

IV-45

時空間データ変換の誤差評価法

○東北大学 学生員 川合健司
東北大学 正 員 内田 敬

1.はじめに

現在多くの都市モデルが開発・適用される中で、モデルに用いる時空間データの不足の問題にしばしば直面する。そのような場合、既存のデータから所要データの推定を行う。例えば、既存の3次メッシュの人口データを、面積に応じて未知の町丁目の人口データに配分したりする。このように既存のデータに何らかの操作をし、新たなデータを作成することをデータ変換と呼ぶことにする。本研究では、最適なデータ変換を決める基準の1つである、データ変換誤差を定量的に評価する方法を明らかにする。

2.データ変換の考え方

データ変換に応じた誤差評価法を考えるには、そのデータ変換がどういった特徴を持っているかを整理しなくてはならない。

(1) データ空間による表示¹⁾

時空間データは「時間」、「空間」、「主題」の3つの要素によって特定化される。これらのサブスキーマは、それぞれ参照システム及び参照システム上の位置が記述される。時間サブスキーマにおいては、年間・年次のような期間や時点などの時間軸上での領域の取り方が時間参照システムであり、その中の1999年といった1期間・1時点が時間位置である。同様に空間サブスキーマにおいても座標系やゾーンのような領域の取り方が空間参照システムであり、その中の1領域・1地点が空間位置である。主題サブスキーマにおいても、同様に主題参照システムと主題位置情報の表現が可能である。

このようなデータ空間において、あるデータはデータ空間内の1点として特定化される。

(2) 参照データによるデータ変換経路の定義

データ変換の元となる既知のデータをソースデータ、推定すべき未知データをターゲットデータと呼ぶ。さらにソースデータとターゲットデータ以外のデータ変換に関わるすべてのデータを参照データと呼ぶことにする。例えば3次メッシュの人口データ

を、面積に応じて町丁目の人口データに変換するのであれば、メッシュ人口がソースデータ、町丁目人口がターゲットデータ、面積が参照データとなる。すべての変換は、ソースデータに参照データを作用させてターゲットデータを得る点で、本質的には同じである。よって、変換形式ではなくどの参照データを用いるかによって変換が定まると考える。

あるターゲットデータを推定する場合、どのような参照データを用いるかによって複数の変換経路が存在する。参照データを作用させることにより各軸方向にデータの移動が可能になる。すべての変換経路は、データ空間内のソースデータからスタートし、この移動の組み合わせによってターゲットデータに到達するものである。あるターゲットデータに対する変換経路の多様性の例を図-1に示す。既存のデータがなくても、参照データを作用させることにより、新たなデータを生み出し、さらにそこからデータ変換を続けて行くことができる。

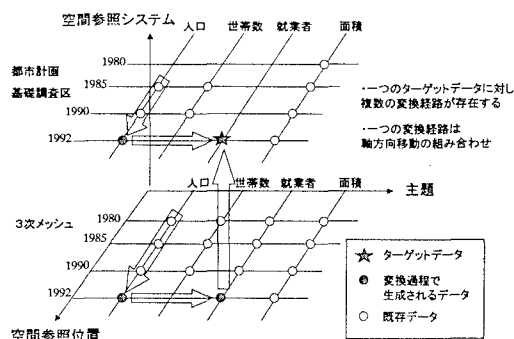


図-1 変換経路の例

このように参照データに着目してデータ変換経路を定義し、それに応じた誤差評価の方法を考える。

3.誤差評価法

(1) 誤差伝播の法則による誤差分布の推定

観測値 $x_1, x_2, \dots, x_n$ と推定値 y があり、これらの間に、

$$y = f(x_1, x_2, \dots, x_n)$$

のような関係があり、観測値の母分散をそれぞれ $\sigma_{x_1}^2, \sigma_{x_2}^2, \dots, \sigma_{x_n}^2$ とするとき、 y の母分散 σ_y^2 は、

$$\sigma_y^2 = \left(\frac{\partial f}{\partial x_1} \right)^2 \sigma_{x_1}^2 + \left(\frac{\partial f}{\partial x_2} \right)^2 \sigma_{x_2}^2 + \dots + \left(\frac{\partial f}{\partial x_n} \right)^2 \sigma_{x_n}^2$$

で、近似することができる。

(2) モンテカルロシミュレーションによる誤差分布の推定²⁾

$N(\mu, \sigma^2)$ に従うような観測値を複数発生させ、それぞれをデータ変換した複数の推定値の分布から、参照システム全体の誤差分布を得る。

4. データ変換と誤差評価法の適用

誤差評価の適用のために、S.61 の札幌市の都市計画基礎調査を基にデータ変換を行い、誤差を評価する。データ空間を図-2 に示す。ただし、存在するデータの中から変換に関係のあるものを取り出し表示した。時間位置情報は S.61 のみであるため、データ空間は主題と空間参照システムの 2 次元で表される。

具体的に変換に関しては、同じターゲットデータに対し、参照データの用い方が異なる 2 通りの変換を考えた。

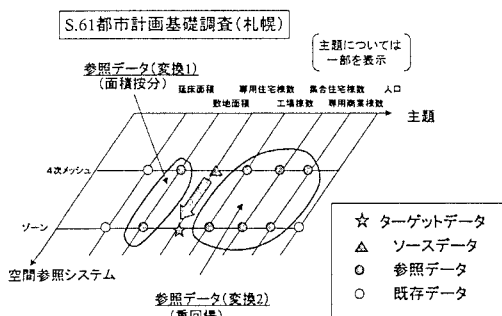


図-2 データ空間と変換経路

この調査は各建物を対象としたものであり、それぞれについてメッシュ番号とゾーン番号の記述がある。ここで、メッシュとゾーンの重ね合わせによって生じる詳細な空間参照システムを最小単位とすると、全てのメッシュとゾーンはこの最小単位をいくつか組み合わせることによって表現することができる。

今回は通常未知であるターゲットデータが既知で

あるため、それを真値として推定値との差をとり、その誤差の分布を得ることができる。しかしターゲットデータは観測誤差を確率的に含むため、この分布も確率的な変動の結果生じた 1 つの誤差分布に過ぎず、これから母分布を推定する必要がある。これを誤差評価法の精度を測る基準とし、他の方法と比較する。

以上の誤差評価の流れを、図-3 にまとめる。

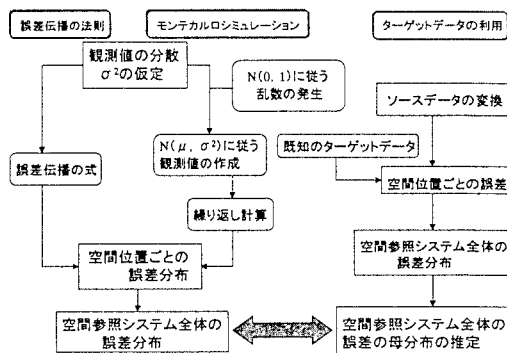


図-3 誤差評価の流れ

5. 結果

以上のデータ変換について、誤差伝播の法則とモンテカルロシミュレーションにより推定値の誤差を評価する。この結果については当日に示す。

観測誤差は結果の精度に影響を与える大きな要因であり、この観測誤差の与え方によって、どれくらい誤差の母分布に近い推定結果を得られるかということをはっきりと示し、またモンテカルロシミュレーションについてはその試行回数が結果の精度に関わってくるため、試行回数と結果の関係を検討する。

データ変換が決定したときに、どのように誤差を評価すべきかが簡単に決まるように、より多くのデータ変換に対して誤差を評価し、そのデータ変換法と誤差評価法との関係を明らかにしていくことが今後の課題となる。

(参考文献)

- 1) 杉木直, 中野光治, 内田敬, 宮本和明; 都市総合支援のための時空間データ変換システム, 地理情報システム講演論文集 Vol.8, pp221-226, 1999.
- 2) Fisher, P.F., M. Langford; Modeling the errors in areal interpolation between zonal systems by Monte Carlo simulation, Environment and Planning A, Vol.27, pp211-224, 1985.