

基于动态加权融合模型的快速路交通量预测

许可心 林婧 杨泽鹏

【摘要】在历史数据量有限的条件下，为提高城市快速路交通量预测的准确性，本文提出基于动态加权融合模型的快速路交通量预测方法。在城市快速路历史交通量数据的基础上，采用自回归差分移动平均（ARIMA）模型与三次指数平滑模型分别对未来交通量进行预测，通过动态权重融合算法对两个模型的预测值进行融合，得到最终的预测交通量。选取重庆市某快速路断面的历史交通量，构建融合模型并对预测结果进行评价，得出模型预测值与实际值基本一致，平均绝对误差为 21.11，平均绝对百分比误差为 8.30%。结果表明：动态加权融合模型对于快速路交通量预测精度较高，可为城市快速路交通规划及管理提供重要参考。

【关键词】ARIMA 模型；三次指数平滑模型；交通量预测；时间序列

0 引言

城市快速路作为实现高效、大容量交通流动的关键，其为城市长距离过境交通提供服务，同时承担着优化城市交通网络结构的重要任务。交通量预测是根据历史与实时的交通量数据对未来的交通量进行预测，准确的预测可为交通领域的多方面工作提供数据支撑，有助于交通拥堵的缓解，以及出行效率的提升^[1]。因此，城市快速路交通量的预测对于评估未来交通需求变化、引导城市空间布局等具有重要意义。

近些年，国内外均在交通量预测方面开展了广泛研究，常用预测方法主要有四类。第一类是统计学方法，ZHOU T等人^[2]对混合双卡尔曼滤波器方法进行优化，预测结果具有较高的准确性与及时性；Tiantian L等^[3]构建了预测流量区间的自回归集成移动平均模型，能够预测流量峰值的需求变化。第二类是机器学习方法，李巧茹等^[4]将支持向量机与自适应时空数据融合，并在此基础上建立了短时交通流量预测模型；此外还包括使用决策树、随机森林等机器学习方法进行交通量预测。第三类是深度学习方法，张海亮^[5]利用了改进GA-BP的模糊神经网络方法，依据该方法所建立的预测模型，提高了在高速公路条件下较大流量预测方面的准确性；刘永乐等^[6]提出的卷积-双向长短期记忆模型，同时考虑到时间及空间因素，该方法更能体现出交通量随时空而不断变化的特征。第四类则是混合模型方法，文献^[7-10]分别将时间序列模型与相关向量机、支持向量机相结合；黄欣等^[11]结合两种神经网络模型，构建了AGRU-Trans融合预测模型；李松江等^[12]以神经网络为基础，结合自适应加权算法提出了数据融合模型。

上述研究大多基于复杂的机器学习或深度学习，通常需要大量的历史数据进行训练以避免模型过拟合，忽略了数据量较少情况下的实际应用。而时间序列模型能够在相对较少的数

据下进行有效的训练和预测，这对于历史数据量有限的交通量预测十分重要。此外，尽管时间序列模型预测数据突变的能力有所欠缺，但在实际应用场景中，根据动态权重相融合的时间序列模型，仍然能够提供较为准确的预测结果。

本文选用重庆市某快速路断面2023年10月23日0点至10月24日6点55分的交通量历史数据（以每5分钟为间隔），提出了一种基于动态加权融合模型的快速路交通量预测方法，通过模型预测及效果评价，验证了融合模型在该快速路的交通量预测中具有较高精度。

1 交通量预测模型

1.1 自回归差分移动平均模型

1.1.1 模型公式

自回归差分移动平均模型（Autoregressive Integrated Moving Average Model）又称为ARIMA模型。是一种重要的时间序列预测模型，模型通过数据的自相关性和差分方式，提取出数据潜在的时间序列模式，并利用这些模式来预测未来的数据。即，利用数据本身的历史信息来预测未来。

ARIMA模型主要由三部分组成，包括了自回归模型、差分过程以及移动平均模型。其中，自回归模型（AR）部分用于处理时间序列的自回归部分，它考虑了过去若干时期的观测值对当前值的影响；差分过程（I）部分用于将时间序列转换为平稳的，通过该部分的处理，能够对序列中的趋势性、季节性等因素进行消除；移动平均模型（MA）部分考虑了过去的预测误差对当前值的影响，用于处理时间序列的移动平均部分。

自回归模型公式为：

$$Y_t = \sum_{i=1}^p \alpha_i Y_{t-i} + \varepsilon_t + m \quad (1)$$

其中， Y_t 为模型的预测量； Y_{t-i} 为过去的时间断点数值； α_i 为过去时间断点数值增加权重后的系数； ε_t 为模型产生的偏差值； m 为常量； p 为自回归模型阶数； i 为过去时间断点的间隔数量。

移动平均模型公式为：

$$Y_t = \sum_{i=1}^q \beta_i \varepsilon_{t-i} + \varepsilon_t + n \quad (2)$$

其中， ε_{t-i} 为过去时间断点 i 上的偏差值； β_i 为偏差值增加权重后的系数； n 为常量； q 为移动平均模型阶数。

由于ARIMA模型将自回归、差分处理、移动平均三个部分进行结合，保证了该模型对数据趋势变化的预测，同时又能够完成对某些突变数据的处理。由此，ARIMA模型的一般表达式为：

$$ARIMA(p, d, q) \quad (3)$$

其中, p 为自回归模型阶数; d 为差分模型阶数; q 为移动平均模型阶数。

1.1.2 模型构建方法

自回归差分移动平均模型的构建方法主要有三个步骤,分别为原始序列趋势分析、序列平稳化、模型参数确定。

(1) 原始序列趋势分析

构建ARIMA模型进行预测时,需要保证用于预测的数据序列处于平稳状态。对于原始序列是否为平稳状态,以及是否具有随机性、季节性、周期性的变化趋势等,常使用图形化检验方法对样本数据的时间序列趋势进行分析,判断原始序列是否满足平稳化要求。

通过对样本数据序列图的观察,对其发展变化的规律总体特征、趋势进行分析。例如,若该时间序列呈现上升或者下降趋势,分析其是否具有周期性;若呈现竖直方向上的波动,分析其波动是否有规律性,或者序列中是否存在异常值点等。一个平稳状态的时间序列将呈现出水平方向上的稳定推进,同时在竖直方向上出现一定范围内的上下波动。

(2) 序列平稳化

在时间序列分析中,差分作为将非平稳序列转化为平稳序列的常用方法,能够减小或消除时间序列的趋势和季节性变化。如序列有长期趋势或季节性趋势等特征,则为不平稳序列,应根据数据特征,对其进行一般或季节性差分运算。所以在利用模型进行预测前,需要通过数据序列图或单位根(ADF)检验等方式,确定出最小的差分阶数,使得数据序列达到平稳化要求。

(3) 模型参数确定

自相关函数(ACF)表示,时间序列与其经过某些阶数滞后形成的序列之间的相关性。偏自相关函数(PACF)表示,在其它序列给定情况下两序列条件相关性的度量函数。通过对上述两种函数图像的分析,可以初步选取ARIMA(p, d, q)中参数 p 、 q 的取值,同时根据序列平稳化的差分次数确定参数 d 的大致取值。通常情况下,参数 d 的取值不超过2,以避免数据发生过度拟合的现象。

通过不断地调试,根据模型赤池信息准则(AIC)值及贝叶斯信息准则(BIC)值的对比分析,对参数进行调整,以确定最终的模型参数取值。两个准则均被用于模型选择,考虑了模型的复杂度和拟合优度。AIC准则常用于衡量统计模型拟合的优良性,类似的BIC准则,其参数相关的惩罚项更大,有助于避免过度拟合的问题。AIC值及BIC值越小,模型越符合要求。

1.2 三次指数平滑模型

1.2.1 指数平滑法

指数平滑法(Exponential Smoothing)常用于对时间序列进行平滑处理并预测。其主要

原理是在历史数据的基础上完成对未来数据的预测,在计算过程中考虑到时间对数据发展趋势的影响,距离预测时刻时间越近的数据权重取值越大,反之则权重较小,以此探究该时间序列数据的发展趋势与变化规律。

指数平滑法主要包含三类,一次、二次、三次指数平滑,分别考虑了 α 、 β 、 γ 三个参数,其中参数 α 对应数据,参数 β 对应数据变化的趋势,参数 γ 对应数据变化的周期性、季节性。一次指数平滑法仅针对时间序列数据本身的特点,所以其公式中仅有参数 α ;二次指数平滑法在数据特征的基础上,增加了时间序列变化的趋势,所以公式中考虑到参数 α 和参数 β ;三次指数平滑法则进一步考虑了时间序列变化的周期性、季节性,所以公式中同时考虑了三个参数 α 、 β 、 γ 。

1.2.2 霍尔特-温特加法模型

霍尔特-温特指数平滑法(Holt-Winters Exponential Smoothing)又称为三次指数平滑法,在二次指数平滑法的基础上增加了季节性因素,该方法引入了 α 、 β 、 γ 三个参数,适用于具有趋势及季节性的时间序列。该方法中包含有三个平滑方程,分别是水平平滑、趋势平滑和季节平滑,各项平滑方程的具体公式如下所示:

$$l_t = \alpha(x_t - s_{t-m}) + (1 - \alpha)(l_{t-1} + b_{t-1}) \quad (4)$$

$$b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1} \quad (5)$$

$$s_t = \gamma(x_t - l_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m} \quad (6)$$

其中, x_t 为时刻 t 的实际观测值; l_t 为时刻 t 的预估水平; b_t 为时刻 t 的预测趋势; s_t 为时刻 t 的季节性成分; m 为周期长度; α 为水平平滑参数; β 为趋势平滑参数; γ 为季节平滑参数。

通过计算上述各项参数,得到霍尔特-温特加法模型:

$$\hat{x}_{t+h} = l_t + hb_t + s_{t+h-m(k+1)}, k = \left\lfloor \frac{h-1}{m} \right\rfloor \quad (7)$$

其中, h 为预测超前期数; $\hat{x}_{t+h}$ 为第 h 期的预测值。

1.3 动态加权融合模型

1.3.1 模型融合方法

由于不同预测模型各有优缺点,各项子模型对数据的适应性有所差别,在不同时刻各项子模型对于时间序列数据的预测精度会存在一定偏差。为了在进行综合预测时,使预测精度更高的子模型发挥主要作用,即赋予预测值与实际值相对误差较小的子模型更大的权重,本文提出了一种基于动态加权的模型融合方法,根据不同时段的具体情况,为两个子模型动态的分配权重。该方法的主要步骤如下:

- (1) 选取两个适用性较好且预测精度满足条件的预测模型,作为融合模型的子模型。

(2) 分别通过两个子模型对时间序列进行预测，得到第 i 个模型在 $\hat{x}_i(t)$ 、 $\hat{x}_i(t-1)$ 、 $\hat{x}_i(t-2)$ 三个时刻的预测值，同时预测出 $t+1$ 时刻的预测值 $\hat{x}_i(t+1)$ 。

(3) 通过子模型预测结果的相对误差计算出对应的初始权重，再根据过去三个时刻的权重值计算得到下一时刻的动态权重，将第 i 个模型在 $t+1$ 时刻的预测值 $\hat{x}_i(t+1)$ 按照动态权重进行加权，最终得到融合模型在 $t+1$ 时刻的预测值，公式如下：

$$x(t+1) = \sum_{i=1}^2 [w_i^*(t+1) \cdot \hat{x}_i(t+1)] \quad (8)$$

其中， $w_i^*(t+1)$ 为第 i 个模型在 $t+1$ 时刻的动态权重； $\hat{x}_i(t+1)$ 为第 i 个模型在 $t+1$ 时刻的预测值。

1.3.2 动态权重确定

融合模型的关键在于两个子模型间权重取值的确定。传统的权重计算方法大多存在判断较为主观的问题，且权重赋值确定后基本不再调整。同时，对于预测精度的问题，尽管可以实时判断预测值与实际值的相对误差，并且将较大的权重值赋予预测准确性更高的模型，但由于子模型间不同的适应性，某个模型在 t 时刻预测精度更高，却无法保证其在 $t+1$ 时刻同样有较高的精度。为了能够灵活调整融合模型的权重分配，考虑过去预测精度的同时又不至于权重修正得过于频繁，引入了动态权重的概念。

分别计算各项子模型预测结果的相对误差，公式如下：

$$e_i(t) = \left| \frac{x(t) - \hat{x}_i(t)}{x(t)} \right|, \quad i=1,2 \quad (9)$$

其中， $x(t)$ 为 t 时刻的实际观测值； $\hat{x}_i(t)$ 为第 i 个模型在 t 时刻的预测值； $e_i(t)$ 为第 i 个模型在 t 时刻预测值相对误差的绝对值。

利用反比例法求出子模型在某时刻的初始权重，即对子模型预测结果的相对误差取倒数。并对两个子模型的初始权重进行归一化处理，以保证权重之和为1。具体公式如下：

$$w_i(t) = \frac{1}{e_i(t)} \quad (10)$$

$$w'_i(t) = \frac{w_i(t)}{\sum_{i=1}^2 w_i(t)} \quad (11)$$

其中， $w_i(t)$ 为第 i 个模型在 t 时刻的初始权重； $w'_i(t)$ 为第 i 个模型在 t 时刻归一化的权重。

动态权重计算时，同时考虑了子模型在 t 时刻、 $t-1$ 时刻和 $t-2$ 时刻的权重取值，在此基础上计算得到各子模型在融合时的动态权重，具体公式如下：

$$w_i^*(t+1) = \frac{[w'_i(t) + w'_i(t-1) + w'_i(t-2)]}{3} \quad (12)$$

其中， $w_i^*(t+1)$ 为第 i 个模型在预测 $t+1$ 时刻的动态权重。

2 模型预测及效果评价

将重庆市某快速路断面2023年10月23日0点至2023年10月24日6点55分的交通量数据,利用Excel软件依据时间顺序建立历史数据库。该断面交通量数据的采集以每五分钟为间隔,并按照采集时间的先后顺序对交通量数据进行编号。

2.1 自回归差分移动平均预测

2.1.1 数据平稳化处理

根据采集的交通量历史数据绘制序列图,如图1(a)所示。该断面交通量受到早、晚高峰的影响,交通量从第78组数据、204组数据(约10月23日6:30、17:00)开始明显增长。同时,根据该序列图的整体趋势变化分析可知,原始数据的时间序列并不平稳。对原始序列进行差分处理,能够将其从不平稳状态转变为平稳状态,以满足数据平稳化的要求。将交通量数据一阶差分处理后的时间序列图,如图1(b)所示。可见该时间序列在水平方向上平稳前移,竖直方向上稳定波动。初步判断经一阶差分后交通量数据的时间序列达到平稳状态,参数 d 暂设定为1。

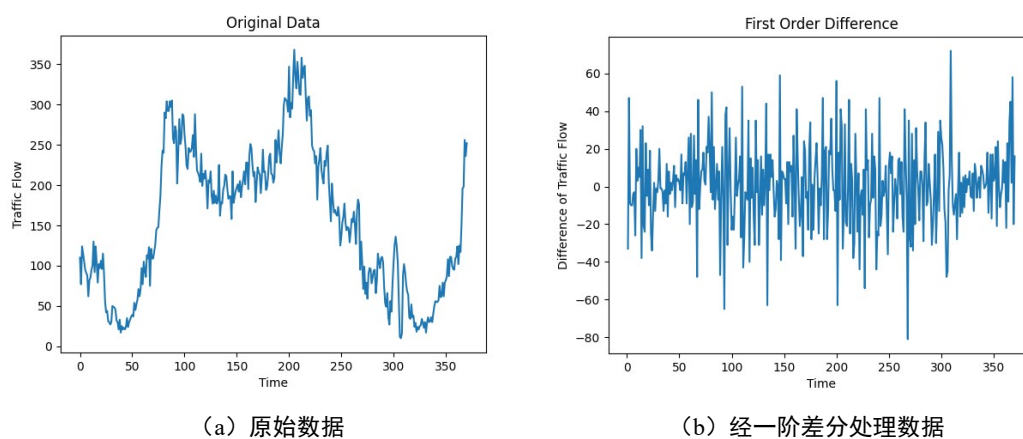


图 1 交通量历史数据时间序列图

2.1.2 模型参数取值

对原始序列进行一次差分处理后,分别绘制得到自相关函数图以及偏自相关函数图,如图2、图3所示。根据ACF、PACF图像特征可初步确定模型中参数 p 、 q 的取值。分析ACF、PACF图得到,两图像均从1阶之后呈现出拖尾现象。由此,初步将参数 p 、 q 的取值定为1。

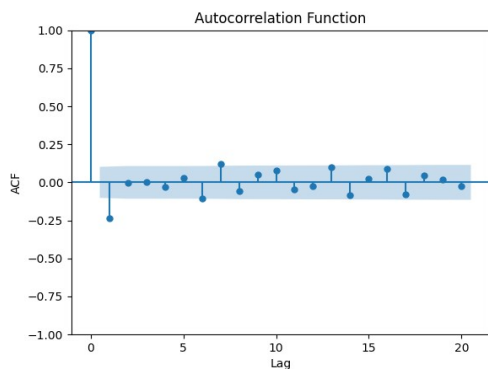


图 2 交通量历史数据经一阶差分后自相关图

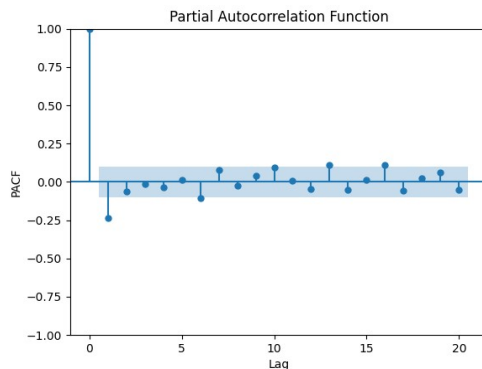


图 3 交通量历史数据经一阶差分后偏自相关图

2.1.3 预测结果

在初步确定各参数的取值基础上，利用 p 、 d 、 q 可能的参数组合分别进行模型构建，并通过不同组合计算所得到的AIC值、BIC值的对比分析，确定最终的预测模型。考虑合适的模型有ARIMA（1,1,0）、ARIMA（0,1,1）、ARIMA（1,1,1）三种，比较各模型间的AIC值以及BIC值，具体指标计算结果，如表1所示。

表 1 模型 AIC、BIC 指数对比

模型	ARIMA(1, 1, 0)	ARIMA(0, 1, 1)	ARIMA(1, 1, 1)
AIC	3319.91	3318.29	3320.26
BIC	3327.74	3326.11	3332.00

三个模型中ARIMA（0,1,1）的AIC值、BIC值分别为3318.29、3326.11，两个指标均为最小，是较为理想的模型。确定模型 p 、 d 、 q 参数取值分别为0、1、1，最终构建ARIMA模型如下所示。

$$ARIMA(0, 1, 1) \quad (13)$$

利用ARIMA（0,1,1）模型对该快速路2023年10月24日早高峰部分时段（7:00-8:40）的交通量进行预测。具体预测结果，如表2所示。

表 2 自回归差分移动平均法预测结果

序号	交通量	序号	交通量	序号	交通量
1	248	8	272	15	289
2	280	9	264	16	267
3	300	10	236	17	282
4	284	11	247	18	279
5	303	12	219	19	247
6	304	13	244	20	230
7	291	14	254	21	226

2.2 三次指数平滑预测

利用python代码对历史交通量的时间序列进行预测。首先，读取历史数据库中的交通量数据。其次，在历史数据的基础上构建指数平滑模型，设置模型的趋势和季节性为加性，即引入霍尔特-温特加法模型。此时水平平滑参数 α 、趋势平滑参数 β 和季节平滑参数 γ 均为待定参数，考虑到交通量在全天存在早、晚高峰和平峰的周期变化，设置该模型季节性周期为24h。最后，根据历史数据对构建模型进行拟合，计算 α 、 β 、 γ 三个参数，以及 l_t 、 b_t 、 s_t 的取值，确定最终预测模型并对下一时刻的交通量进行预测。预测结果如表3所示。

表 3 三次指数平滑法预测结果

序号	交通量	序号	交通量	序号	交通量
1	252	8	272	15	289
2	297	9	265	16	278
3	323	10	247	17	282
4	284	11	248	18	273
5	297	12	223	19	249
6	305	13	243	20	231
7	278	14	244	21	227

2.3 融合模型预测

通过公式（9~11）分别计算两个子模型预测数据与实际数据的相对误差，利用反比例法得到子模型在某时刻对应的初始权重并对其归一化处理。根据公式（12）分别得到两个子模型对应时刻的动态融合权重。最后，根据公式（8）得到融合模型预测结果，如表4所示。表中，模型1为自回归差分移动平均模型，模型2为三次指数平滑模型。

表 4 融合模型预测结果

序号	模型1预测值	权重1	模型2预测值	权重2	融合模型预测值
4	284	0.47	284	0.53	284
5	303	0.48	297	0.52	300
6	304	0.66	305	0.34	304
7	291	0.60	278	0.40	286
8	272	0.55	272	0.45	272
9	264	0.45	265	0.55	265
10	236	0.44	247	0.56	242
11	247	0.41	248	0.59	248
12	219	0.41	223	0.59	221
13	244	0.40	243	0.60	243

14	254	0.50	244	0.50	249
15	289	0.51	289	0.49	289
16	267	0.52	278	0.48	272
17	282	0.45	282	0.55	282
18	279	0.44	273	0.56	276
19	247	0.42	249	0.58	248
20	230	0.49	231	0.51	231
21	226	0.51	227	0.49	226

2.4 预测效果评价

利用动态加权融合模型对未来交通量进行预测,将其预测结果与采集数据进行对比,通过式(14)、(15)计算预测值与实际值的平均绝对误差(MAE)和平均绝对百分比误差(MAPE),以此评价文本所构建的融合模型的预测精度是否满足要求。计算得到融合模型各个时刻的预测结果误差如表5所示。

平均绝对误差:

$$MAE = \frac{1}{n} \sum_{j=1}^n |x_j - \hat{x}_j| \quad (14)$$

平均绝对百分比误差:

$$MAPE = \frac{1}{n} \sum_{j=1}^n \left| \frac{x_j - \hat{x}_j}{x_j} \right| \times 100\% \quad (15)$$

其中, x_j 为第 j 个时刻的实际值; $\hat{x}_j$ 为第 j 个时刻的预测值。

表 5 融合模型预测结果误差

序号	实际值	预测值	MAE	MAPE (%)
4	310	284	26	8.39
5	305	300	5	1.64
6	287	304	17	5.92
7	266	286	20	7.52
8	262	272	10	3.82
9	228	265	37	16.23
10	251	242	9	3.59
11	211	248	37	17.54
12	253	221	32	12.65
13	258	243	15	5.81
14	300	249	51	17.00
15	261	289	28	10.73
16	287	272	15	5.23

17	278	282	4	1.44
18	237	276	39	16.46
19	225	248	23	10.22
20	225	231	6	2.67
21	232	226	6	2.59

绘制融合模型预测值与实际值对比折线图，如图4所示。分析对比图可知，融合模型预测得到的交通量变化趋势与实际交通量基本吻合。同时，计算得到整体预测值与实际值两者间平均绝对误差为21.11，平均绝对百分比误差为8.30 %。

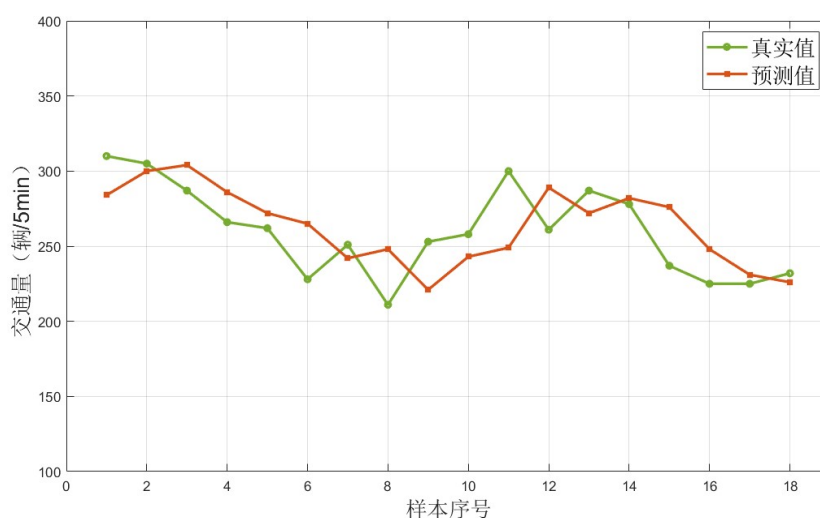


图 4 实际值与融合模型预测值对比图

尽管由于时间序列预测模型的局限性，导致其难以对交通量的突变情况（突然增加或减少）进行准确及时的预测，在某些数据转折点可能存在预测延迟的现象。但是该融合模型的预测趋势与实际趋势基本吻合，且整体相对误差在可接受范围内（10%以内）。由此可见，本文构建的动态加权融合模型的预测精度较好。

3 结语

本文对城市快速路交通量数据分析得到，受到早、晚高峰等因素影响，交通量数据具有时间序列性及周期性。因此，选择对时间序列具有较好适用性的自回归差分移动平均模型和三次指数平滑模型。以两个预测子模型作为基础，结合动态权重计算方法，构建了一种基于动态加权融合模型的快速路交通量预测模型。选用重庆市某快速路断面的交通量数据，对构建模型的准确性进行评价，结果表明该模型具有较好的预测精度，其预测值可以为城市交通规划及管理提供有效的数据支持。

本文构建的融合模型更适合历史数据量较少情况下的交通量短期预测，由于时间序列模型的局限性，对于时间序列数据突变的预测仍有欠缺，下一阶段可以将模型与机器学习、深

度学习等方法相结合，以提高融合预测模型的综合性能。

参考文献

- [1] 温惠英,元昱青,赵胜.基于改进 GRU 模型的高速公路短时交通量预测[J].广西大学学报(自然科学版),2023,48(02):459-468.
- [2] ZHOU T, JIANG D Z, LIN Z Z, et al. Hybrid dual Kalman filtering model for short-term traffic flow forecasting[J]. IET Intelligent Transport Systems, 2019,13(6):1023-1032.
- [3] Tiantian L, Yanqin W, Le Z. A Traffic Interval Prediction Method Based on ARIMA[J]. Journal of Physics: Conference Series,2021,1880(1):
- [4] 李巧茹,赵蓉,陈亮.基于 SVM 与自适应时空数据融合的短时交通流量预测模型[J].北京工业大学学报,2015,41(04):597-602.
- [5] 张海亮.基于改进 GA—BP 的模糊神经网络在高速公路交通量预测中的研究[J].公路,2021,66(09):292-295.
- [6] 刘永乐,谷远利.基于 CNN-BiLSTM 的高速公路交通流量时空特性预测[J].交通科技与经济,2022,24(01):9-18.
- [7] 韦凌翔,陈红,王永岗等.基于 RVM 和 ARIMA 的短时交通流量预测方法研究[J].武汉理工大学学报(交通科学与工程版),2017,41(02):349-354.
- [8] 陈昆,曲大义,贾彦峰等.基于 ARIMA-SVM 模型的短时交通流量预测研究[J].青岛理工大学学报,2022,43(05):104-109.
- [9] 黄欣,谢文红,陈耀鹏,等.基于 AGRU-Trans 融合模型的特长海底隧道交通量预测[J].青岛理工大学学报,2024,45(01):118-125.
- [10] 李松江,弓晋霞,丁岩,等.基于数据融合的高速公路短时交通量预测模型[J].计算机工程与设计,2018,39(10):3208-3213.

作者简介

许可心，女，硕士研究生，重庆交通大学。电子邮箱：1309035878@qq.com

林婧，女，硕士研究生，重庆交通大学。电子邮箱：273494631@qq.com

杨泽鹏，男，硕士研究生，重庆交通大学。电子邮箱：1362059260@qq.com