

CS-5

都市解析における時空間データ変換の誤差評価法

地域未来研究所 正 員 中野 光治  
 東北大学 学生員 杉木 直  
 東北大学 正 員 内田 敬  
 東北大学 フェロー 宮本 和明

1.はじめに

都市モデルを扱う際には、多くのデータ入力により将来予測が行われるが、その際の入力データは必ずしも既存の統計調査で整備されているとは限らないために、解析を行う前の段階で何らかの変換が行われていることが多い。本研究では、この変換をデータ変換、およびこのデータ変換による誤差を変換誤差と呼ぶ。この変換誤差については、これまで十分な考慮がされてこなかった。そこで本研究では、1) データモデルを用いたデータ表記の一般化。2) データモデルに基づくデータ変換の表現。3) 変換誤差の一般的な評価。の順序をとることで、データ変換誤差の体系的評価を行い、最良なデータ変換を行うための手法を提案した。

2.データモデルによるデータ表記

都市モデルが扱うデータの属性は、「時間属性」「空間属性」「主題属性」の三つに分類される。このとき、データ変換はこれらの属性の変換として捉えることができる。例えば、時系列回帰は時間属性の変換、面補間は空間属性の変換である。さらに、これらの属性を「位置情報」と「サブスキーマ」とで以下のように捉えることで、データ空間の枠組みを考える。

・位置情報

属性の位置を特定化するための情報

・サブスキーマ

属性の位置情報の構成を定義する情報

以下、この概念を、個々の属性について適用する。

(1)時間属性

データは収集に際して5年おき・毎月などのある時間単位に基づいて計量されている。「時間サブスキーマ」は、このような単位を表し、「時間位置情報」は、その中の一時点もしくは一期間を表す（図1）。

(2)空間属性

「空間サブスキーマ」は、空間集計単位など、対象地

域の空間領域の取り方を定義するもので、「空間位置情報」は、その中の一領域もしくは一地点を表す（図2）。  
 (3)主題属性

主題属性についても、図3のようなデータ表を用いてデータを整理することで、主題（位置情報）と主題サブスキーマにより主題を表現する。

以上をもとに、それぞれの属性の位置情報とサブスキーマを形式的に互いに直交する軸として配置することで、データ空間が表現できる。データの値は、このデータ空間上の格子点に与えられていると捉える。このときデータ空間全体の構成は図4のようになる。

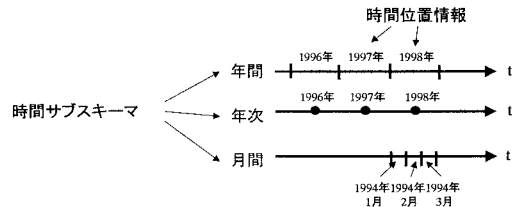


図-1 時間位置情報と時間サブスキーマ

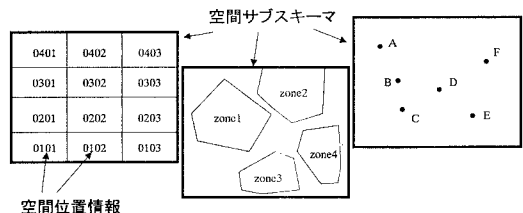


図-2 空間位置情報と空間サブスキーマ

| 主題サブスキーマ |       |     | 産業別(大分類, 小分類)<br>年齢階層別 など | 主題        |
|----------|-------|-----|---------------------------|-----------|
| 都市活動要素   |       |     | 対象 数地面積                   | 人数        |
| 大分類      | 中分類   | 小分類 |                           |           |
| 立地主体     | 第1次産業 | ... | 農地・山林等面積                  | 第1次産業就業人数 |
|          | 第3次産業 | ... | 第3次産業事業所敷地面積              | 第3次産業就業人数 |
|          | 卸売業   | ... | 卸売業事業所敷地面積                | 卸売業就業人数   |
|          | 金融業   | ... | 金融業事業所敷地面積                | 金融業就業人数   |
| 居住       | 住宅    | ... | 住宅建物床面積                   | 夜間人口      |
| 道路       | 道路    | ... | 道路面積                      |           |
| ...      | ...   | ... | ...                       | ...       |

図-3 主題と主題サブスキーマ

キーワード：GIS、都市モデル、時空間データ、誤差評価、データ変換

連絡先：〒520-0055 滋賀県大津市春日町5番11号 REC 大津3階 TEL 077-522-6163 FAX 077-522-6231

### 3. データ変換の表現

都市モデルで扱うデータが、データモデルを用いてデータ空間内の一点で表されたことによって、データ変換はデータ空間上の移動として表現できる。この移動は、移動方向に平行する軸が「位置情報」か「サブスキーマ」かで異なる。このとき「位置情報」の変換は、内挿などの回帰式で表され、「サブスキーマ」の変換は、例えば、式(1),(2)などのような単純化と詳細化を組み合わせたものとして表される。

$$\text{単純化 集計: } d(x_j) = \sum_{x_i \in S_j} d(x_i) \quad (1)$$

$$\text{詳細化 分割: } d(x_i) = \frac{c(x_i)}{c(x_j)} d(x_j) \quad (2)$$

( $c$  は配分指標,  $x_i$  は  $x_j$  に含まれる集合)

### 4. 変換誤差の評価

変換誤差の評価には、変換式の誤差分散を用いる。本研究では観測誤差については所与のものとする。データ変換後の誤差は、観測誤差と変換誤差が、誤差伝播の法則に従って伝播したものとして表現される。このような表現を用いることで、複数の変換を行って算出されたデータの誤差も、簡単に分散による表現が可能となる。データ空間を用いてこの複数の変換を一つの変換経路として表現(図5)したとき、それぞれの変換経路による誤差が評価されるため、このときの最も分散の小さい変換経路を選択することで、最良のデータ変換が行われる。

### 5. 適用

最良なデータ変換手法のケーススタディとして、80,85,90年の国勢調査および都市計画基礎調査の人口・世帯数・就業者数のデータから、1992年の仙台都市計画基礎調査区の就業者数を求めた。システムとしては、ATOM(朝日航洋(株))のGISを用い、都市計画基礎調査区と3次メッシュ、さらにこれらを重ね合わせたゾーン図の面積のデータを作成した。対象地域には、仙台都心部の5km四方を選んだ。

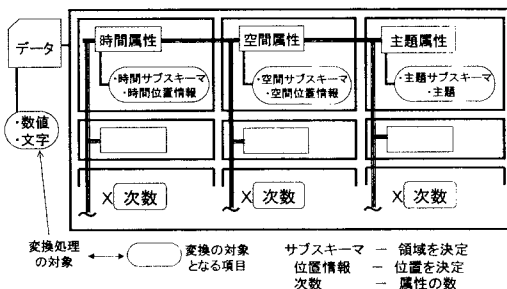


図-4 データモデル

以上の結果として、変換経路は34通り検出された。

図5にその一例を示す。最良変換経路の決定には、対象地域全体で同一とする場合の他に、ゾーン個別の経路選択を行うという2つの場合を考えた。前者の場合は、対象地域全体で経路32が選択された。今回の条件では、空間サブスキーマの変換の影響を強く受ける変換について、比較的大きな誤差を生み出す傾向の持つ条件であったが、図6は、この影響を受けた一部の経路が大きな誤差分散を持つことを示しており、恣意的な変換経路の選択の危険性を十分に表していると言える。また、後者の場合は、対象地域北西部で経路32以外の経路が選択された。この地域は、人口が対象地域内での平均に近かったために、就業者数との回帰を行う変換の誤差が小さかったためである。

### 6. 結論

本研究では、都市モデルの予測値の誤差について、観測誤差、変換誤差、モデルに起因する誤差に分類した上で、変換誤差に着目し、この誤差を体系的に表現するためのシステムを構築し、仙台市への適用を行った。その結果、データ変換の際の、変換経路の不用意な選択の危険性が示され、また、対象地域の特徴に応じた、個別の変換経路の選択という新たな手法の実用性が示された。

### 参考文献

- 1) GISの標準化に関する調査報告書(国土地理院技術資料 E・1-No.247)

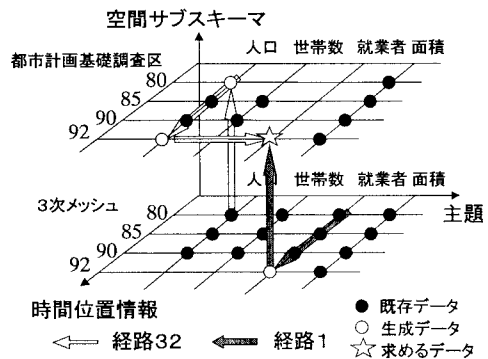


図-5 検索された変換経路の一部

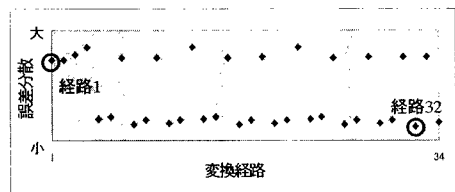


図-6 それぞれの変換経路における誤差分散