッファリング距離が非常に小さくなるデータが出現し、

連担すべき集積に分断が発生してしまう場合がある。本研究では様々な R の値を設定して適切な R 値を検証した結果、R=30[m]が適切と判断した。

以上の方法を用いて作成したバッファリングの結果を図 18 に示す。図中の三角形は (2) で収集したデータから作成した農大通り商店街振興組合加盟店の分布である。本手法では同じ商店街を構成する店舗群が同一のポリゴン内に収まっているだけでなく、商店街から遠距離にある店舗との連担も見られない。表 15 は同様のチェックを (2) で収集した商業集積地域全てにおいて行った結果である。

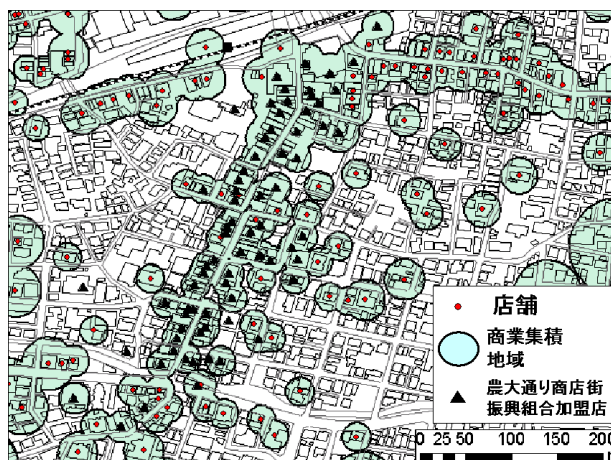


図 18 本研究のバッファリング手法にて抽出した商業集積地域の分布

最後に収録件数が少ないポリゴンを取り除く。商業統計調査表によると 20 店舗以上の店舗集積が見られる地域を「商店街」と定義している。ただし郊外のロードサイド型集積や衰退傾向にある地方の小規模な商店街では 20 件の集積の構成は期待出来ない。そこで本研究では 10 件を閾値として用いることにする。すなわち作成されたポリゴンに含まれる店舗数が 10 件未満の場合、そこは商業集積地域と見なさない。このことは同時に集積内の件数が 20 件を超えている集積だけ抽出することで、商業統計調査表が定義した集積だけを抽出することが出来ることも意味している。

表 15 商業集積地域ごとの連担率一覧

商店街名	タイプ	総件数	連担件数	連担率(%)
上石神井商店街	駅前商店街	169	144	85.21
えぼらまち商店街	駅前商店街	112	110	98.21
フロムナードお花茶屋	駅前商店街	84	84	100.00
三田商店街	ロードサイド・学生街	80	67	83.75
下北沢南口商店街	駅前商店街・ロードサイド	185	183	98.92
九段商店街	駅前商店街・ロードサイド	100	94	94.00
人形町商店街	ロードサイド	176	176	100.00
代々木八幡商店街	駅前商店街	76	75	98.68
浅草国際通り(8商店会)	ロードサイド・観光地	120	120	100.00
小台大通商店街	駅前商店街	81	81	100.00
小岩フラワーロード商店街	駅前商店街・ロードサイド	221	221	100.00
巢鴨地藏通り商店街	観光地	186	182	97.85
戸越銀座商店街	駅前商店街	395	389	98.48
早稲田商店街	ロードサイド・学生街	59	59	100.00
本郷真弓商店街	ロードサイド・学生街	63	63	100.00
永福町商店街	駅前商店街・ロードサイド	149	137	91.95
池尻大橋商店街	駅前商店街・ロードサイド	100	100	100.00
深川仲町通り商店街	ロードサイド	117	93	79.49
王子銀座商店街	ロードサイド	91	50	54.95
田園調布商店街	駅前商店街・ロードサイド	100	95	95.00
神楽坂商店街(2商店会)	ロードサイド	149	147	98.66
神田駅東ロー番街商店街	駅前商店街	61	56	91.80
経堂農大通り商店街	駅前商店街・学生街	174	165	94.83
関原不動商店街	ロードサイド	64	45	70.31
高島平商店街	駅前商店街	64	59	92.19
高輪台商店街	ロードサイド	80	65	81.25
鳩の街通り商店街	ロードサイド	39	39	100.00
鷺宮商店街	駅前商店街	92	83	90.22
合計		3387	3182	93.95

※アドレスマッチング精度の低いデータや商業集積地域から遠く離れているデータは除外

以上の手法によって得た南関東全域における商業集積地域の分布を図 19 に示す。また図 20、図 21 に図 19 のいくつかの地域を拡大したものを示す。東京都、特に 23 区とそれ以外の県でその分布の様子が大きく異なることが分かる。集積の数は東京都 4022 箇所、神奈川県 1879 箇所、千葉県 1134 箇所、埼玉県 1387 箇所であった。図 22 に各都県の集積規模の分布を示す。グラフ中の直線は各都県のデータの回帰直線であり、この傾きが大きいほど高密度な集積が多いことを意味する。

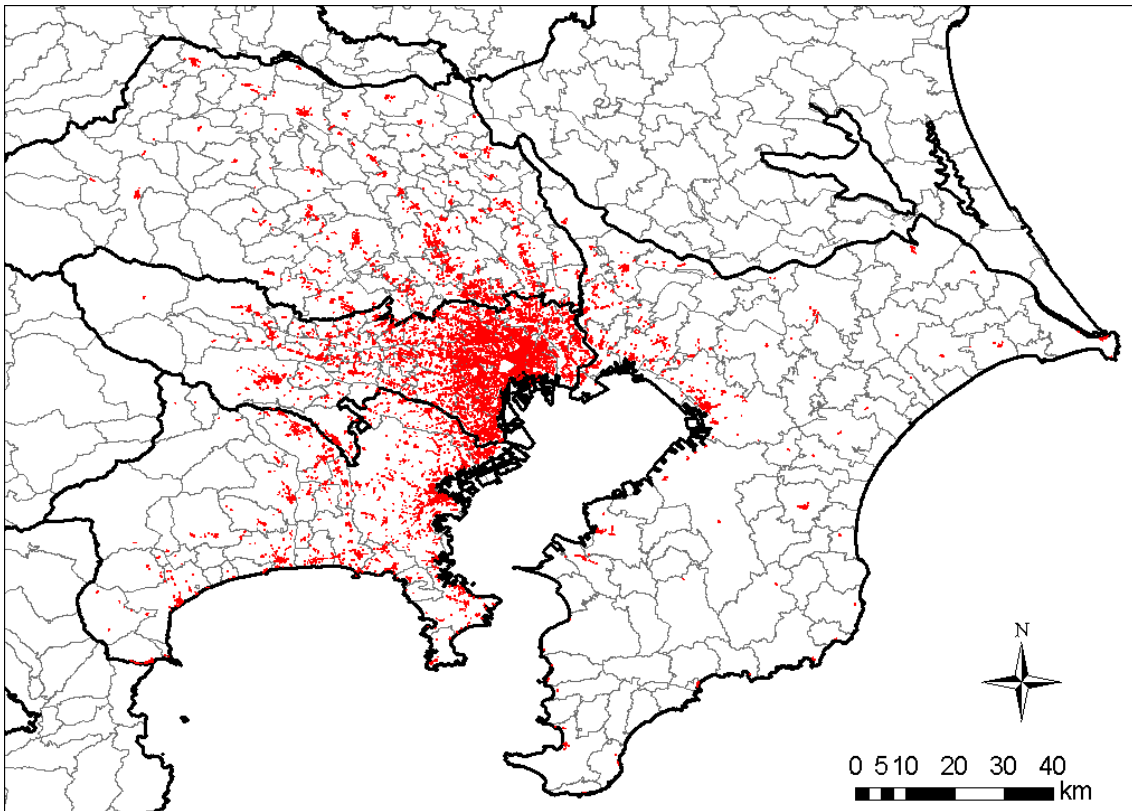


図 19 南関東地域全域における 2005 年の商業集積地域の分布 (赤色のエリア)

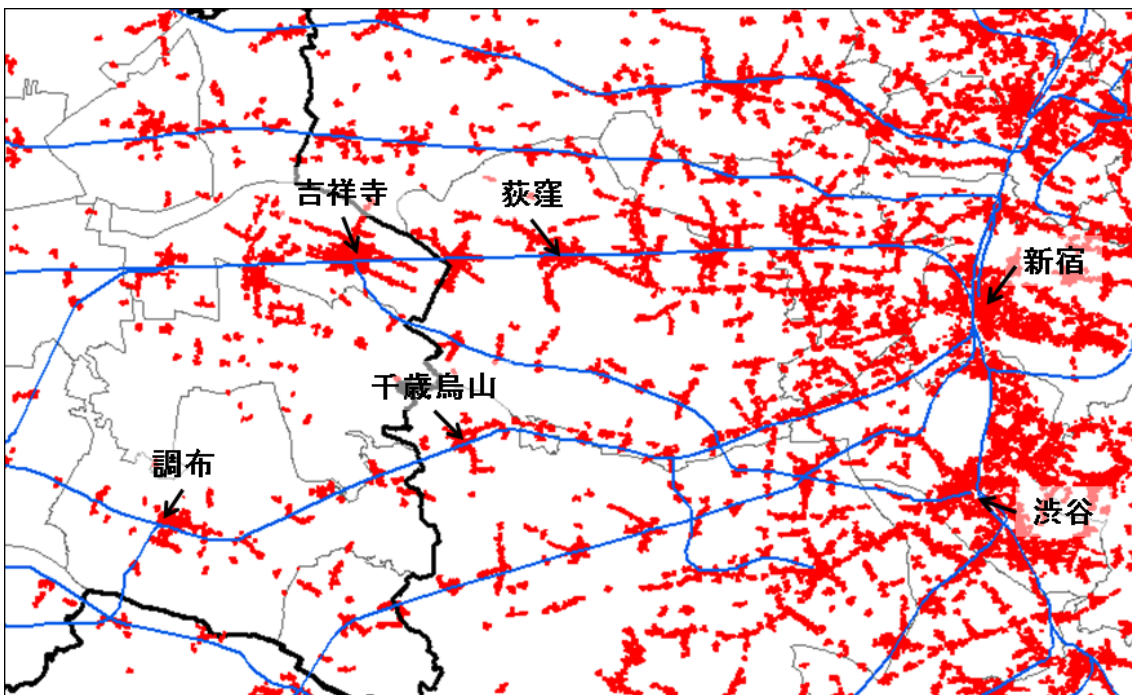


図 20 東京都中西部における商業集積の分布 (図 19 を拡大表示)

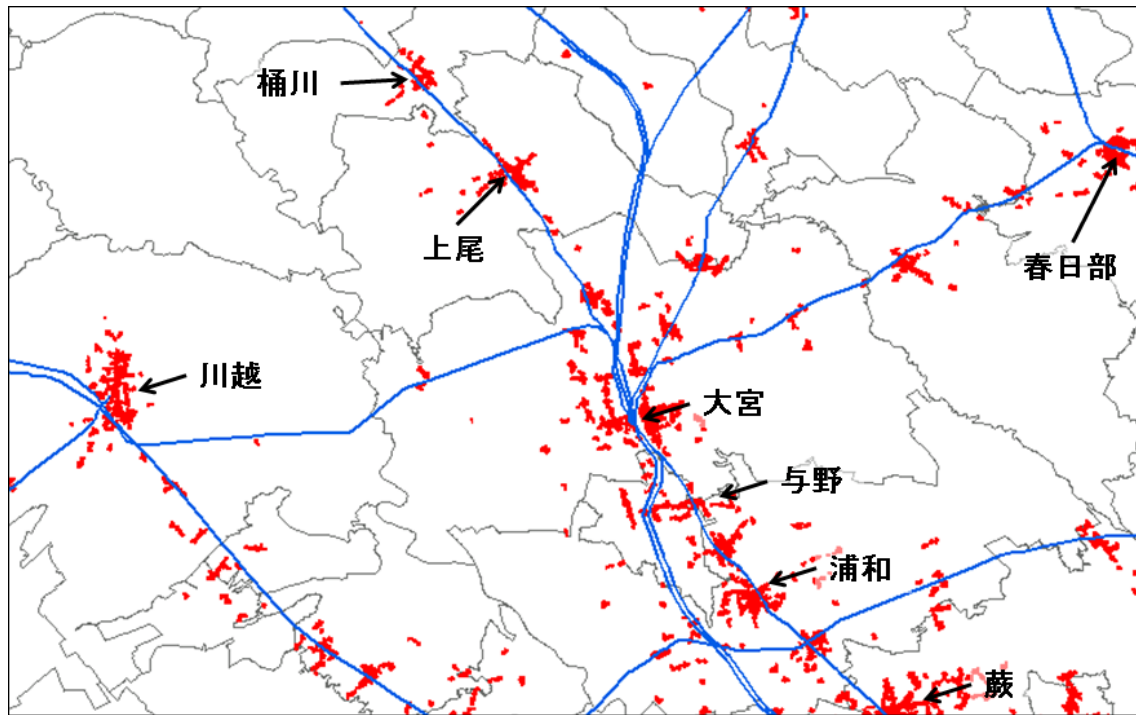


図 21 埼玉県南部における商業集積の分布 (図 19 を拡大表示)

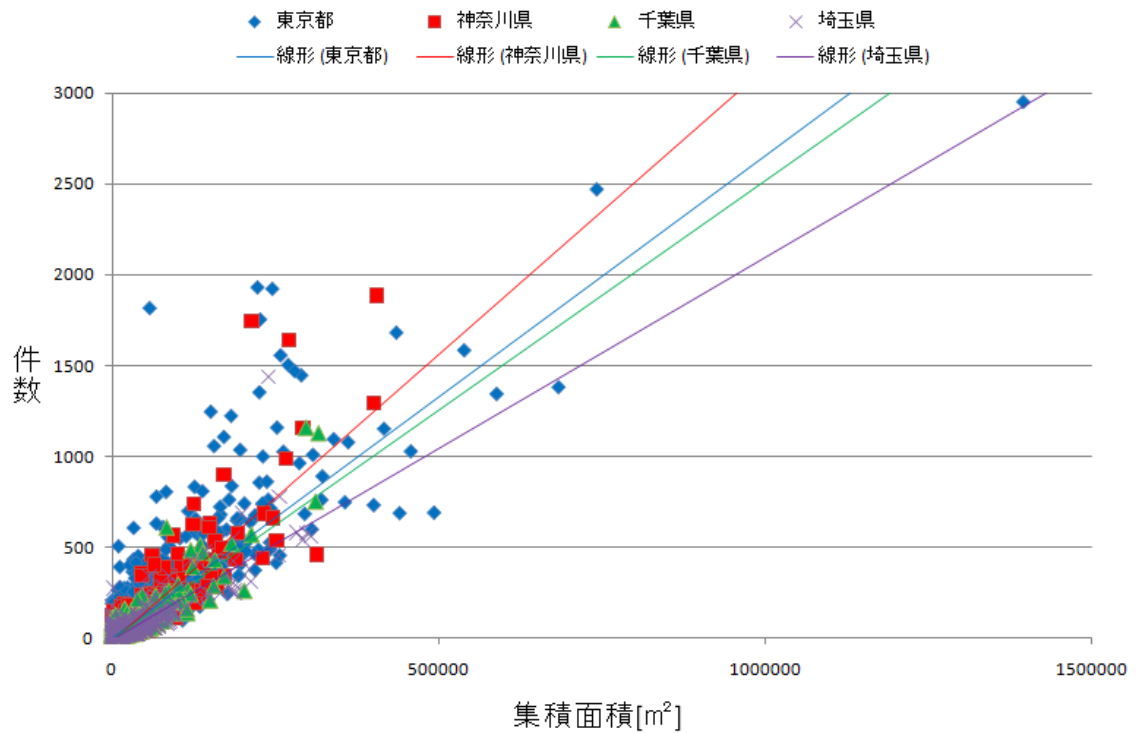


図 22 南関東地域全域における 2005 年の商業集積地域の規模

### ③商店街名鑑との突き合わせによる本手法の妥当性の検証

商業集積地域を抽出する手法を提案したが、この手法の妥当性を明らかにしておく必要がある。そこで本研究では 2004 年の東京都の商店街名鑑（全国商店街振興組合連合会発行）を用いて、東京都全域の商店会が組織されている明らかに商店街が存在すると考えられる地点を収集した。商店街名鑑には商店会の代表組織や代表者の住所と商店会に加盟している店舗の件数が収録されている。この店舗件数を近似的にその商店街の店舗数とする。また住所情報があるためアドレスマッチングを行うことで経緯度情報に変換出来る。すなわち商店会に加盟している店舗数情報を保有するポイントデータを作ることが出来る。

一方我々のデータはポリゴンデータである。そこでこれらのポリゴン内に商店街名鑑のポイントデータが重複して存在した場合に、それを我々のモデルが空間的に一致したと見なせる。2004 年の商店街名鑑ポイントデータと 2005 年の商業集積ポリゴンを統合した結果、商店街名鑑に登録されている 2184 件のポイントデータのうち、1617 件 (74.04%) が商業集積ポリゴンに内包された。図 21 に商業集積データと商店街名鑑の掲載店舗数の相関を示す。なお一つのポリゴンに複数の商店会が含まれた場合にはその総和をとって比較している。相関係数は  $R=0.6643$  とかなり強い正の相関が見られる。我々の商業集積地域の抽出手法が空間的な位置という意味でも、またその規模を捉える意味でも適した手法であることが分かる。なお我々の商業集積地域データと空間的に一致しなかった商店街名鑑が存在する理由については資料 2 を参考にされたい。

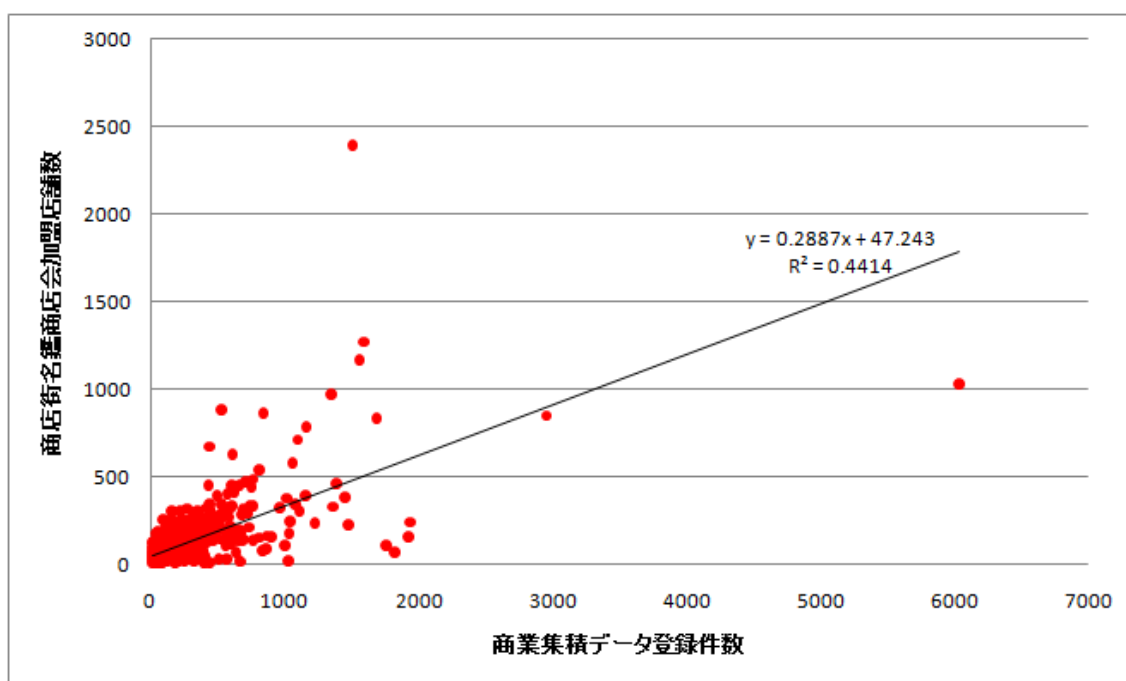


図 23 商業集積データ登録件数と商店街名鑑商店会加盟店舗数の相関関係（東京都）

## 2-2 南関東地域における商業集積地域の実態とその変化動向の分析

### (1) 商業集積地域の分布

本項では南関東地域全域の商業集積地域の分布を観察する。2000年の商業集積の分布と2005年の分布データを空間結合することで、その規模の変化を明らかにする。また消滅、あるいは新規出現した集積の発見も可能である。新規出現あるいは消滅した集積の特徴についても触れる。なお本節の商業集積ポリゴンに与える電話帳データは商業集積データを作る際に用いた業種を厳選したデータでは無く、全データとする。

#### ① 規模の変化

2000年と2005年の商業集積分布ポリゴンを作成し、それらを空間結合することで2005年の分布データと同一地点に2000年に存在した集積の件数を取得することが出来る。その差を得ることで集積の件数変化を得ることが出来る。

図25に集積規模の変化を、図26に都県別の結果を示す。また図27には都県別の集積規模変化の割合を示す。図25の全体的な傾向としてはほぼ全域の集積で店舗・事業所の件数が減少傾向にあることが分かる。一方東京23区の主なターミナル駅周辺の多くではその面積、収録件数がともに増加傾向にある。また23区以外では東京都では吉祥寺駅周辺や立川駅周辺など、神奈川県では横浜市中心部を含む横浜市南区や中区付近、平塚駅や本厚木駅周辺など、千葉県では本八幡駅、船橋駅、柏駅周辺など、埼玉県では大宮駅周辺や熊谷駅周辺などでは集積面積をほとんど拡大すること無く収録件数だけを伸ばしている。この結果から東京都心部以外でも地方の中核都市への一極集中が進んでいることがうかがえる(図24)。また図23に見られる東京都の収録件数が大幅に増加しているデータは、2000年段階に複数の別々の集積だったものが2005年には1つに統合しているためである。一方地方の中核機能を持たない小規模な都市の商業集積の多くが現状維持、あるいは減少に転じている。特に千葉県の房総地方や埼玉県北東部、東京都の多摩地方などで顕著である。

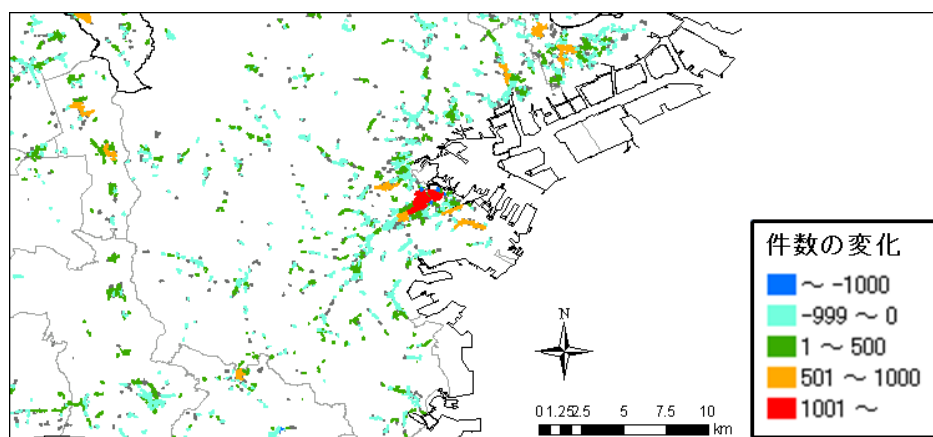


図24 横浜市付近の商業集積の分布とその件数変化。横浜市中心部とその周辺の差が鮮明。

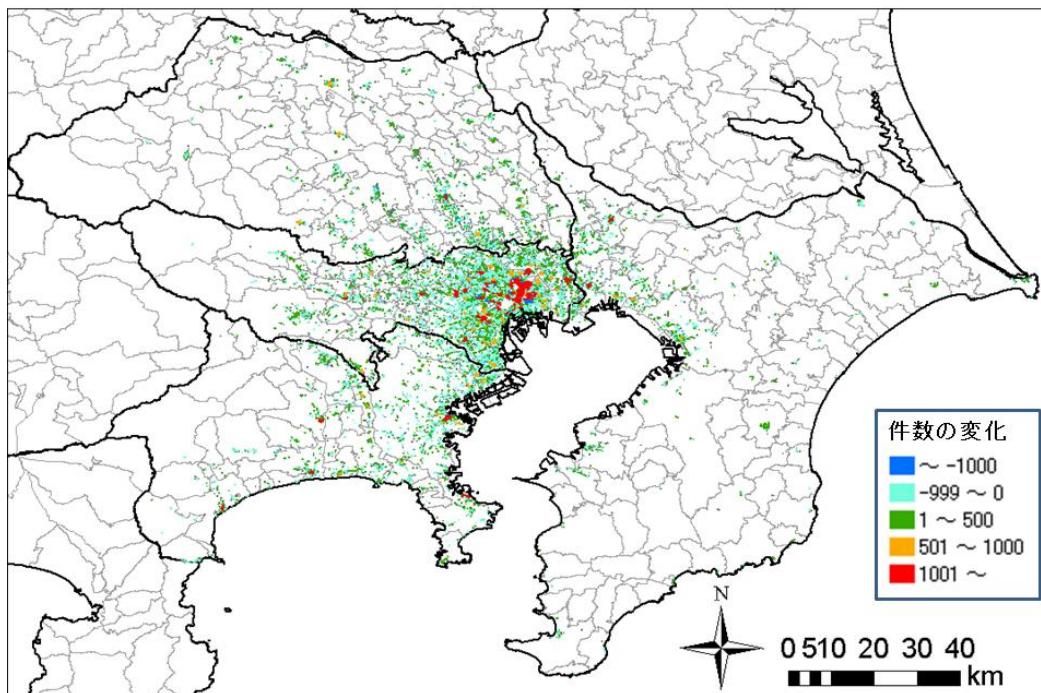


図 25 商業集積の集積規模の変化（2000～2005 年）

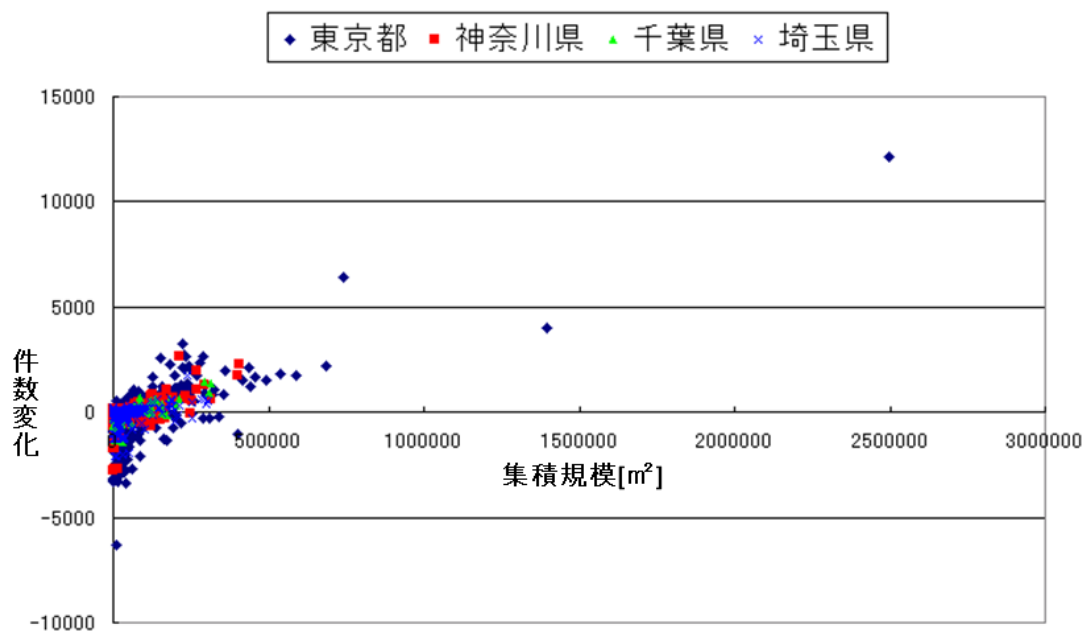


図 26 都県別集積規模と件数変化の関係性

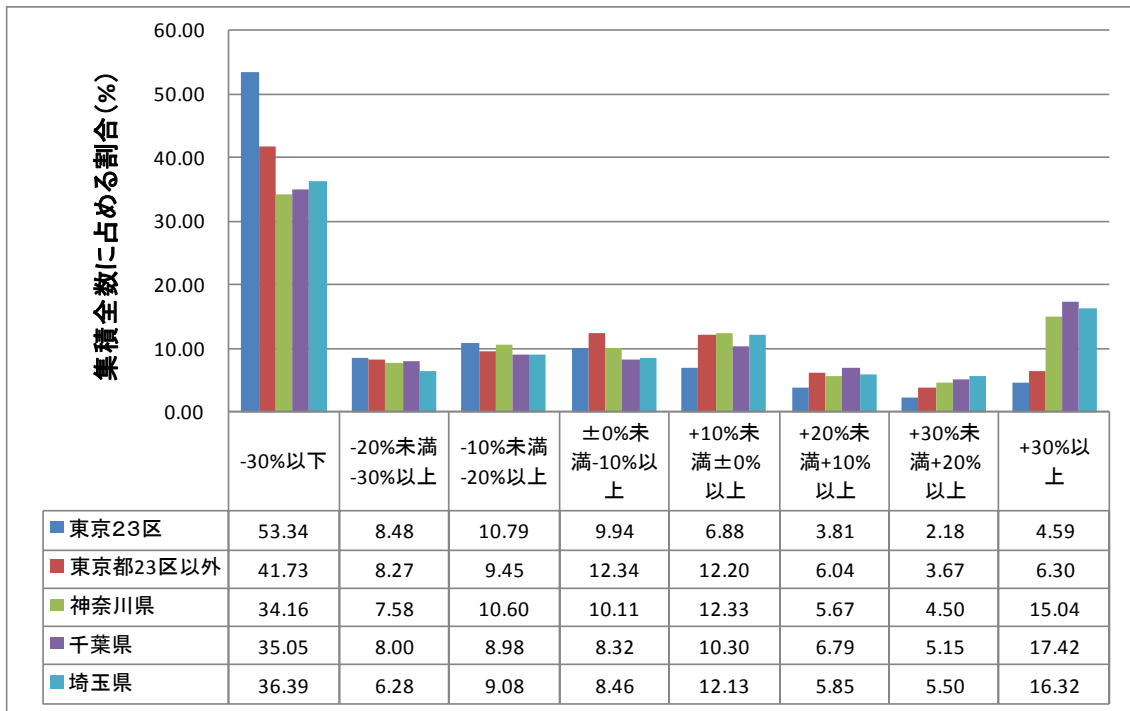


図 27 都県別集積規模変化の様子

## ②新設集積と消滅集積の分布

2005年のデータを2000年と空間結合する際に、統合先が存在しない集積を見つけることが出来る。このような集積は2000年には存在せず、2005年までの間に10件以下の集積から成長、あるいは店舗群が全く存在しなかった場所に新規に出現した集積である。またその逆を見ることで2000年に存在したが2005年には消滅してしまった集積も発見出来る。図28に新規出現した集積と消滅した集積の分布を示す。

東京都や神奈川県では新設集積と消滅集積が混在した分布になっている。一方千葉県の房総地方や埼玉県北部では消滅集積が目立つ。なお新設・消滅の集積もその集積面積、包含する件数ともに全ての集積の平均値よりも小さくなっている。

また興味深い点としてスーパーマーケットや大規模ショッピングセンター、デパート等の大規模商業施設との位置関係が挙げられる（大規模店舗の抽出方法は本項(4)のチェーン店抽出の手法を利用。スーパーマーケット、デパート、総合スーパーに該当するものを抽出した。）。新設、消滅どちらの集積も新規に出店した大規模商業施設の近隣に存在するケースが多い。表16は新規出現した集積が大規模商業施設から1km圏内と2km圏内に何件存在するかを示す。同様に表17には消滅した集積の場合を示す。例えば東京都の場合、2000年から2005年の間に新規に出店した大規模商業施設から1km圏内に新設商業集積366件中、159件が存在し、消滅商業集積771件中、371件が存在する。更に2kmまで広げると新設商業集積は283件、消滅商業集積は575件となる。他の県でも同様の傾向が見られた。とりわけ埼玉県ではこの傾向が鮮明である。



これは大規模商業施設の出店の影響で消滅してしまった集積がある一方で、それに便乗してその周辺に新設店舗が出現し、新たな商業集積を作り出したと予想される。

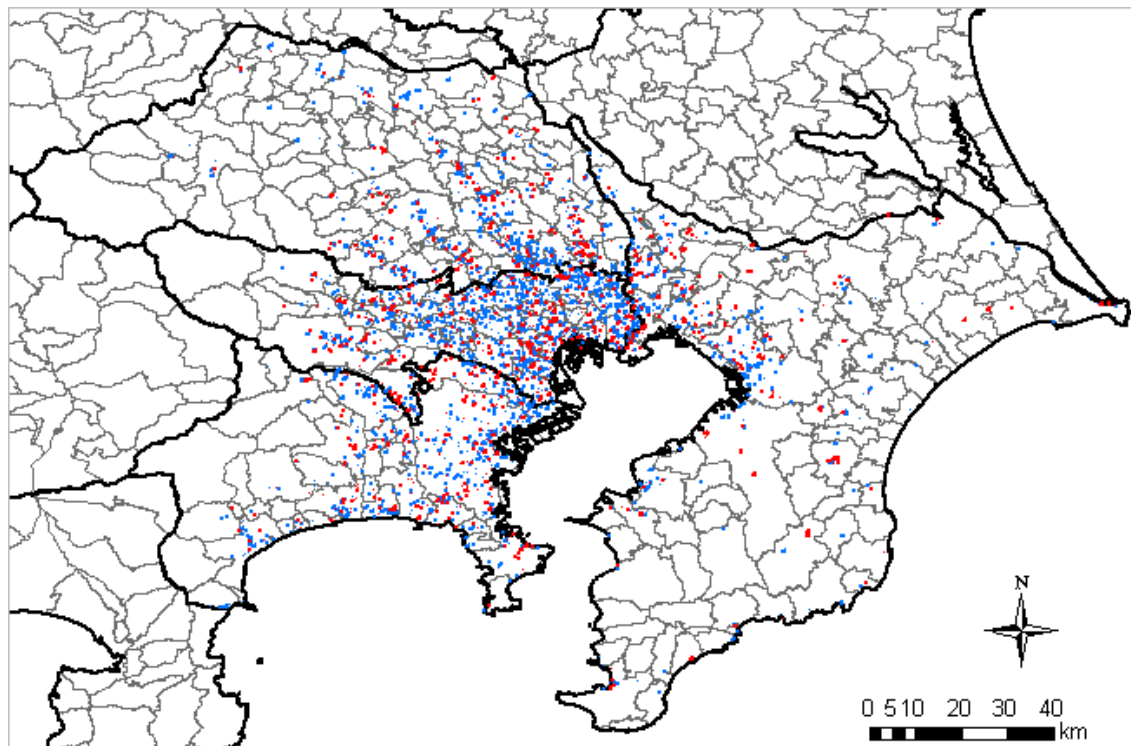


図 28 新設集積と消滅集積の分布 (赤：新設 青：消滅)

表 16 新設集積と大規模商業施設立地の空間的關係

	東京都	神奈川県	千葉県	埼玉県
総件数	366	259	222	242
大規模商業施設から1km圏内	159	85	80	131
割合(%)	43.44	32.82	36.04	54.13
大規模商業施設から2km圏内	283	167	137	204
割合(%)	77.32	64.48	61.71	84.30

表 17 消滅集積と大規模商業施設立地の空間的關係

	東京都	神奈川県	千葉県	埼玉県
総件数	771	466	347	410
大規模商業施設から1km圏内	371	126	103	200
割合(%)	48.12	27.04	29.68	48.78
大規模商業施設から2km圏内	575	277	212	330
割合(%)	74.58	59.44	61.10	80.49

## (2) 集積ごとのテナントの回転率

続いて商業集積ごとのテナントの回転率を観察する。「テナントの回転率」とは、ある一定の期間にどの程度の数のテナントが入替わり、あるいは新規に出現しているか、という意味である。本研究ではテナントの回転率を以下の式で定義する。

$$M_i = \frac{\sum C_i + \sum E_i}{\sum N_i} \quad (4)$$

$M_i$ : 商業集積  $i$  の回転率

$C_i$ : 商業集積  $i$  の入替テナント

$E_i$ : 商業集積  $i$  の新規出現テナント

$N_i$ : 商業集積  $i$  のテナント

店舗・事業所等の入れ替わりや新規出現の情報は1節で開発した空間データの時系列化システムを用いれば良い。ここでは各集積の回転率を求めるために商店街・商業集積を構成する業種だけや、電話帳に掲載されているデータだけに限るべきではない。商業活動が活発な商業集積ではそれに付随する各種産業に関連するテナントの動きや、マンション・アパート等の不動産物件の回転率も高まることが予想される。そこで現実空間の動向をより網羅的にカバー出来る住宅地図の入居者データを使ってテナントの回転率を観察する。なお既に説明したが住宅地図の入居者データの掲載件数は電話帳データの2倍以上であり、実空間の網羅性という点で明らかに優れている。またマンションやアパートの入居者データも可能な限り収集しているため、不動産物件の回転率も考慮に入れることが出来るものと判断した。

2000年と2005年の南関東全域の住宅地図を時系列化し、それらを商業集積ポリゴンデータと統合する。その後、商業集積地域ごとに4式を適用し、全ての集積のテナントの回転率を求める。図29にその結果を示す。また都県別のデータ分布を図30に示す。図29より回転率の比較的低い集積が東京都心部を囲むようにドーナツ状に形成されていることが分かる。ドーナツ状の低回転率地帯の外側では東京都心部ほどではないが、比較的回転率の高い集積が点在するようになる。また東京都心を中心に見られるドーナツ状の低回転率地帯は小規模ながらも千葉市や横浜市の周辺にも構成される。都県別に見ると埼玉県に特異な現象が見られる。他の都県に比べて集積規模と回転率の正の相関が強い。これはさいたま新都心の再開発に伴う変化だと考えられる。さいたま新都心の再開発による影響か、さいたま市中心部付近では横浜市や千葉市に見られるドーナツ地帯も見られない。

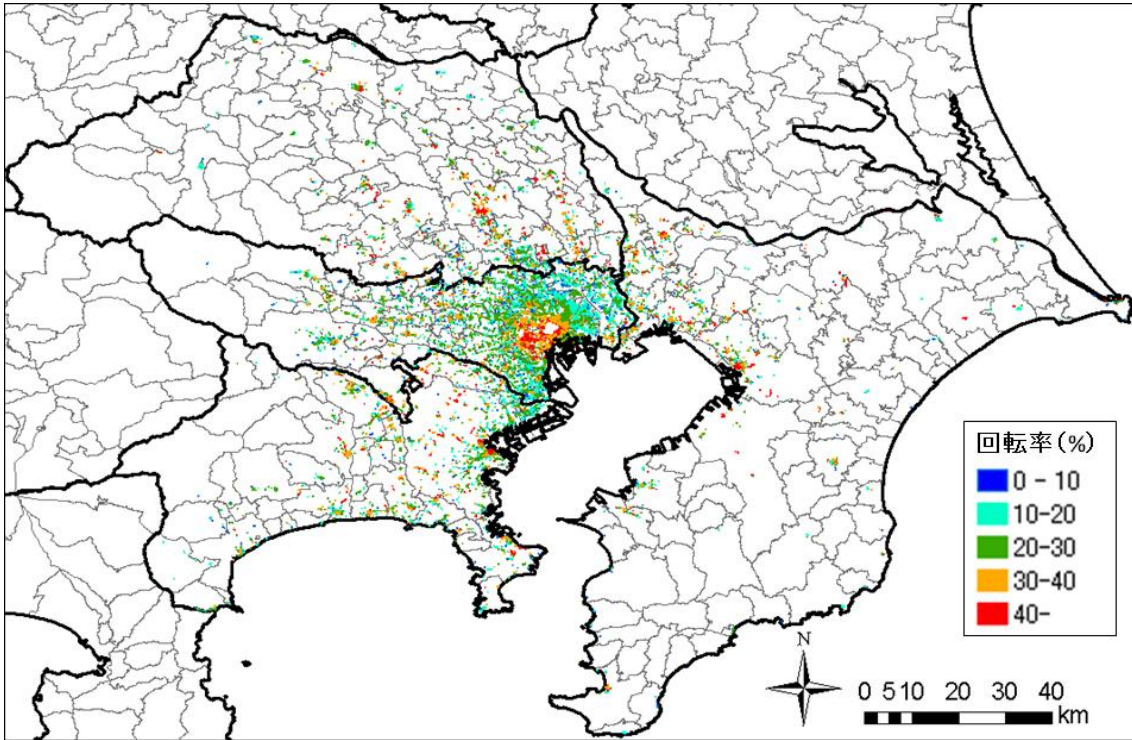


図 29 テナントの回転率

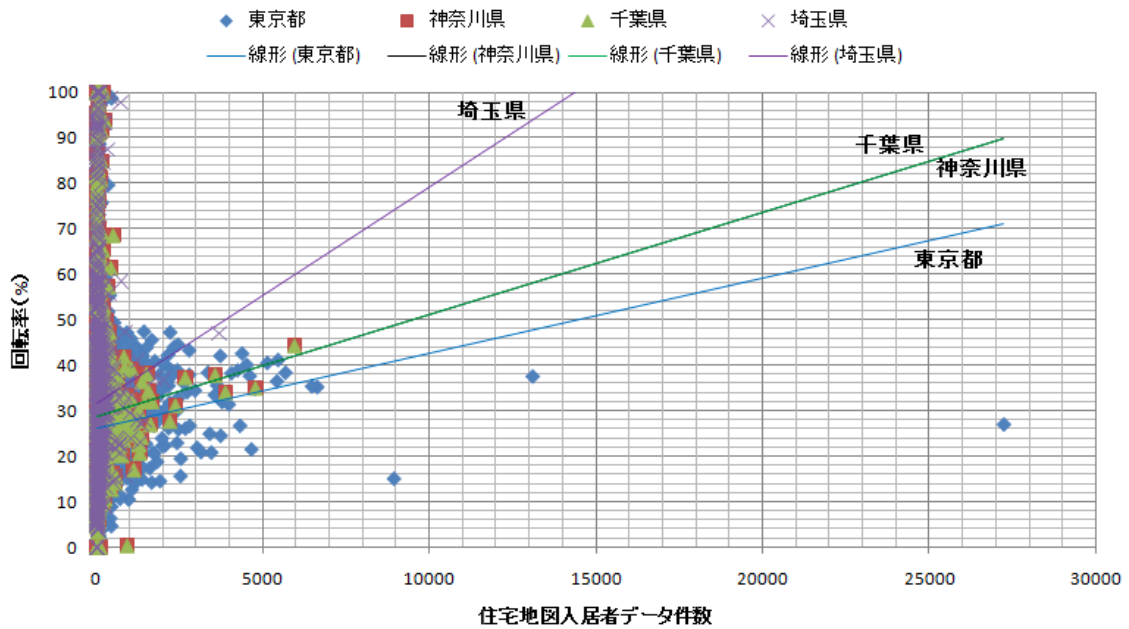


図 30 都県別回転率と近似曲線（線形近似）

### (3) 生鮮三品の充実度

利用しやすい商店街の指標の一つとして生鮮三品を扱う店舗が揃っているか否か、という点である。生鮮三品とは魚、肉、野菜を意味しており、その充実度とはそれらを扱う魚屋、肉屋、青果店が商店街の中に存在しているか、という意味である。生鮮三品は日々の食生活に欠かせないものであり、かつての商店街には必ず存在していた業種である。しかしながら近年ではスーパーマーケット等の大規模商業施設が出現した影響で廃業、あるいは業種転換する魚屋、肉屋、青果店が増えており、それがますますその商店街の魅力と利便性を低下させる一因となっているとも言われている。

そこでここでは電話帳の業種情報を用いて商業集積地域ごとの生鮮三品の充実度を明らかにする。各集積に魚屋、肉屋、青果店全てが最低でも1件は存在する集積の数とその分布を観察する。また同様に2000年には生鮮三品が揃っていたが、2005年にはそれらの何れかが欠けてしまった集積を発見する。電話帳の時系列データを用いて、全ての集積の新旧の業種を集計し、2000年には生鮮三品が揃っており、かつ2005年にはそれらの何れかが欠けている集積を抽出すれば良い。図31は南関東地域で生鮮三品業種全てが最低でも1件は存在する集積の分布と、かつては生鮮三品が揃っていたが、2005年時点ではそれが崩壊してしまった集積の分布を示す。

生鮮三品が揃っている集積は東京都398箇所（/4022箇所 9.90%）、神奈川県198箇所（/1879箇所 10.54%）、千葉県75箇所（/1134箇所 6.61%）、埼玉県77箇所（/1387箇所 5.55%）が検出された。一方生鮮三品が揃っていた状態が崩壊した集積は東京都270箇所、神奈川県100箇所、千葉県46箇所、埼玉県57箇所が発見出来た。表18には都県別の生鮮三品業種が揃っている集積とそれが崩壊した集積の件数と割合を示す。生鮮三品が揃っている集積は東京都と神奈川県は比較的高い値を示すが、千葉県と埼玉県では他2都県の半分ほどの割合しか存在しない。これは埼玉県や千葉県で大規模商業施設の進出が盛んなことや、それによる影響を受けた集積が多いためと考えられる。一方生鮮三品の充実が崩壊した集積の件数の割合は概ね同程度であるが、東京都では特に23区でその件数が多くなっている。これは千葉市や横浜市の中心部でも同様に見られる傾向である。一方千葉県や埼玉県ではこの値は若干小さい。東京都心部では「生活のための商店街」という機能を捨てて、特定業種が偏在した専門店街や遠方からの集客力が大きい買回品中心の業種構造にシフトしつつあるものと考えられる。一方千葉県や埼玉県では、2000年の段階で既にかつては生鮮三品が揃っていた商業集積の淘汰が進んでいたものと考えられる。

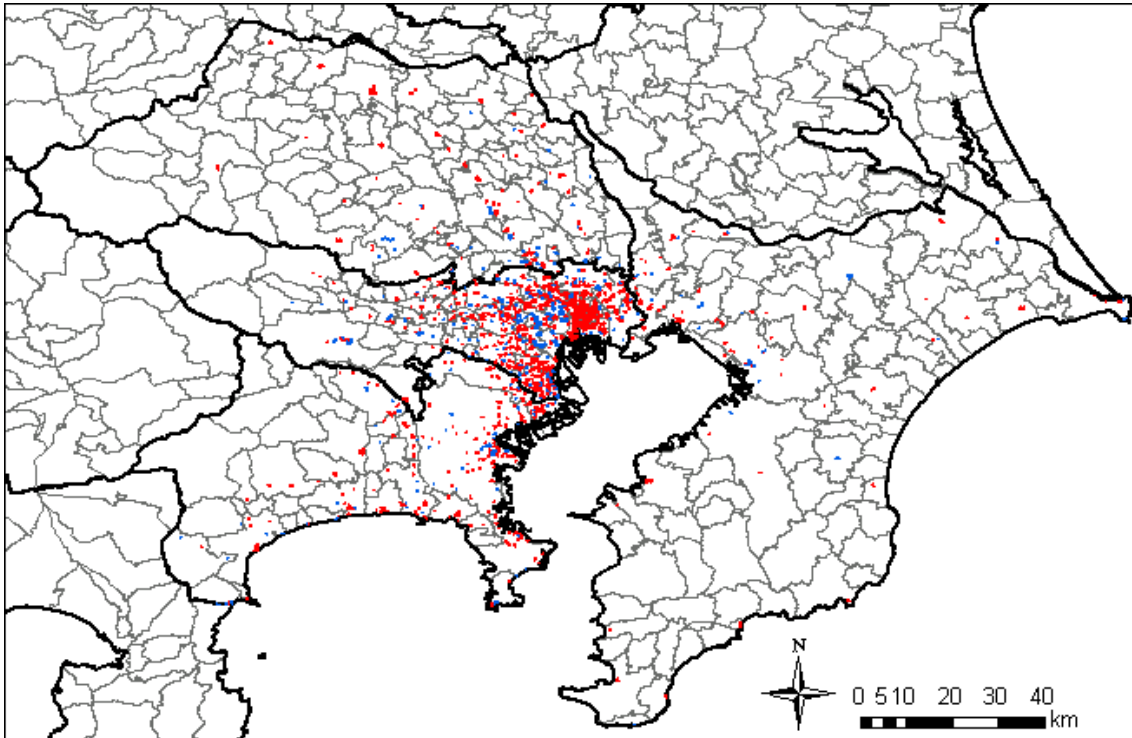


図 31 生鮮三品業種が全て揃っている商業集積の分布（赤色）と 2000 年には生鮮三品が全て揃っていたが現在は何れかが欠けてしまった商業集積の分布（青色）

表 18 生鮮三品業種が揃っている集積とそれが崩壊した集積の都県別割合

	東京都		神奈川県	千葉県	埼玉県
	23区	23区外			
集積総数(2005年)	3135	887	1879	1134	1387
生鮮三品が全て揃う集積	325	73	198	75	77
割合(%)	10.37	8.23	10.54	6.61	5.55
かつては全て揃っていた集積	220	50	100	46	57
割合(%)	7.02	5.64	5.32	4.06	4.11

#### (4) チェーン店進出状況

近年空きテナントへのチェーン店の進出が著しい。チェーン店の進出は空きテナントの減少につながるという点では歓迎すべきことであるが、チェーン店が増え過ぎた商店街は無個性化が進み、結果的に本来持っていた魅力を失う恐れがある。また平成 19 年の商店街実態調査分析（全国商店街振興組合連合会）によると商店会費を全く収めていないチェーン店は商店街内にある全チェーン店の 14.6%にのぼる。すなわち過剰なチェーン店の進出は商店会の運営にも影響を与えかねない。

そこでここでは商業集積地域ごとにチェーン店が何件あるかを明らかにし、総件数で除すことで集積ごとのチェーン店率を明らかにする。チェーン店の抽出のための簡

単な手法は電話帳データが持つ業種情報を利用することである。しかし商業統計との突き合わせの際に紹介したように、電話帳データの業種は事業主の自己申告である為その情報だけからチェーン店を抽出するのは困難である。より確実なチェーン店抽出手法が必要である。

そこで本研究では電話帳に掲載された店舗名を利用した。まず店舗名内にチェーン店であることを示す文字列（例：セブンイレブン、吉野家など）を含むものを抽出する。チェーン店の名称一覧データはテレポイントデータ（株式会社ゼンリン）がもつ業種情報を利用した。テレポイントデータは主なチェーン店の業種はその店舗名で登録されている（例えば「ローソン下北沢店」なら「コンビニエンスストア」ではなく「ローソン」という業種で登録される）。まずここから主なチェーン店の名前を集めた。コンビニエンスストア、スーパーマーケット、デパート、大規模ショッピングセンター、ドラッグストア、レストラン等が含まれる。またこれだけの情報では心もとないため、全国フランチャイズ加盟店協会ホームページ（<http://www.fcjapan.gr.jp/>）や日本フランチャイズチェーン協会ホームページ（<http://jfa.jfa-fc.or.jp/>）等から上記の手法だけで見つけられなかったチェーン店を追加で収集するとともに、それらの文字列を南関東地域全域の電話帳データの名称情報と照らし合わせて行くことで、その出現パターンをライブラリ化した。その結果、その殆どが「チェーン店舗名」＋「語尾が店」で構成されていることが分かった（資料3参照）。また語尾の「店」の前に地名や駅名、大規模商業施設名がくるケースが多いことも明らかになった。

以上より上記のライブラリを使って抽出した結果に加えて、データ時系列化のシステムで利用している地名と駅名のライブラリを利用して、語尾が「地名か駅名＋店」で表記される店舗を抽出する。図32は集積ごとのチェーン店率である。都市の中心部ほどチェーン店率は低く、郊外に向かうにつれてその値が大きくなっていることがよく分かる結果となった。特に東京都では東京都心部や八王子駅・吉祥寺駅周辺、神奈川県では横浜市や川崎市の中心部、藤沢駅や小田原駅周辺、千葉県では千葉市中心部や銚子市中心部、船橋駅周辺、埼玉県では川越市や熊谷市中心部が低い値を示した。

都県別に見てみると特に千葉県や埼玉県でチェーン店率の高い集積が多くなっていることが分かる（図33、図34）。一方東京都は他の県と比べてチェーン店率の低い集積が特に多くなっていることが分かる。またチェーン店率が50%以上の非常に高い値を示す集積の割合は東京都1.34%に対し、神奈川県2.50%、千葉県2.03%、埼玉県2.02%となっている。このような集積はそのほぼ全てが大規模商業施設や大規模なオフィスビル、あるいはそこに入居するテナントである。例えば東京都の場合、チェーン店率50%以上の集積に含まれる電話帳データは631件あるが、住宅地図データと照らした結果そのうちの518件は大規模なビルやショッピングセンター、デパート等のデータかそこに入居している店舗のデータであった。この結果からも東京都以外の県で大規模店舗の出現が積極的に行われていることが分かる。

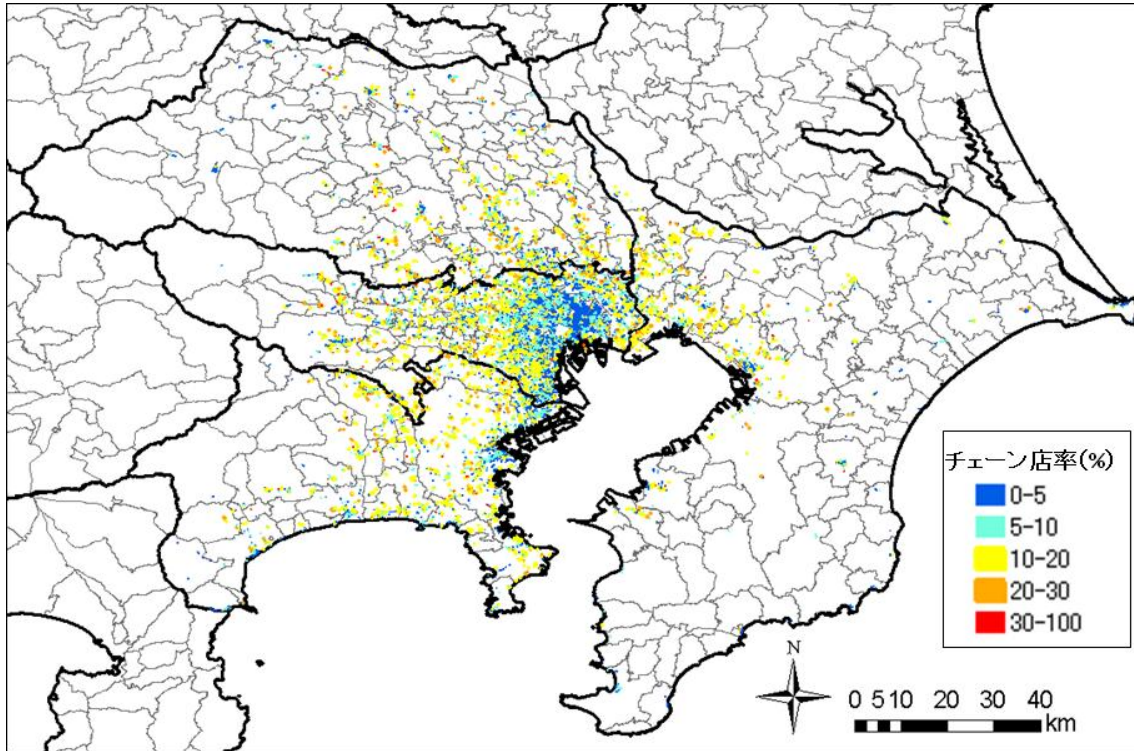


図 32 商業集積ごとのチェーン店率

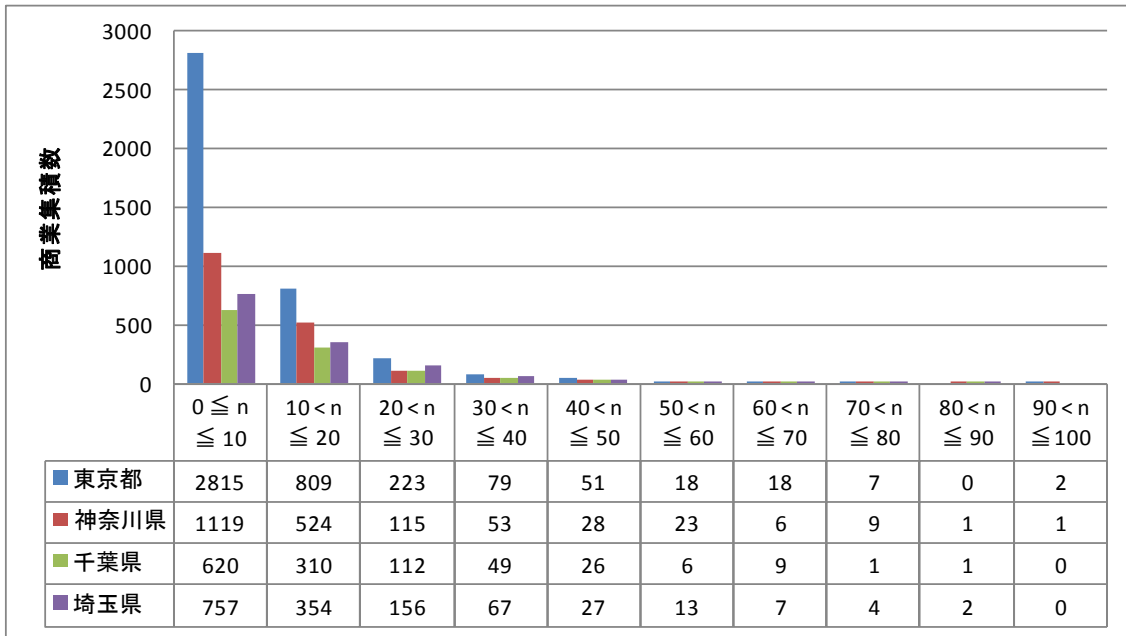


図 33 チェーン店率別集積の件数（都県別）

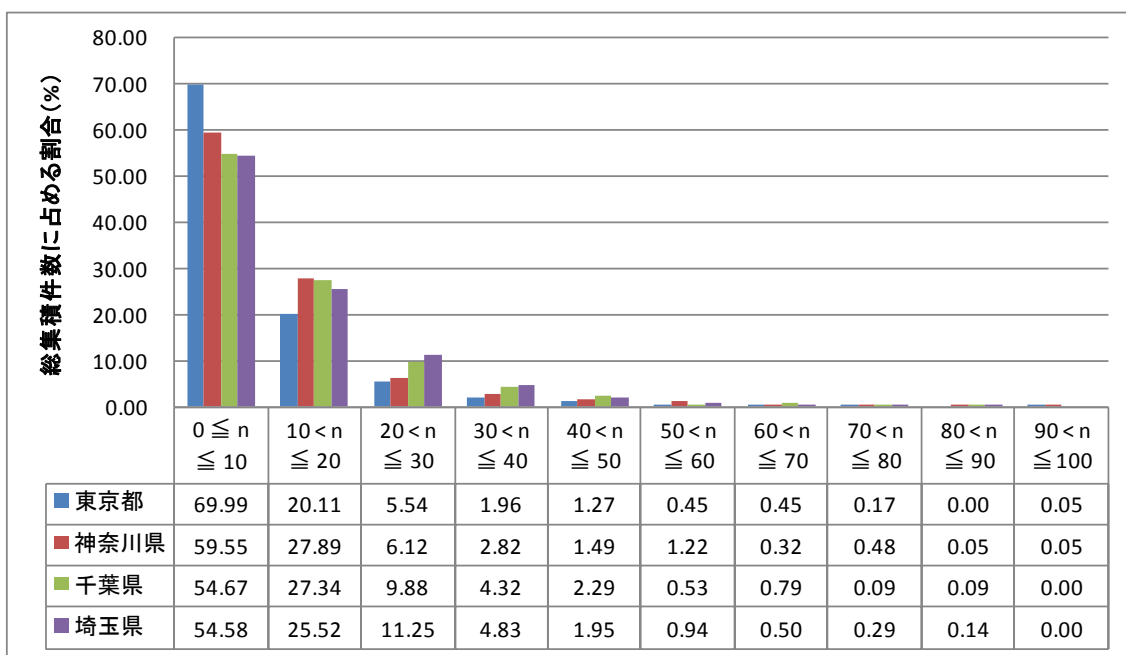


図 34 チェーン店舗率別の総集積件数に占める割合（都県別）

### (5) まとめ

南関東地域における商業集積地域の現状と変化を以下の点に注目して分析した。

- ・集積の分布とその盛衰と規模の変化
- ・集積ごとのテナントの回転率
- ・生鮮三品の充実度
- ・チェーン店進出状況

集積規模は特に東京都心部でその規模、件数ともに増加傾向であった。東京都心部では以前は複数の別々の集積であったものが統合して巨大な集積に成長するケースが多く見られた。東京都心部以外でも地方の中心都市や県庁所在地都市の中心部では集積規模はそれほど大きくなってはいないが、件数が増加しているケースが多く、集積内の店舗・事業所密度が高まっていることがうかがえる。一方で東京都区部や東京都以外の県の中心都市以外では規模も件数も減少傾向であった。

テナントの回転率では特徴的な分布が見られた。東京都心や地方中心都市の中心部では非常に高い回転率となる。その近傍の周辺地域では回転率が急激に小さくなり、さらにその外では都心部と近傍周辺地域の間くらいの値をとるという構造が見られる。大都市周辺部にドーナツ状の回転率が低い地域が出現するという構造である。大都市中心部での回転率の高さはそれらの地域の不動産価値の高さや商業活動における競争率の高さに起因すると思われる。一方前述したドーナツ状地域の外側でもまずまずの



回転率の高さが見られるのは、チェーン店の進出に起因するものと予想される。

生鮮三品業種の充実度では生鮮三品業種が全て揃っている集積を抽出した。その結果、東京都心部周辺や東京都市部に多くの該当集積が見られた。また東京都と神奈川県では全集積の約 10%が該当するのに対し、千葉県や埼玉県では 6%前後と低い値を示す事も分かった。千葉県や埼玉県では東京都や神奈川県よりも郊外化が進展しているためだと考えられる。また 2000 年は三品業種が揃っていたが 2005 年にはそれが崩壊した集積の分布を見ると、大都市の中心部にそうした集積が集まっていることが分かった。生活のための商店街という機能を捨てて、特定業種が偏在した専門店街や遠方からの集客力が大きい買回品中心の業種構造、あるいは観光客向け店舗群という機能にシフトしたものと考えられる。

チェーン店の進出状況は東京都心や地方中心都市の中心部ほどチェーン店率は低く、郊外に向かうに連れてその値が大きくなることが分かった。とりわけ千葉県や埼玉県ではチェーン店率の高い集積が多くなっており、郊外化が進んでいることがよく分かる結果となった。

以上の結果を表 19 に簡略化してまとめた。東京都心部や地方中心都市の中心部では生鮮三品の充実度が低い集積が多く、生活のための商店街では無くなりつつあることが分かる。大都市の周辺部(上述のドーナツ状地域)では集積規模が安定しており、生鮮三品の充実度も高く、テナントの回転率も低い。これらの地域では日々の生活に密着した商店街が機能していると言えよう。郊外では全体的に規模も件数も縮小傾向である。また東京都市部以外では生鮮三品業種も不足しておりチェーン店が増加し続けている。明らかな郊外化傾向が確認出来る。

表 19 南関東地域における商業集積地域の特徴のまとめ

地域	商業集積		テナント 回転率	生鮮三品 充実度	チェーン店 の割合
	規模	件数			
東京都心部	拡大	増加	高い	低い	低い
地方中心都市中心部	微増	増加	高い	低い	低い
大都市周辺部	維持	維持	低い	高い	やや低い
東京都市部	縮小	減少	中程度	やや高い	やや高い
その他の地域	縮小	減少	やや高い	低い	高い

### 3. 結論

#### (1) 結論

本研究ではまずテナントの変遷を広域に渡って詳細に観察出来るデータセットの開発技術を実現した。経緯度のみならず住所や建物情報といった非定量的な位置情報の統合および、曖昧さを持つテキスト同士の同一性判定技術が提案され実用化された。作成したデータについては既存の統計情報との突合せを行い、多くの地域と業種において良好な結果が得られることが分かった。

また電話帳データの業種情報及び、それらの店舗間距離を用いることで商業集積地域を特定することが可能になった。商業集積地域の抽出のために独自のバッファリング手法を提案した。本手法を用いることで商業集積地域の分布の実態を容易に抽出することが可能である。

抽出した商業集積地域データに電話帳や住宅地図の時系列データを組み合わせることで南関東地域全域における商業集積地域の現状とその変化の様子を、その集積の規模、テナントの回転率、生鮮三品業種の充実度およびチェーン店の割合を使って観察した。東京都心部や地方の中心都市の中心部など高度に都市化が進んだ地域、その周辺の近距離郊外地域、さらに郊外の地域でその状況に大きな違いがあることが明らかになった。

今後の課題はまず商業集積地域抽出の精度を更に高めるための手法を開発することである。現状の手法では都市部における集積の検出は優れているものの、郊外化が進んだ地域のロードサイド型店舗群の検出性能はまだそれほど高くない。現在、住宅地図の建物形状データと組み合わせることで精度の向上を出来る可能性があることが明らかに成り始めており、今後はこの手法の具体化を行う。

商業集積地域の分析についてはまだ大いに検討すべきであろう。本研究で紹介した分析はまだまだ初歩的な内容であり、今後更に興味深いファクターを盛り込んだ分析や、複数のファクターを統合して現象を説明出来る手法を探っていく予定である。複数の有意なファクターの組み合わせで各集積の「活性化度」のようなものを全国規模に渡って観測出来る環境作りが商業集積地域分析の最終的な目標である。また活性化に成功した事例や衰退の一途を辿っている商店街に見られる特徴を、我々のデータを用いてどこまで説明できるかというケーススタディ的な検討も行っていく予定である。

なお本報告書では分析地域が南関東地域全域に限られていたが、我々は既に他の地域への拡張も進めている。少なくとも商業集積地域の分布データだけなら既に日本全国を整備出来る状況にあり、またデータの時系列化システムも全国対応で設計されている。今後はデータの調達が完了した地域から順次エリアを拡大していく予定である。

## (2) 国土政策への活用

最後に本手法の国土政策への具体的な貢献の可能性について紹介する。

まず本研究の前半部分で紹介した空間データの時系列化システムと、それを用いて作成出来る時空間データは、本研究内でも紹介したように既存の統計データの代替データとして活用出来る。国土政策のみならず様々な政策や事業の立案時において統計データは多用されているが、既に紹介したように既存の統計データの欠点はデータ更新の間隔が長いこと、そして少なくとも無料で公開されているものについては集計単位が荒い点が挙げられる。これまでの統計データが持ち得ない性能である、可能な限り新しい、そして非常に細かい空間単位のデータが登場することで、政策立案にも良い意味で影響を与えるものと考えられる。本研究では電話帳データを処理することで事業所・企業統計調査や商業統計の代替データとしての可能性を示したが、今後の本研究の進捗によっては更に様々な既存データから、統計データの代替データ作成の手法が示されるであろう。誰でも入手可能なデータ（無料で公開されているものや購入可能なもの等）から政策立案に有意義な統計データを作成出来ることは、国土政策への貢献に寄与するものと言える。

後半の商業集積地域の抽出と分析手法は、今日の日本で大きな課題の一つとなっている、地方都市の衰退とその再生のための基板データと成り得よう。今後本手法に関する研究が進めば、商業集積・商店街ごとの活性化や衰退化を定量化するインデックスが示されるようになり、その結果からより早急に対策を講ずるべき地域を国土スケールで明らかにすることが可能になる。また一方で活性化に成功した地域を収集することで、活性化のための手法を定量的に示すことが出来るかもしれない。

なお現政権において平成 21 年 12 月 30 日に閣議決定された新成長戦略では、「観光立国・地域活性化戦略」が成長戦略の一つとして掲げられており、その中で「地域資源の活用による地方都市の再生、成長の牽引役としての大都市の再生」という課題に対して、我々のデータが貢献出来る可能性は非常に高いものと考えられる。例えばある地域特有の商業集積あり方や、特定業種の集積と分布はまさに地域資源となりうるリソースであり、こうした情報を定量的かつ時系列的に明らかに出来るデータは現在のところ存在しないため、本手法は現政権での国土政策・国家戦略的にも大いに貢献出来る可能性を秘めていると言えよう。

本手法で得られる商業集積地域データに、前述した時系列化データや既存の空間データ等を統合することで、様々な知見を得ることが出来る。収集出来る情報量は非常に膨大かつ多岐に渡るため、我々だけの手ではそれらを十分に解析仕切ることが困難であると予想される。そのような視点から本研究のあり方を考えれば、我々が今後目指すべきもう一つの道は、商業集積地域データの全国整備（過年度も含めて）と、空間データの時系列化システムを広く公開し、多くの研究者・有識者等に利用してもらうことである。本研究のフランチャイズ化を行うことは我々にとっても、また他の多

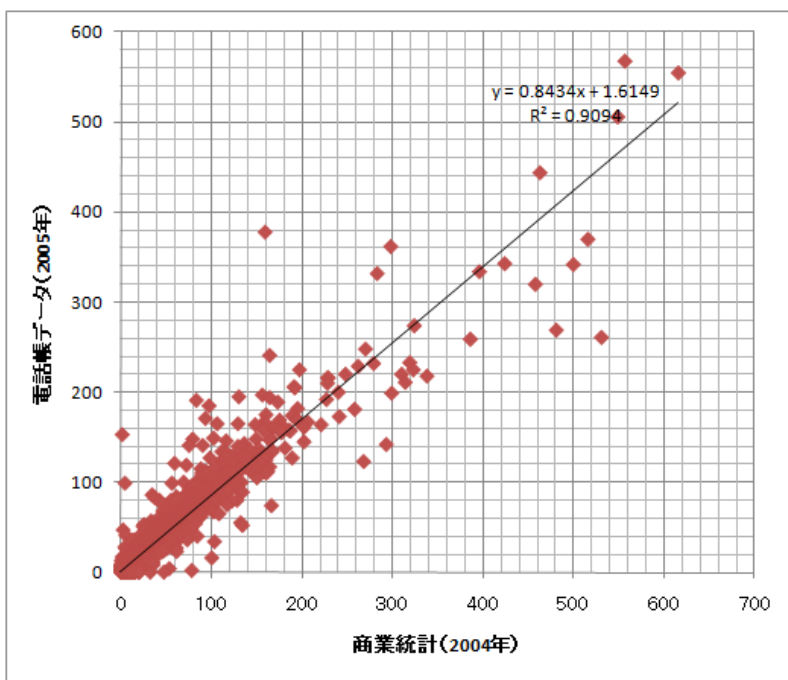
くの研究者らにとっても有意義な環境と結果を創り出すことになるだろう。空間データの時系列化システムは一般的なパソコンで動かすことが出来る設計がなされており、公開に向けての準備を現在進めているところである。また時系列化データや商業集積地域データについてはデータ提供企業や大学側との調整を行い、可能な範囲で公開を検討して行きたいと考えている。なお作業の進捗状況、データやシステムの利用や公開については我々の研究プロジェクトのホームページを参考にされたい。

(<http://shiba.iis.u-tokyo.ac.jp/akiyama/>)

資料1：電話帳データと商業統計メッシュデータの突き合わせ集計結果

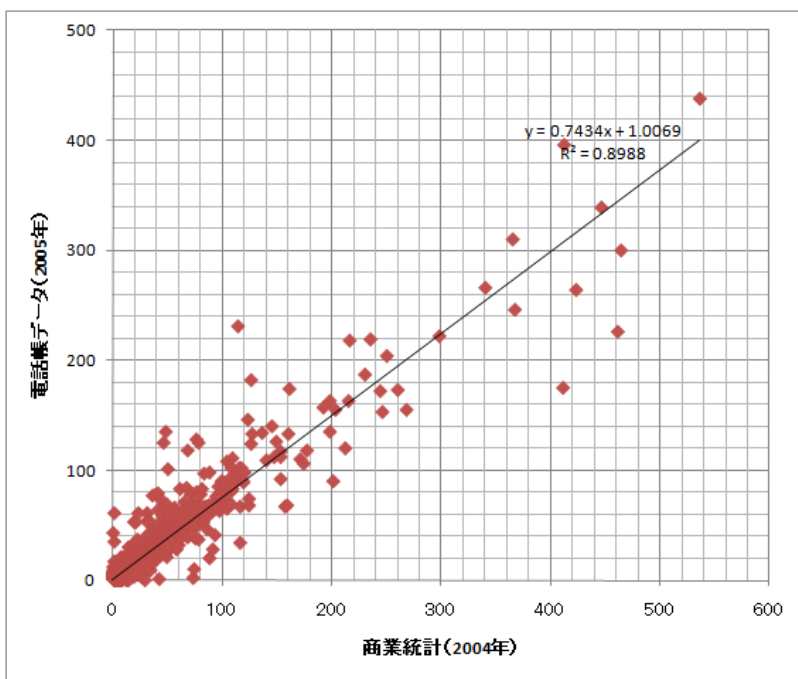
○業種別

・小売店（全業種）総計



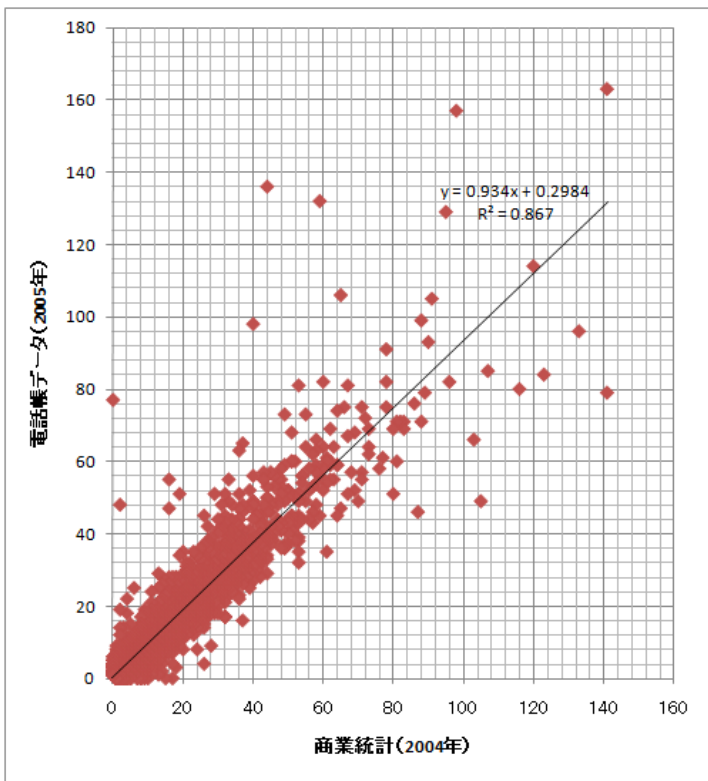
回帰統計	
重相関 R	0.953642
重決定 R2	0.909433
補正 R2	0.909413
標準誤差	11.25174
観測数	4613

・買回品総計



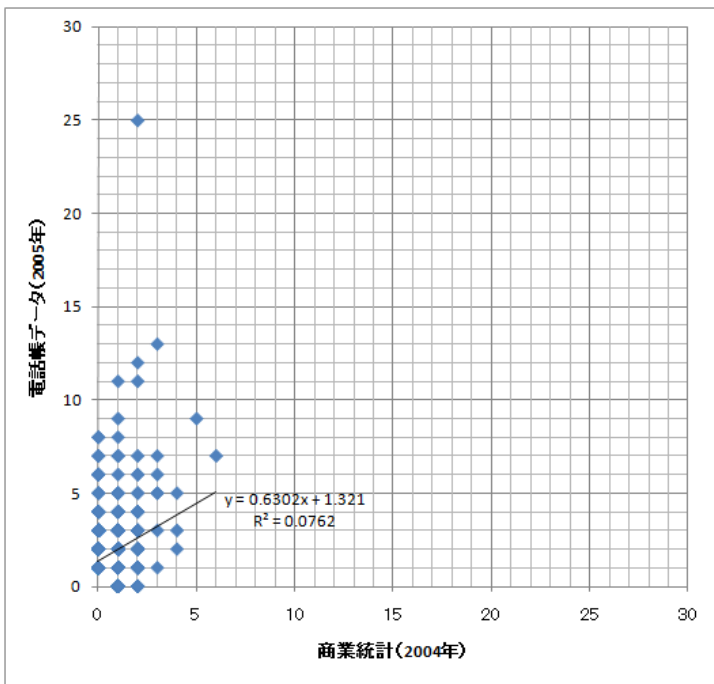
回帰統計	
重相関 R	0.948056
重決定 R2	0.89881
補正 R2	0.898786
標準誤差	7.81251
観測数	4138

・最寄品総計



回帰統計	
重相関 R	0.931122
重決定 R2	0.866988
補正 R2	0.866957
標準誤差	5.300952
観測数	4275

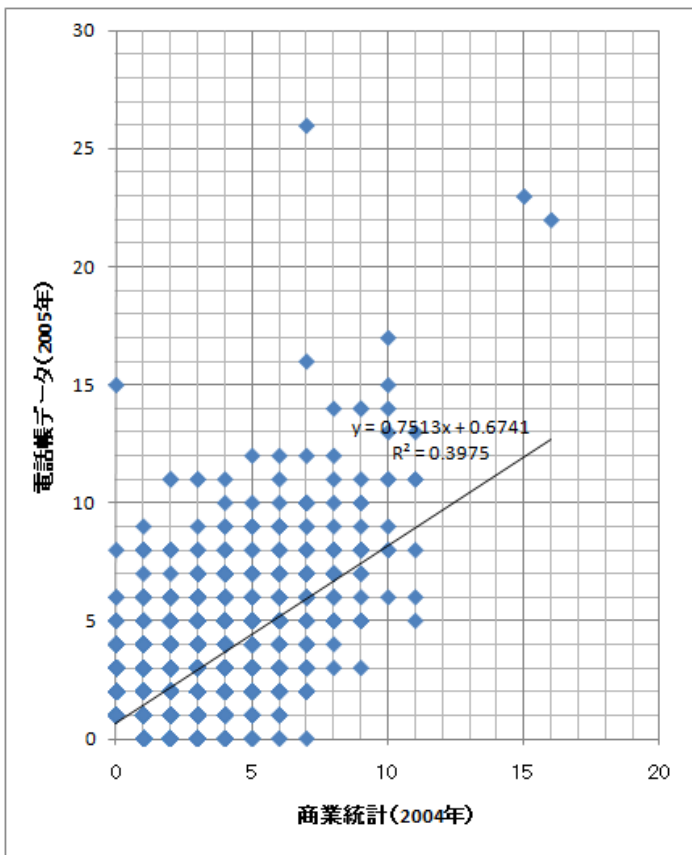
・各種商品総計



回帰統計	
重相関 R	0.276027
重決定 R2	0.076191
補正 R2	0.075586
標準誤差	1.271794
観測数	1528

完全一致は 102 件 (6.68%)  
 ±1 件以内に 1115 件 (79.30%)

・その他総計



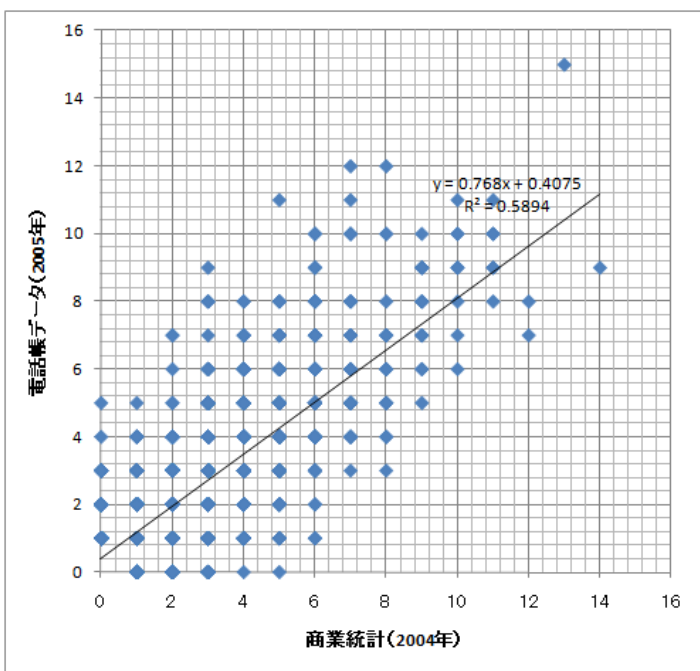
回帰統計	
重相関 R	0.630475
重決定 R2	0.397499
補正 R2	0.397315
標準誤差	1.750358
観測数	3284

完全一致は 692 件 (21.07%)

±1 件以内に 2261 件 (68.85%)

○業態別

・コンビニエンスストア

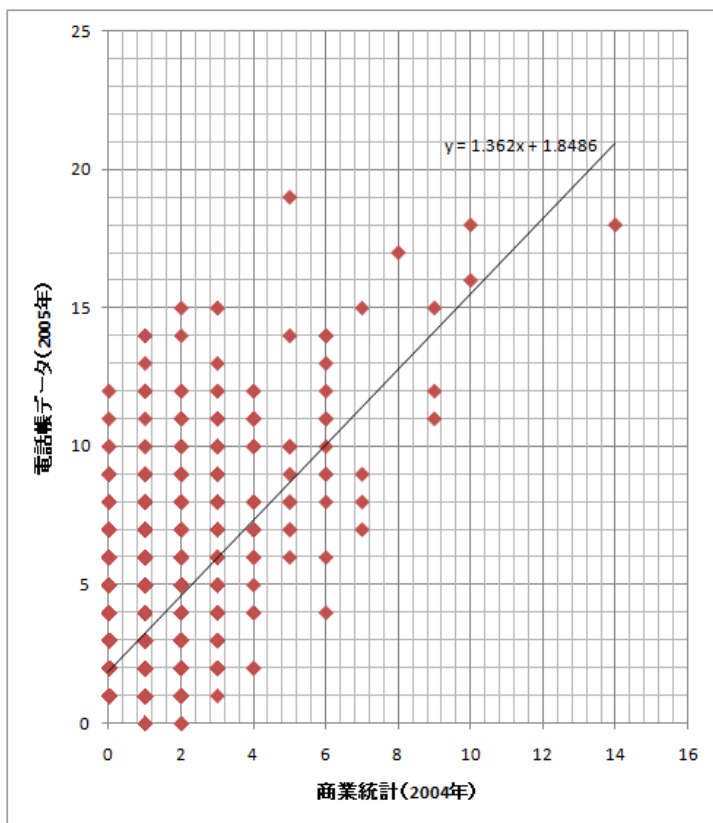


回帰統計	
重相関 R	0.767722
重決定 R2	0.589397
補正 R2	0.589254
標準誤差	1.153335
観測数	2871

完全一致は 979 件 (34.10%)

±1 件以内に 2410 件 (83.94%)

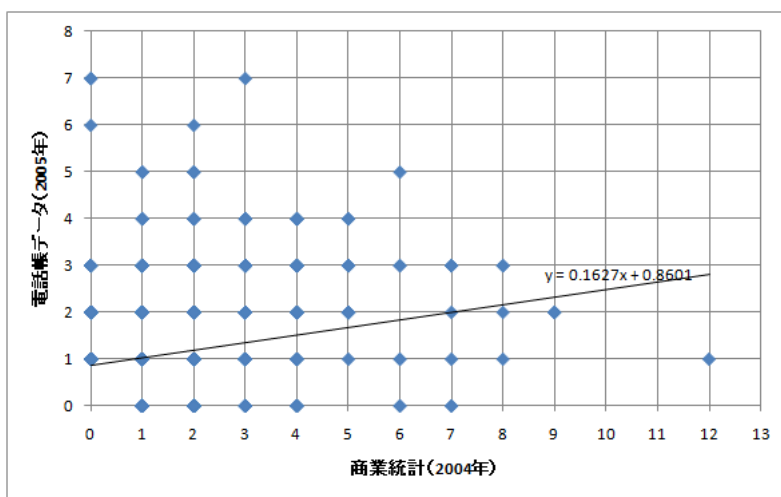
・ドラッグストア



回帰統計	
重相関 R	0.619816
重決定 R2	0.384172
補正 R2	0.383914
標準誤差	1.990702
観測数	2392

完全一致は 174 件 (7.27%)  
 ±1 件以内に 1195 件 (49.96%)

・スーパーマーケット

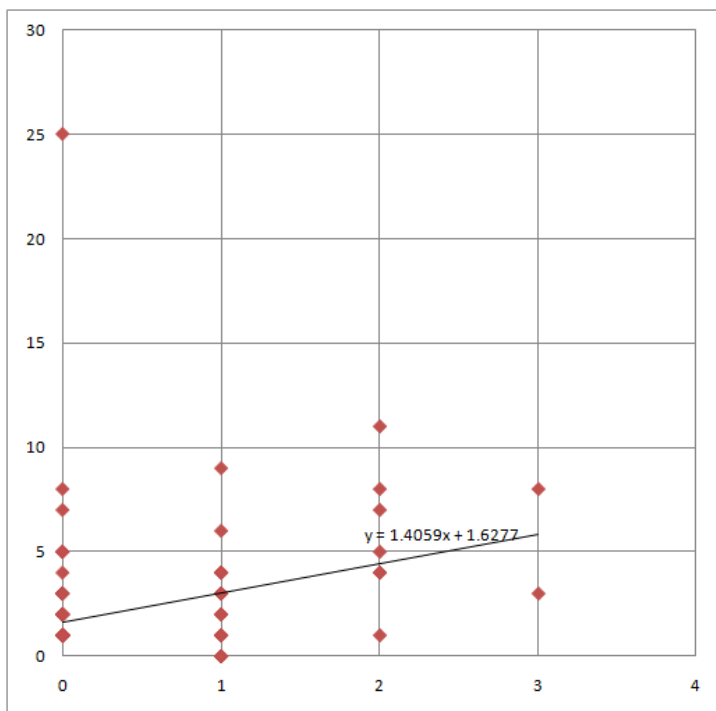


回帰統計	
重相関 R	0.061611
重決定 R2	0.003796
補正 R2	0.003261
標準誤差	1.128951
観測数	1866

完全一致は 532 件  
 (28.51%)  
 ±1 件以内に 1589 件  
 (85.16%)



・百貨店



回帰統計	
重相関 R	0.331713
重決定 R2	0.110033
補正 R2	0.104291
標準誤差	2.429474
観測数	157

資料 2：商業集積地域データと空間的に一致しない商店街名鑑が存在する理由

## 電話帳のもれのケース ⇨ 集積ができていない

商店街はあるが、電話帳データがないために集積地が分断されている

商店街エリア

レイヤ	位置	ファイル	値
H16商店街名鑑_東京23区	[139° 40'15.24" E 35° 39'24.12" N]	FID	691
H16商店街名鑑_東京都/世田谷区/代沢/4*		Shape	ポイント
Join_Output		商店会	代沢高栄会
japan_ver62_GCSStokyo		区	世田谷区
japan_ver62		住所	代沢4-41-7
ku		建物	東京都世田谷区代沢4-41-7
commercial2000_20		LocName	東京都/世田谷区/代沢/4丁目/41番
commercial2005_20		fX	139.6709
commercial2005_2998		fY	35.66067
commercial2005_2998		iConf	5
commercial2000_0		iLvl	7

## 個人宅・個人店が登録住所となり、拾い切れない

商店街名鑑

電話帳

レイヤ	位置	ファイル	値
H16商店街名鑑_東京23区	[139° 40'10.66" E 35° 40'32.74" N]	FID	349
H16商店街名鑑_東京都/中野区/南台/4*		Shape	ポイント
CA_2005_point		商店会	南台4丁目商店会
tokyo_23_2000_CAポイント		区	中野区
沖山橋子		住所	南台4-53-11
commercial2005_20_point		建物	東京都中野区南台4-53-11
tokyo_23_2000		会員数	20
沖山橋子		LocName	東京都/中野区/南台/4丁E
tokyo_23_2005		fX	139.66946
沖山橋子		fY	35.67669
japan_ver62_GCSStokyo		iConf	5
東京都		iLvl	8

資料3：チェーン店・電話帳データにおける表記パターン（一部）

表記パターン	業種
^Doit@店\$	ホームセンター
^Sガスト@店\$	ファミリーレストラン
^アピタ@店\$	総合スーパー
^イトーヨーカドー@店\$	総合スーパー
^オートバックス@店\$	自動車用品店
^オリジン弁当@店\$	その他食料品店
^カインズホーム@店	ホームセンター
^ガスト@店\$	ファミリーレストラン
^株式会社@そごう\$	デパート
^株式会社イトーヨーカ堂@店\$	総合スーパー
^株式会社いなげや@店\$	スーパーマーケット
^株式会社エイブル@店\$	その他
^株式会社小田急百貨店@ショップ\$	デパート
^株式会社小田急百貨店@店\$	デパート
^株式会社カスミ@店\$	スーパーマーケット
^株式会社京王ストア@店\$	スーパーマーケット
^株式会社京急ストア@店\$	スーパーマーケット
^株式会社さいか屋@店\$	デパート
^株式会社さくらや@店\$	家電量販店
^株式会社さくらや@館\$	家電量販店
^株式会社ジョイフル本田@店\$	ホームセンター
^株式会社西武百貨店@店\$	デパート
^株式会社セイミヤ@店\$	スーパーマーケット
^株式会社西友@店\$	スーパーマーケット
^株式会社ダイエー@店\$	総合スーパー
^株式会社大丸ピーコック@店\$	スーパーマーケット
^株式会社東急ストア@店\$	スーパーマーケット
^株式会社東急百貨店@店\$	デパート
^株式会社東武ストア@店\$	スーパーマーケット
^株式会社東武百貨店	デパート
^株式会社長崎屋@店\$	スーパーマーケット

- ・「^」は先頭マッチングを意味する。例えば「^株式会社」なら先頭が株式会社で始まっている必要がある。
- ・「\$」は後方マッチングを意味する。例えば「店\$」なら語尾が店で終わっている必要がある。
- ・「@」は不特定の文字列を意味する。ここにはどのような文字列が入っても構わない。