

模糊模式识别 在交通量预测模型选择中的应用

晏 秋 杜 文

西南交通大学，交通运输学院，成都 610031

摘 要：本文将模糊模式识别应用于交通规划预测模型的评价。以出行受约束的重力模型与双约束重力模型为例，说明常用的一些检测方法存在的不严密性，并用模糊模式识别方法定量确定预测结果与实际出行量的拟合程度。本文的分析结果能为交通规划预测模型的选择提供可靠的依据，并为预测结果的评价提供新的方法。

关键字：交通量预测；模糊模式识别；重力模型法；贴近度

中图分类号：U491

文章编号：A

文章编号：1672-4747(2006)01-0041-04

Application of the Fuzzy Model Identification in the Selection of Traffic Volume Forecast Models

YAN Qiu DU Wen

College of Traffic and Transportatiion ,

Southwest Jiaotong University , Chengdu 610031 , China

Abstract : This paper introduces the application of the Fuzzy Model Identification to the evaluation of traffic planning forecast models , and by taking as examples the single constrained gravity model and the doubly constrained gravity model , clarifies the inaccuracy of some common detection methods simultaneously , defines the fitting degree of the prediction results to the origin destination volume with the Fuzzy Model Identification approach. The analysis results of this paper can serve as the basis for the option of the optimal traffic planning forecast models , and this method provides a new approach to the evaluation of the prediction results.

收稿日期：2005-06-29.

作者简介：晏 秋（1970-），女，四川邛崃人，西南交通大学交通运输学院讲师，在读博士。

Key words : Traffic volume forecast , fuzzy model identification , gravity method , fitting degree

0 引言

在交通量预测中,我们常靠数据拟合或常用的经验模型来决定模型的构建形式。我们可以选择的经验模型较多,比如,在公路交通量的预测中,对于交通量增长预测,常采用一元或二元线性回归模型,即认为交通量的增长与经济增长等因素存在着线性关系;也有人认为,根据过去的的数据,其增长特性符合指数规律,因此,应采用指数模型来进行交通量的预测。

对所选择的模型是否合理常采用统计学的知识来决定,如线性回归模型拟合良好的统计标准是相关系数 τ 接近1,重力模型的验证也常用相关系数 τ 、 x^2 检验或分布加权平均时间的相对误差等。只要在要求的精度范围内,一般我们就认为可以选择此模型。

1 模糊模式识别

我们希望所选择的模型应该是对实际问题最真实的反映,到底我们的实际数据更接近于哪一种模型,可以运用模糊模式识别来判断调查数据真正反映的函数关系(B)与各类可选模型(A_k)之间的贴近度来决定模型的最佳选择。比如,贴近度的计算可以采用式(1)^[1]。

$$(B \cdot A_k) = \frac{\sum(\mu_B(x) \wedge \mu_{A_k}(x))}{\sum(\mu_B(x) \vee \mu_{A_k}(x))} \quad (1)$$

式中, $\mu_B(x)$ 、 $\mu_{A_k}(x)$ 分别为模式 B 、 A_k 的隶属函数; x 属于论域 x ; $(B \cdot A_k)$ 为介于[0, 1]之间的数,即表示模式 B 与模式 A_k 的贴近程度。

下面我们用四阶段法的出行分布预测常用的重力模型法来说明。

2 重力模型法的两种模型及算例

重力模型法又可分为四种类型的出行分布模型。

其中包括出行受约束的重力模型(公式(2))和双约束重力模型(公式(3))^[2]。为说明方便,以下分别简称模型1和模型2,并分别将两种模型所表示的函数关系定义为模式 A_1 和模式 A_2 。

$$\hat{T}_{ij1} = \frac{K_i P_i A_j}{f(t_{ij})} \quad (2)$$

式中, \hat{T}_{ij1} 为模型1计算的出行分布量; K_i 调整系数, $K_i = \left[\sum_j \frac{A_j}{f(t_{ij})} \right]^{-1}$,出行吸引模型计算的和调查数值不相等时,可用控制 K_i 值为加以约束,使其出行总数相等; P_i 、 A_j 分别表示出行产生与出行吸引; t_{ij} 是交通区 i 与交通区 j 之间的交通阻抗参数,如两区间交通的距离、时间或费用等; $f(t_{ij})$ 是区间阻抗函数,最常用 $f(t_{ij}) = t_{ij}^\alpha$ 形式, α 为待调整的系数^[2]。

$$\hat{T}_{ij2} = \frac{K_i K'_j P_i A_j}{f(t_{ij})} \quad (3)$$

式中, \hat{T}_{ij2} 为模型2计算的出行分布量; K_i 和 K'_j 为调整系数, $K_i = \left[\sum_j \frac{K'_j A_j}{f(t_{ij})} \right]^{-1}$; $K'_j = \left[\sum_i \frac{K_i P_i}{f(t_{ij})} \right]^{-1}$ 。在出行产生和出行吸引的模型计算与调查数据都不相等时,用控制 K_i 和 K'_j 的互相约束来保证出行总数的相等; $f(t_{ij})$ 意义同公式(2)^[2]。

假设一个调查地区分成2个城郊居住分区(1分区和2分区)及3个就业中心(3、4、5分区),家庭到工作地点的日出行模式表示为调查起讫点矩阵见表1,区间距离用交通分区之间的交通时间 t_{ij} 表示,

表1 OD矩阵的调查分布

Tab.1 surveyed OD matrix

起 点	讫 点			P_i
	3	4	5	
1	150	100	50	300
2	400	100	200	700
A_j	550	200	250	1 000

如图 1 所示。选用交通阻抗函数为距离的幂函数, 即 $f(t_{ij}) = t_{ij}^\alpha$ (指数 α 采用试算法, 可先假定为 1.0), 并假设用模型 1 与模型 2 计算得出的出行分布矩阵分别为 OD_1 和 OD_2 。

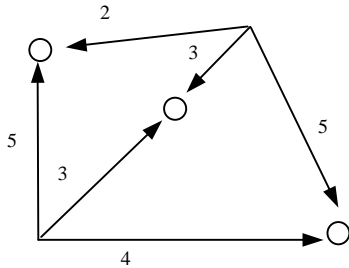


图 1 调查分区及小区间出行的交通时间 (min)

Fig.1 Travel-time between surveyed zones

当我们采用模型 1 时, 按照精度要求, 即 OD_1 分布中加权平均行程时间 $\bar{t}(OD_1)$ 与 OD 分布的加权平均行程时间 $\bar{t}(OD)$ 的相对误差不超过 3%, 通过对指数 α 的不断改动, 最后得到 OD_1 分布如表 2^[3]。

表 2 用模型 1 计算的分布表 (OD_1)

Tab.2 OD_1 matrix produced by model 1

起点	讫点			P_i
	3	4	5	
1	165	90	45	300
2	449	98	153	700
A_j	614	188	198	1 000

当采用模型 2 时, 经过 α, K_i, K'_j 的不断修正, 最后收敛 (可信度为 95%), 得到计算结果 OD_2 分布如表 3^[2]。

表 3 用模型 2 计算的分布表 (OD_2)

Tab.3 OD_2 matrix produced by model 2

起点	讫点			P_i
	3	4	5	
1	147	96	57	300
2	403	104	193	700
A_j	550	200	250	1 000

3 模型不同检验方法对比

(1) χ^2 检验

这是重力模型常使用的模型验证方法。对模型和调查的起讫点矩阵的检验为:

$$\chi^2 = \sum_i \sum_j \frac{(T_{ij} - \hat{T}_{ij})^2}{T_{ij}} \quad (4)$$

式中 T_{ij} 为 OD 调查分布表中交通区 i 到 j 出行量, \hat{T}_{ij} 为模型得出的交通区 i 到 j 出行量。

计算 χ^2 值并与其临界值比较, 如果计算的 χ^2 比临界值小, 说明是一个拟合良好的模型。临界值可由自由度和显著性水平两者查临界值分布标准表^[2]得到。在本例中, 自由度 $DF = 2 \times 3 - 1 = 5$, 显著性水平 $\alpha = 100\% - 95\% = 5\%$, 查表可以获得 ρ^2 (χ^2 的临界值) = 11.07。

$$\text{对模型 1, } \chi_1^2 = \sum_i \sum_j \frac{(T_{ij} - \hat{T}_{ij1})^2}{T_{ij}} = 20.0875 > \rho^2, \text{ 不}$$

满足精度要求;

$$\text{对模型 2, } \chi_2^2 = \sum_i \sum_j \frac{(T_{ij} - \hat{T}_{ij2})^2}{T_{ij}} = 1.5449 < \rho^2, \text{ 满}$$

足精度要求。

(2) 分布加权平均时间误差 (min)

精度要求: 模型计算的 OD 分布中加权平均行程时间与调查 OD 分布中加权平均行程时间的相对误差不应超过 3%。

OD 分布加权平均行程时间为:

$$\bar{t}(OD) = \frac{5(100+50) + 4 \times 200 + 3(150+400) + 2 \times 100}{1000} = 3.4$$

对模型 1, OD_1 分布加权平均时间为

$$\bar{t}(OD_1) = \frac{5(98+45) + 4 \times 153 + 3(165+449) + 2 \times 90}{1000} = 3.35$$

两者的相对误差为 $0.015 < 0.03$, 因为, 此时假定 $\alpha = 1$, 因此, 参数 α 被确定。

对模型 2, OD_2 分布加权平均时间为

$$\begin{aligned} \bar{t}(OD_2) &= \frac{5(56.7+104.3) + 4 \times 193.3 + 3(147.6+402.4) + 2 \times 95.7}{1000} \\ &= 3.4196 \end{aligned}$$

两者的相对误差为 $0.005\ 764\ 705 < 0.03$ 。

从以上的检验可以认为模型 2 比模型 1 对数据的拟合更好,但是,至少还存在以下两个问题:一是我们发现不同的检验方法对模型的认可结果不同,如模型 2 当采用 χ^2 检验时,是不能被认可的;二是模型与数据所反映的真实关系的贴近程度怎样,这两种检验方法都不能明确回答。而这些都可以通过调查数据反映的函数关系与模型的贴近度来准确地反映出来。

(3) 贴近度的计算

按照模糊模式识别,比较调查数据所反映的函数关系,即模式 \tilde{B} 与模式 A_1 和模式 A_2 贴近度。模式 \tilde{B} 、 A_1 和 A_2 所表示的数据关系分别用 OD 、 OD_1 和 OD_2 矩阵的各元素按公式(5)正规化后表示。

$$x_{ij} = \frac{T_{ij}}{\sum_i \sum_j T_{ij}} \quad x_{ijk} = \frac{\hat{T}_{ijk}}{\sum_i \sum_j \hat{T}_{ijk}} \quad k=1,2 \quad (5)$$

式中, x_i , x_{ijk} 分别表示模式 \tilde{B} 与模型 A_k 的论域中的数;其它符号含义同式(2)~式(4)。

用公式(1)表示贴近度,则有:

$$(\tilde{B} \cdot A_k) = \frac{\sum_{i=1}^2 \sum_{j=3}^5 (x_{ij} \wedge x_{ijk})}{\sum_{i=1}^2 \sum_{j=3}^5 (x_{ij} \vee x_{ijk})} \quad k=1,2 \quad (6)$$

参考文献

[1] 黄天民. 应用模糊数学(讲义)[M]. 成都:西南交通大学, 2005: 27-32.
 [2] 肖秋生等. 城市交通规划[M]. 北京:人民交通出版社, 1999: 84-85, 90-100.

式中符号含义同公式(1)。

经计算得: $(\tilde{B} \cdot A_1) = 0.879\ 7$; $(\tilde{B} \cdot A_2) = 0.973\ 6$

即模型 1、模型 2 与实际交通分布的贴近度分别为 0.879 7 和 0.973 6。

4 结束语

对于本例, χ^2 检验和分布加权平均时间误差检验方法对模型 1 和 2 得到的 OD 出行分布所得出的检验结果是不同的,从而会使我们对模型结果的可靠性产生怀疑,同时也很难确定预测结果对实际出行量的拟合程度。而用模糊模式识别方法则较准确地反映出模型 1 和 2 与实际出行量的贴近度,定量地对不同模型结果进行比较,使我们更科学地对预测结果进行评价,并作出合理的选择。

另外,四阶段交通量预测法中存在着很多模型和模型参数的最优选择的问题,比如模型 1、2 中区间阻抗函数 $f(t_{ij})$ 就可分为五种基本形态^[2],设其中的一种形态 k 对应的模型为模式 A_k ,通过模糊模式识别判断调查数据与各模式的贴近度,从而判断最佳模型或确定选择这些形态是否合理。

[3] 刘建军. 交通工程学基础[M]. 北京:人民交通出版社, 2000: 182-183.
 [4] 林方城. 若干常用模糊数学方法的实质[J]. 系统工程, 1998, (11): 38-4.