IV-67 交通モデルと予測誤差について—信頼度分析の概念—

東京大学都市工学教室 正員 太田勝敏

1. 研究の背景と目的

交通の需要予測においては，通常，発生・分布・交通機関別分担・配分の4段階に分割し，それぞれのサブモデルを直列に連結した予測システムがとられている。このため，発生モデルのアウトプットが分布モデルのインプットのひとつとなるというような構造となっており，予測誤差はつぎつぎと累積的に伝播してゆくと考えられる。この研究は，このような予測誤差を予測システムの構造との関係で分析する方法として，工学の分野で発展した信頼度分析の概念の応用可能性について若干の考察を加えたものである。

2. 単一予測モデル式についての予測誤差分析[1]

$Y=F(X;A)$ 但し，Yは予測対象，$F(X;A)$はXを独立変数，Aをパラメータとする構造式，といった数学的モデルにより予測を行なう場合の誤差の原因は，① モデルの理論的構造の誤り，② 予測モデル式$F(X;A)$の誤差，③ パラメータ推定値$\hat{A}$の誤差，④ 独立変数推定値$\hat{X}$の誤差，にあるといえる。予測に伴なう最大の誤差原因は，①であるが，将来の予測時点でのモデルの理論的構造を予め推定することは困難である。一方，②,③,④の誤差については，基になったデータの精度，パラメータの推定方法，現状との適合度などからある程度それらの大いさを推定できる場合が多い。これらの3要因による予測誤差については，それらの誤差の確率分布・モーメント（分散等）・範囲（信頼限界）等の情報があれば，予測値$\hat{Y}$の誤差は，確率論における"変数変換"の問題として理論的には推定できる。このような予測誤差分析では①のモデルの理論的構造の誤り以外の原因について，具体的にX，Aおよびそれらの誤差の値を与えれば，個々の予測値の誤差を推定できる。

一方，いくつかのサブモデルを構成要素とする予測システムを設計する場合には，個々のサブモデルの精度（誤差）を表わす指標およびそれらのサブモデルを組み合わせた時のシステムの精度を解析する方法が必要となる。この種の問題については，上述した予測誤差分析は必ずしも適切ではないため，予測モデルないし予測システムの予測誤差（精度）を分析するマクロ的アプローチとして，信頼度の概念の適用が考えられる。

3. 信頼度（Reliability）の概念

信頼性工学の分野では，信頼度は，システムがその通常の稼動条件下で所定の期間満足に機能する確率として定義されている。[2] 予測の問題にこの概念を適用するために，先づ，ある予測値が一定の許容誤差の範囲内にある場合，その予測は"成功"と定義すると，"予測値の信頼度"はその予測の成功の確率と定義することができる。簡単のために，モデルの理論的構造に誤りはないものとし，また予測値に偏りがなく対称的確率分布に従うものとすれば，e_mを最大許容誤差とする場合，予測値$\hat{y}$の信頼度$r(\hat{y}:e_m)$は次式で表わされる：

$$r(\hat{y}:e_m)=p(-e_m\le e\le e_m)=p(\hat{y}-e_m\le y\le\hat{y}+e_m) \qquad (1)$$

但し，yはYの真の値，pは確率である。

この定義による信頼度は，信頼限界（Confidence Interval）の概念と密接な関係をもつことは明らかである。即ち，(1)は，$\hat{y}-e_m$を下限，$\hat{y}+e_m$を上限とする$\hat{y}$の信頼限界に対する信頼水準が$r(\hat{y}:e_m)$であると云える。

この概念を，予測値の信頼度から"予測モデルの信頼度"に拡張することができる。ここでは，モデルの通常の稼動条件を独立変数X_iの個々の値とその誤差，さらにそれらの組み合わせが通常の範囲にある場合と考え，そのような独立変数の状況において予測が成功する確率をモデルの信頼度と定義する。この予測モデルの信頼度は現状データについての予測成功率より事後的に推定する，あるいは，X_iを適宜与えて上述した予測誤差分

析により信頼度を多くのケースについて計算して推定することができる。

4. 予測システムの信頼度

次に，いくつかの予測モデルから構成される予測システムについての信頼度を考えるために，サブモデル M_i のパーフォーマンスは予測の成功 S_i，失敗 F_i のいづれかで表わされるものとすると，M_i の信頼度 R_i は予測の成功確率によって表わされる。

$$R_i = p(S_i) = 1 - p(F_i) \qquad (2)$$

一方，"予測システムの信頼度"は，その構成要素であるすべてのサブモデルの状態（F_i, S_i）についてのあらゆる組み合わせについて，システムの最終的予測値の誤差が許容範囲内にある（成功 S_s）か否（失敗 F_s）かを検討し，システムの予測が成功するサブモデルの組み合わせの確率と定義する。即ち，サブシステムが2個の場合には，システムの信頼度は次式のようになる：

$$R_s = p(S_s) = \sum p(S_s|*_1 *_2)\, p(*_1 *_2) = \sum p(S_s|*_1 *_2)\cdot p(*_1|*_2)\cdot p(*_2) \qquad (3)$$

但し，$*_i$ はサブモデル i の状態（S_i, F_i）を表わし，$\sum$ は，すべての可能な状態の組み合わせ（$S_1S_2, S_1F_2, F_1S_2, F_1F_2$）についての和である。

さて，上述した4段階の交通需要モデルのように，サブモデルが直列に結ばれている場合，サブモデルのいづれかが失敗するとその後に続くサブモデルも失敗し，システムも失敗すると仮定することができる。このような場合にはすべてのサブモデルが成功である場合に限り，システムが成功することになる。さらに，あるサブモデルが成功である場合には，その予測値と予測誤差はそれに続くサブモデルの通常の範囲内にあると仮定すれば，サブモデルの予測成功確率はその信頼度と等しくなるため，このような直列構造の予測システムの信頼度は，サブモデルの信頼度の積で表わすことができる：即ち，サブシステムが3個の場合には次式で示される：

$$R_s = p(S_s|S_1 S_2 S_3)\cdot p(S_1 S_2 S_3) = p(S_1 S_2 S_3) = p(S_3|S_2 S_1)\cdot p(S_2|S_1)\cdot p(S_1) = R_1 R_2 R_3 \qquad (4)$$

5. 予測システムの並列構造と信頼度

信頼性工学の分野では，同一の機能に対し複数の要素をあてるという並列構造は信頼度を高めることが知られている。予測システムの場合には，いくつかの代替的サブモデルがある場合（例えば，OD分布予測について，重力モデルとエントロピーモデル），それらを並列して用いると信頼度を高めることができる（例，信頼度が R_1, R_2 の独立した2つのサブモデルを並列して用いると，信頼度は $R_1 + R_2 - R_1R_2$ と向上する）。

一般に並列構造の場合，n ケのサブモデルの予測値 $\hat{y}_i$ を組み合わせて，より信頼度の高い新しい予測値 $\hat{y}$（$=\sum_{i=1}^{n} w_i \hat{y}_i$，$w_i$ は重味）を求めるには，$\sum_i w_i = 1.0$ の条件の下で $V(\hat{y}) = \sum_i \sum_j w_i w_j \mathrm{Cov}(\hat{y}_i, \hat{y}_j)$ を最小とする w_i を求めればよく，この解は $w = \frac{eA^{-1}}{eA^{-1}e'}$，但し，$w$ は重味ベクトル（$w_1, w_2 \cdots w_n$），e は単位ベクトル（1, 1, …, 1），e' は e の転置ベクトル，A^{-1} は $n \times n$ の共分散 $\mathrm{Cov}(\hat{y}_i, \hat{y}_j)$ マトリックス A の逆行列，であることが知られている。[3] この時，$V(\hat{y})$ は $1/eA^{-1}e'$ である。従って，特殊なケースとして，予測値が相互に無相関の場合，$w_i = \frac{1}{\sigma_i^2} \Big/ \sum_{i=1}^{n} \frac{1}{\sigma_i^2}$ で，$V(\hat{y}) = 1 \Big/ \sum_{i=1}^{n} \frac{1}{\sigma_i^2}$（但し，$\sigma_i^2$ は $\hat{y}_i$ の分散）となり，新しい予測値 $\hat{y}$ の分散は，各サブモデルの予測値 $\hat{y}_i$ の分散のいづれよりも小さいことが容易に示される。

さらに複雑な構造の予測システムについても (3)式の原理により信頼度を計算することができ，信頼度分析はこのような意味で，予測システムの構造の分析や設計に利用できる。

以上のように信頼度分析は，予測の行なわれる一般的状況においてモデルないし予測システムの予測の成否を分析するマクロ的手法であり，一方，予測誤差分析は，誤差の原因に基づきその伝播過程を分析することによって予測値の誤差を個別に分析するミクロ的手法であるが，両者は相互に補完的な誤差分析の手法である。

参考文献

1) 太田勝敏「交通需要予測の誤差分析」第28回土木学会年次学術講演会概要集．1973年．
2) M.L. Schooman, Probablistic Reliability: An Engineering Approach, McGraw-Hill. 1968年.
3) Des Raj, Sampling Theory, McGraw-Hill. 1968年.