

電子地図と電話帳情報の時空間結合による店舗及び事業所の広域立体変動モニタリングデータセットの開発

Development of 3D Tenant Monitoring dataset in Broad Urban Area by Spatio-temporal Integrating Digital Maps and Yellow Page Data

秋山 祐樹 [1]; 澁木 猛 [2]; 柴崎 亮介 [3]

Yuki Akiyama[1]; Takeshi Shibuki[2]; Ryosuke Shibasaki[3]

[1] 東大・新領域・社会文化環境; [2] 東大・工・社基; [3] 東大・空間セ

[1] Frontier Science, The University of Tokyo; [2] Civil, Univ. of Tokyo; [3] CSIS,UT

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

これまで都市空間の研究のために多種多様なデータセットが開発されている。しかし広域に渡る高い空間的精度と信頼性を持つデータセットの作成は困難であった。都市の多様な変化の把握には詳細な都市空間の時系列データが必要である。このようなデータの必要性は数多くの既往研究でも指摘されている。そのため日本全土を対象にした時空間都市データセットの開発は、日本の都市で起こっている現象の把握に多大な貢献が出来る。

本研究では都市空間で起こる現象把握のために「テナント」の変化に注目する。テナントとは個人宅を除く店舗、事業所の総称である。テナント情報の収集には電子住宅地図と電話帳情報を用いる。本研究で開発するデータセットは、全てのテナントがテナントの存続、変化、新規出現、消滅といった時系列変化情報を持つ。加えて建物情報まで含む詳細な位置情報と業種情報がほぼ全てのテナントに付加される。

業種情報は電話帳情報のみ保有する。そのためまず電話帳情報と電子住宅地図を結合させる。その後、電話帳情報を持つ複数年分の電子住宅地図を時空間結合することでデータセットが構成される。以上の技術の実現には性質の異なるデータ同士の結合技術が必要となる。また異なる年のデータの時系列変化を特定する技術も必要である。両データの結合および時系列化を行うためには、経緯度、住所、建物情報を用いた位置情報の空間結合技術、および結合後データのテナント名称の同一性判定技術の開発が課題となる。

位置情報は建物情報や住所が欠損する場合もある。そこで位置情報の空間結合では保有する位置情報の情報量に応じて最も確からしい結合先を見つける処理を行う。本研究の位置情報の結合が一般的なGISソフトウェアと大きく異なる点は、経緯度のみでなく、住所や建物情報といった定量化されていない位置情報にまで拡張された部分である。また同一地点に複数のテナントが存在する場合、それらの中から最も名称が類似するものを検索し結合する機能も持つ。高層建築物のような同一地点に複数のテナントが存在する建物の場合、経緯度情報だけでは正確な空間結合は不可能である。

位置情報に基づいて結合されたテナントの同一性はテナント名称に基づいて判定される。このような判定の自動化にはテナント名称の類似度の定量化技術が必要である。これはテナント名称の表記に微妙な揺れが生じる場合があるためである。本研究では名称の類似度をn-gramを用いて定量化し、同一性判定を行った。n-gramは単語や文章の類似度を定量化出来る自然言語処理の手法のひとつである。そのため近年では文学や言語学の領域でも注目されつつある。

ただしn-gramだけでは正確な同一性判定は困難である。原因は「株式会社」「不動産」といった頻出語や、地名などのノイズとなる文字列である。本研究ではこうした頻出語をライブラリ化し削除する処理を可能とした。また英語表記の名称を日本語の読みに変換することも可能とし、処理精度を大幅に向上させた。

システムの処理精度およびソースデータの信頼性の検証も行った。システムの処理精度検証では本システムにて作成したデータと、手作業にて作成したデータの比較を行った。サンプルは東京23区から5地区500件(各地区100件)をランダムに取り出した。その結果、両データの結合処理結果の94.00%、電話帳時系列化の95.88%、住宅地図時系列化の93.00%が正確に処理された。

ソースデータの信頼性検証では赤羽駅東口付近および下高井戸駅周辺にて現地調査を行い、ソースデータとの比較を行った。その結果、赤羽駅周辺では住宅地図の場合で88.16%、電話帳データで87.74%のデータが実空間と一致した。下高井戸駅周辺では、住宅地図で89.51%、電話帳データで91.51%が実空間と一致した。ソースデータと現地調査の調査時期のずれ(約2年半)を考慮すれば、ソースデータが広域に渡る実空間を詳細に観察出来るデータ、という要求に耐えるデータと言える。

最後に本データの適用事例を示す。図は2005年と2000年の時空間データを用いて、そのテナント変遷の様子を3次元的に視覚化したものである。このような3次元時空間データが日本全土において整備可能である。

本研究で開発した技術を用いることで、テナントスケールの詳細さを持つ時空間データセットの整備が可能となった。本技術の開発過程にて住所や建物情報の処理にまで拡張された空間結合技術、およびテキストの曖昧性を考慮したテキスト同定技術を開発した。そしてこれらを組み合わせることでほぼすべてのテナントが業種情報および時系列変化情報を保有する詳細な時空間データセットが実現した。これらの処理は自動化されており、都市の変動を低コストかつ継続的にモニタリングするための実用的な手段といえよう。

