数値地図2500を用いた街区代表点の算出

島田貴仁・原田 豊

Calculating Representative Point of Block Using Digital Map 2500

Takahito SHIMADA and Yutaka HARADA

Abstract : This paper points out that the records of “block level spatial reference data” don’t reflect the centroid of plural polygons for unique address. To improve accuracy of block’s centroid, we suggest to recalculate centroid of block using “Digital Map 2500”. This method involves dissolving polygons that have identical address using ArcView8.1. Generated multipart polygon improved accuracy by an average of 30m for non-unique, 5.6m for unique records. Importance of data maintenance is also discussed.

Keywords : 街区 (block), 代表点 (centroid), 住所照合 (address matching)

1. 問題の所在

1. 1 住所照合

住所照合（アドレスマッチング）は「住所表記からその位置を同定すること」と定義される（地理情報システム学会, 2000）。自然科学分野のGISでは緯度・経度など直接空間参照可能なデータが用いられることが多いが、人文・社会科学では住所や郵便番号といった間接的な空間参照に頼ることが圧倒的に多い。このため、住所照合は人文・社会科学分野でGISを活用する際の重要な技術といえる。社会調査結果や犯罪発生地点に関するデータも住所照合により空間データ化されている。

これまで、島田・原田(1997)は数値地図2500を用いて住所照合を行うための問題点を整理し、原田・島田(1998)は ArcView3.0 環境で、数値地図2500の町丁目・街区データを用いて住所照合を行う住所照合プログラムを開発している。

また、相良・有川(2000)は、現在の住所照合につ

いて、(1)前処理として住所を都道府県名、市区町村名、町丁目名といった分かち書き表記に分解する必要がある、(2)住所照合を行うためのテーブルが容易に作成できず、高価であるという問題点を指摘した上で、分かち書き不要なアドレスマッチングサービスを提供している。

住所照合には、住所と、その住所を代表する座標（代表点）とを対応させる住所照合テーブルが必要となる。本稿では、現在、最も容易に入手しやすい空間データの一つである、街区レベル位置参照情報を住所照合テーブルに用いる際の問題を指摘し、その解決法を検討する。

1. 2 街区レベル位置参照情報

街区レベル位置参照情報は、国土交通省国土計画局により整備が進められている街区単位のポイントデータであり、同省の Web サイトで公開されている(国土交通省国土計画局総務課国土情報整備室, 2002)。CSV 形式の表であり、分かち書き形式での住所（都道府県名、市区町村、大字・町丁目名、街区符号・地番）と、平面直角座標での XY 座標、緯度経度といったフィールドを有している。

島田：〒277-0882 千葉県柏市柏の葉 6-3-1
科学警察研究所犯罪行動科学部犯罪予防研究室
Crime Prevention Section, National Research Institute of
Police Science
6-3-1, Kashiwanoha, Kashiwa, Chiba, 277-0882, JAPAN
E-mail: takajin@nrips.go.jp

表形式であるため住所照合テーブルとして加工しやすく、GIS 上ではポイントデータとして容易に扱うことができる。先述した相良・有川(2000)でもこの街区レベル位置参照情報が用いられている。

街区レベル位置参照情報は、2000年度から整備を開始し、2001年度末に全国の都市計画区域について整備が完了している。

街区レベル位置参照情報は、街区代表点の作成と、街区符号・地番の付与という2段階を経て作成されている。街区代表点は、原則として数値地図2500から取得され、街区代表点がない場合には数値地図2500の道路中心点と行政界から街区ポリゴンを認識させることにより取得される。また、街区符号・地番は、住居表示実施区域の場合には住所表示実施図、住居表示街区位置図、住宅地図から付与される。未実施区域の場合には、ブルーマップ、住宅地図、公図を参照し、街区相当範囲に含まれる地番の本番から付与される。

図1 街区レベル位置参照情報と数値地図2500

1. 3 住所照合に用いるデータの要件

住所照合の目的は、ある住所に対応する座標の特定であるので、住所照合テーブルの住所フィールドが一意であることが求められる。住所が一意でない、すなわち、住所と座標が1対多関係の場合は、住所から座標を特定することができなくなる。

街区レベル位置参照情報の作成元である数値地図2500では、同一住所を属性データとして有する複数の街区ポリゴンの存在が指摘されている(原田・島田, 1998)。街区レベル位置参照情報においても、同一の住所を持つ複数レコードが一般的に存在している。住所が一意でないので街区レベル位置参照情報をそのまま住所照合テーブルに用いることはできない。

そこで、街区レベル位置参照情報では、住所の一意性を確保するために「代表フラグ」フィールドが用意されている。同一住所を有する複数レコードの中で1レコードのみが「代表フラグ」フィールドの値が1となっている。したがって、クエリによって、一意な住所と座標を有する住所照合テーブルを容易に作成できる。

1. 4 街区レベル位置参照情報の代表点の正確性

しかし、代表フラグにより一意化した街区レベル位置参照情報ではその正確性に問題が生じる。

図1は、ポイントとして空間データ化した街区レベル位置参照情報と数値地図2500の街区ポリゴンとを重ねあわせたものである。図中の数字は街区符号を示している。ここで、図中の破線で囲んだ「田島一丁目10番」や「巽北一丁目12番」のように、街区レベルの同一住所に対応するポリゴンが複数存在する場合、そのうちの1ポリゴンの中心が代表点となり、それ以外のポリゴンの中心は非代表点となる。ところが、この代表点は、同一住所に対応する複数ポリゴン全体を図形的に代表するものではないため、視覚的にも誤解を生じる可能性が強い。

例えば、「田島一丁目10番」の代表点は、非代表点を含む5つのポリゴンの中でかなり南にずれているし、「巽北一丁目12番」の代表点は、非代表点を含む2つのポリゴンの中では東にずれることになる。

街区レベル位置参照情報を用いて住所照合したポイントデータ（例えば犯罪発生地点）は、代表点のみに落ちることになる。この代表点は、一見したところ代表点を包含するポリゴンの重心にあるように見えるため、かえって深刻な影響をもたらす。というのは、街区ポリゴンデータと重ね合わせて表示した場合に、代表点に対応するポリゴンのみで犯罪が発生し、非代表点に対応するポリゴンでは犯罪が発生していないというように見えるからである。

この誤解を回避するためには、街区レベルでの住所の代表点をより正確に表現する住所照合テーブルを作成する必要がある。ここで、廉価で入手可能な空間データからより正確な街区代表点を求める手段として、1) 街区レベル位置参照情報の代表点と非代表点から新しい点を合成する、2) 街区レベル位置参照情報の作成元である数値地図2500の街区ポリゴン情報から代表点を再計算する、といった方法が考えられる。方法1)は簡便ではあるが、ポリゴンの図形的な情報が捨象されるため、同一住所に対応する複数ポリゴン間で面積や形状が異なる場合、良質な結果が得られないと考えられる。

そこで本研究では、方法2)に従い、数値地図

2500の街区ポリゴン情報から、より正確度が高い街区代表点を算出する手法を提案する。

さらに、提案された手法により再計算した街区代表点と、街区レベル位置参照情報に由来する街区代表点との比較を行い、提案された手法により代表点の正確度がどのくらい改善されるかを計量的に評価した。

2. 方法

2. 1 数値地図2500データを用いた街区代表点の算出

数値地図2500の街区データは、属性データとして「市区町村名」「町丁目名」「街区符号」等を有するポリゴン形式の情報である。

このデータには、図2に示すような同一住所を有する複数ポリゴンが一般的に存在する。そこで同一住所を有する複数ポリゴンを1ポリゴンに統合して、住所の一意性を確保した上で、重心を求めることにした。その手順を以下に示す。

1. 市区町村、町丁目名、街区符号を連結した「住所」フィールドを作成する。
2. 「住所」フィールドを基準に、同一値を持つ

	FID	Shape*	FILEID	CODE	ID	JISCODE	CYCODE	CYNAME	GAINAME
▶	41651	Polygon	0600994	A1108	40219	27116	1407	箕北一丁目	10
	41674	Polygon	0600994	A1108	40242	27116	1430	箕北一丁目	10
	41649	Polygon	0600994	A1108	40217	27116	1405	箕北一丁目	11
	41675	Polygon	0600994	A1108	40243	27116	1431	箕北一丁目	11
	41652	Polygon	0600994	A1108	40220	27116	1408	箕北一丁目	12
	41676	Polygon	0600994	A1108	40244	27116	1432	箕北一丁目	12
	41653	Polygon	0600994	A1108	40221	27116	1409	箕北一丁目	13
	41677	Polygon	0600994	A1108	40245	27116	1433	箕北一丁目	13
	41654	Polygon	0600994	A1108	40222	27116	1410	箕北一丁目	14
	41678	Polygon	0600994	A1108	40246	27116	1434	箕北一丁目	14
	41650	Polygon	0600994	A1108	40218	27116	1406	箕北一丁目	9
	41673	Polygon	0600994	A1108	40241	27116	1429	箕北一丁目	9

図2 数値地図2500の属性データ（処理前）

	FID	Shape*	Ave_JISCOD	address	x	y
▶	18105	Polygon	27116	大阪市生野区箕北一丁目10	135.547445	34.651670
	18106	Polygon	27116	大阪市生野区箕北一丁目11	135.547453	34.651986
	18107	Polygon	27116	大阪市生野区箕北一丁目12	135.547463	34.652294
	18108	Polygon	27116	大阪市生野区箕北一丁目13	135.547457	34.652634
	18109	Polygon	27116	大阪市生野区箕北一丁目14	135.547463	34.652950
	18133	Polygon	27116	大阪市生野区箕北一丁目9	135.547495	34.651348

図3 数値地図2500の属性データ（処理後）

複数ポリゴンを1ポリゴンにマージする。

3. マージされたポリゴンについて重心を算出して属性データとして付与する。

本手法はArcGIS8.1を用いることを想定している。

ArcGIS8.1には、フィールドに同一値を持つポリゴンをマージする「ディソルブ」機能が用意されており、上記の作業をGUIで対話的に行うことができる。

マージされたポリゴンは「マルチパートポリゴン」と言われ、見かけ上は複数のポリゴンであるが、実際は1つの地物として扱われる。

本手法によりマージされた街区ポリゴンの属性データを図3に示す。処理により住所が一意となり、座標データが付与されている。この属性データはそのまま住所照合テーブルとして利用可能である。

図4には、本研究が提案する手法により数値地図2500から算出された代表点が○で、既存の街区レベル位置参照情報の代表点が×で示されている。

住所とポリゴンが1対1対応する場合には両代表点の位置にはほとんど差がない。ある住所に対応するポリゴンが複数存在する場合は、両者に違いが生じる。今回提案した手法により算出された代表点は、

図4 既存の街区レベル位置参照情報の代表点と数値地図2500から再計算した代表点との比較

複数のポリゴンにより構成される領域の中心に位置している。既存の代表点に比べて、住所に対応する領域をより忠実に反映しているといえる。

2. 2 対象地域とデータ

実証データによる検討では、数値地図2500データが市域全域をカバーしている大阪市を対象地域として選定した。

街区レベル位置参照情報は国土交通省サイトからダウンロードした。数値地図2500はCD-ROM版「大阪-2」「大阪-3」に日本地図センターが提供しているパッチを当てて使用した。

街区レベル位置参照情報については、代表フラグに基づきクエリをかけ、住所が一意的な代表点を抽出した。数値地図2500データについては、2.1節で示した手法により、同一住所を有するポリゴンをマージし、その重心を求め代表点とした。

3. 結果

本研究では、住所と地物が1対多対応しているデータの問題を指摘し、住所の一意化を行った。そこで、以下の分析では、住所と地物との対応関係が1対1対応であるもの（以下「1対1対応データ」と表記）と、1対多対応であるもの（以下「1対多対応データ」と表記）に分けて結果を検討する。なお、今回行った住所の一意化処理では、1対多対応データのみが処理の対象になり、1対1対応データはもともと住所の一意性が確保されているので処理による変化は生じない。

3. 1 地物数の変化

図5は、各データの処理前後における地物（ポイント、ポリゴン）数を、住所と地物の対応状況別に示したものである。

総数で見ると、住所の一意化処理により街区レベル位置参照情報のポイント数が34105から25093へ、数値地図2500のポリゴン数が35624から24860へと、それぞれ3割程度圧縮されている。

街区レベル位置参照情報においては、クエリによって15037あった1対多対応データの中から6025

図5 処理前後の地物数の変化

の一意なデータが選択された。これは、全レコード数25093の中の24%を占める。すなわち、街区レベル位置参照情報から作成される既存の住所照合テーブルのうち2割強は、代表点の正確度に問題があるといえる。

一方、数値地図2500では、1対多対応データの17914ポリゴンが7150のマルチパートポリゴンに圧縮された。このポリゴンの重心を計算することでより正確度の高い代表点が得られることになる。

なお、処理前の数値地図2500と街区レベル位置参照情報の地物数を比較すると、総数で約5%、1対多対応データに限ると20%ほど、数値地図2500の方が多くなっていた。これは、数値地図2500では、図郭線にまたがるポリゴンが複数ポリゴンに分割されていることが原因だと考えられる。

表2は、住所の一意化処理が行われた両データの各レコードを、処理前の住所と地物の対応状況から集計したものである。住所と地物が1対1対応しているデータは全体の75%ほどに過ぎなかった。一方、同一住所に対応する地物が4つ以上見られるケースが全体の約1%ほどを占めていた。

表2 住所と地物の対応関係による分類

地物数	街区レベル位置参照情報		数値地図2500	
	度数	累積比率	度数	累積比率
1	19068	76.0%	17710	71.2%
2	4230	92.8%	5070	91.6%
3	1156	97.5%	1239	96.6%
4	404	99.1%	539	98.8%
5	123	99.6%	155	99.4%
6	55	99.8%	71	99.7%
7	18	99.8%	24	99.8%
8	11	99.9%	26	99.9%
9	9	99.9%	8	99.9%
10	6	99.9%	7	100.0%
11以上	13	100.0%	11	100.0%
合計	25093		24860	

3. 2 街区代表点のずれの評価

本研究で提案した手法により、街区代表点の正確度がどの程度改善されたかを定量的に評価するため、街区レベル位置参照情報と数値地図2500から計算した街区代表点データのずれを算出した。

この分析には、街区レベル位置参照情報と数値地図2500が正しく対応していた24729件のデータを用いた。街区レベル位置参照情報の364レコード、数値地図2500の131ポリゴンは対応するデータがなかったため分析から除外した。

表3は、代表点のずれの基本統計量を、住所と地物との対応関係別に算出したものである。本研究では、街区レベル位置参照情報の住所と地物とが1対多対応の場合、街区レベル位置参照情報と数値地図2500の街区代表点のずれは平均で35mほどのずれが見られた。本手法を用いることでこのずれが改善することが見込まれる。

なお、1対1対応データの中でも、街区レベル位置参照情報の代表点と、数値地図2500ポリゴンの重心との間で平均5.64mのずれが見られた。大阪市の臨海部等の面積が大きな街区では100mオーダーのずれが見られたため、ずれの平均値は中央値に比べて大きくなっている。

表3 代表点のずれの基本統計量

		標本数 (個)	平均値 (m)	中央値 (m)	標準偏差 (m)
対応関係	1対1	18752	5.64	1.76	16.08
	1対多	5977	34.48	26.45	39.27
	合計	24729	12.61	2.92	26.86

図6 代表点のずれの分布

図6は、代表点のずれの分布を、住所と地物の対応別に示したものである。

全体で見ると、約2割のデータで、ずれが20m以上になっていた。すなわち、街区レベル位置参照情報の代表フラグに従って住所照合テーブルを作成すると、およそ2割でずれが20m以上になることが見込まれる。

1対1対応データに限ると、ずれは小さく、9割近くのデータで10m以内に収まっていた。

1対多対応データでは逆にずれが顕著で、ずれが10m以内だったのは13%に留まり、およそ半数で、ずれが25m以上となっていた。本研究で提案した手法を用いることにより、特に住所と街区が複雑な対応関係にある地区で、代表点の正確度の改善が期待できる。

4. まとめ

本研究ではまず住所照合の重要性を述べ、インターネットで入手可能な街区レベル位置参照情報を、住所照合に用いる際の注意点を指摘した。

街区レベル位置参照情報では、一意な住所を得るために代表フラグによるクエリをかけることにより、

住所が一意な代表点データを得ることができる。しかし、この方法では、1つの住所に複数レコードが対応する場合、代表点の正確度に問題が生じることが懸念される。

そこで、本研究では、街区レベル位置参照情報と並んで廉価に入手できる数値地図2500を用いて、住所の一意性を確保した上で、住所と街区ポリゴンが1対多対応する場合に、複数のポリゴンが占める領域をより正確に反映した代表点を求める方法を提案した。

さらに、大阪市を対象地域として、本手法を用いて数値地図2500から再計算した街区代表点と、既存の方法で街区レベル位置参照情報から算出した街区代表点とを比較し、正確度の改善が期待される程度を定量的に評価した。

本方法により、住所という間接的な参照データがより正確に空間データ化されることが期待される。

また、住所照合の精度が高くなればなるほど、住所照合テーブルのメンテナンスが問題になってくる。住居表示の実施、市町村合併による住所変更、都市再開発による街区の改変などで、街区レベルの位置参照情報は絶えず変化しつづけている。空間

データだけではなく、住所照合テーブルも適切にメンテナンスしなければ、すぐ陳腐化して照合の成功率が低下することが懸念される。

数値地図2500や街区レベル位置参照情報はともに地理情報の中でも中核と位置付けられ、年に1回程度の更新が期待されている（地理情報システム関係省庁連絡会議，2002）。

本研究で提案した方法を用いることにより、数値地図2500の更新に応じて、住所照合テーブルを常に最新の状態に保つことが可能になる。

謝辞

本研究は、科学技術振興調整費・知的基盤「空間情報科学の確立のための空間情報のデータベース化に関する研究開発」の一環で行われた。

参考文献

国土交通省国土計画局総務課国土情報整備室(2002) 街区レベル位置参照情報ダウンロードサービス, < [http://](http://nfltp.mlit.go.jp/isj/)

nfltp.mlit.go.jp/isj/>.

相良毅・有川正俊(2000) 日本の住所体系に適した分散アドレスマッチングサービス, 「地理情報システム学会講演論文集」, **9**, 183-186.

島田貴仁・原田豊(1997) 地域安全情報システムのための住所照合ソフトウェアの設計, 「科学警察研究所報告(防犯少年編)」, **38**(2), 1-12.

地理情報システム学会用語・教育分科会編(2000) 『地理情報科学用語集』(第2版) .

地理情報システム関係省庁連絡会議(2002) GISアクションプログラム2002-2005, < http://www.mlit.go.jp/kisha/kisha02/02/020220_.html>.

原田豊・島田貴仁(1998) 数値地図2500のための住所照合ユーティリティの開発, GISA 第1回バーチャルカンファレンス, <http://www.asahi-net.or.jp/~RW4Y-HRD/vcgis98/>

〔原稿受付：2002年5月21日〕
〔採用決定：2003年7月7日〕