

引用格式:孙秀娟,陆新秀,刘法胜,等. 基于熵权法的交通流组合预测模型研究[J]. 山东科技大学学报(自然科学版), 2018, 37(4):111-117.

SUN Xiujuan, LU Xinxiu, LIU Fasheng, et al. Research on combination prediction model of traffic flow based on entropy weight method[J]. Journal of Shandong University of Science and Technology (Natural Science), 2018, 37(4): 111-117.

# 基于熵权法的交通流组合预测模型研究

孙秀娟<sup>1</sup>, 陆新秀<sup>1</sup>, 刘法胜<sup>2</sup>, 任传祥<sup>2</sup>

(1. 山东科技大学 电气与自动化工程学院, 山东 青岛 266590; 2. 山东科技大学 交通学院, 山东 青岛 266590)

**摘要:**针对基于智能算法的组合预测模型算法复杂、计算周期较长等问题,将信息熵理论应用于交通流量预测,利用客观数据所包含的隐藏信息,建立基于熵权法的变权重组合预测模型。以具体某路段的交通流量为基础数据,分别用基于熵权法的组合预测模型和其他几种预测模型进行分析比较。仿真实验结果表明,该组合预测模型的预测精度和运算速度优于其他几种模型。

**关键词:**智能交通;熵权法;MATLAB仿真;交通流预测;遗传算法

中图分类号:U491.1<sup>+</sup>

文献标志码:A

文章编号:1672-3767(2018)04-0111-07

DOI:10.16452/j.cnki.sdkjzk.2018.04.014

## Research on Combination Prediction Model of Traffic Flow Based on Entropy Weight Method

SUN Xiujuan<sup>1</sup>, LU Xinxiu<sup>1</sup>, LIU Fasheng<sup>2</sup>, REN Chuanxiang<sup>2</sup>

(1. College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao, Shandong 266590, China;

2. College of Transportation, Shandong University of Science and Technology, Qingdao, Shandong 266590, China)

**Abstract:**In view of the complexity and long period of calculation in most combination prediction model based on intelligent algorithm, this paper applied information entropy theory to traffic flow prediction. A new combination prediction model based on entropy weight method was established by using the implicit information generated by objective data. With the actual traffic volume data measured on a certain street as the essential data, simulation experiments were implemented by using the new model and other different methods for comparison. The results indicate that the proposed new model is superior to other different methods in both prediction accuracy and calculating speed.

**Key words:** intelligent transportation; entropy-weight-method; MATLAB simulation; traffic flow forecast; genetic algorithm

在智能交通研究领域中,无论是进行交通规划,还是实施交通诱导与控制,都需要提供交通信息并对交通状态进行动态估计。因此,对交通流量的预测一直是智能交通理论研究和工程应用的热点。

目前,在交通流量预测的理论和方法已经有许多研究成果,主要有基于统计理论的模型<sup>[1]</sup>、基于神经网络理论的模型<sup>[2]</sup>、基于动态分配理论的模型<sup>[3]</sup>、基于非线性预测理论的模型和基于交通模拟仿真的模型等<sup>[4]</sup>。常见的具体方法有:ARIMA法、历史趋势法、卡尔曼滤波模型、神经网络模型等。相关研究表明,单

收稿日期:2016-11-06

作者简介:孙秀娟(1974—),女,山东新泰人,副教授,博士,主要从事交通信息处理、电池管理、嵌入式系统应用等方面研究。

E-mail:13863967594@163.com

一模型或方法在预测精确性、快速性、动态反馈性、鲁棒性或实用性等方面各有优缺点<sup>[5-7]</sup>。

1969年, Bates等<sup>[8]</sup>首次提出组合预测方法, 指出它能发挥各个模型的优点, 相互弥补缺陷, 有效提高预测精度。此后, 组合预测法一直是国内外预测界研究的热点课题, 在智能交通预测领域也得到了广泛应用。

目前较多的组合模型预测方法主要思想是应用各种智能算法, 模拟人的思维机制来进行预测。这种组合方法虽然能提高预测精度, 但往往以增加运算时间和提高算法复杂度为代价<sup>[9]</sup>。在实时性要求高或者缺少足量样本的情况下并不适用。而实际的交通系统, 因信息采集物理设备及覆盖面的限制、外在环境的复杂性及不确定性等因素的影响, 容易导致测量信息的不完整、实时性不高或信息量不足。因此, 当只有小样本或信息匮乏时, 要得到实时的交通流量预测值, 需寻找一种算法简单又满足精度要求的新方法。

信息熵因为完全建立在原始数据基础上, 可以充分利用样本数据的显性信息和隐性信息, 解决小样本或信息匮乏的问题。因此, 提出基于信息熵原理, 将熵权法(entropy weight method, EWM)用于短时交通流的组合预测并做了数据验证。

### 1 常用交通流量组合预测法介绍

组合预测模型的数学描述通常为:

设在  $t$  时刻的交通参数实测值是  $y(t)$ ,  $t = 1, 2, \dots, n$ ,  $m$  种模型在  $t$  时刻对交通参数的预测值为:  $\hat{y}_i(t)$ ,  $t = 1, 2, \dots, n$ ,  $i = 1, 2, \dots, m$ ,  $\omega_i$  为第  $i$  种预测模型在组合预测模型中的权重, 则组合预测值为:

$$\begin{cases} \hat{y}(t) = \sum_{i=1}^m \omega_i \hat{y}_i(t), & t = 1, 2, \dots, n \\ s. t. & \sum_{i=1}^m \omega_i = 1, \omega_i \geq 0 (i = 1, 2, \dots, m) \end{cases} \quad (1)$$

显然, 确定各预测模型的权重系数  $\omega_i$  是组合预测的关键, 常见的确定权重方法有以下几种。

#### 1.1 等权重组合预测法

这是一种直接给各独立预测模型分配相同权重的方法, 若采用  $m$  种独立模型做组合预测, 则每种模型的权重系数为  $\omega_i = \frac{1}{m}$ 。此法虽然计算简单快速, 但不能体现高精度的预测模型在组合预测中的比重, 未考虑各种模型的差异和优劣, 实际中很少应用。

#### 1.2 用误差最优确定权重的组合预测法

该方法一般以组合预测的误差最小为原则, 如取绝对值误差最小或方差最小等<sup>[10]</sup>, 所取误差不同则解出的权重也不同。

以最小方差组合法为例, 设组合预测误差为  $e_t$ ,  $e_t = \hat{y}(t) - y(t)$ ,  $t = 1, 2, \dots, n$ , 则确定组合预测模型的权重转化为求解下述优化问题:

$$\begin{cases} \min J = \sum_{t=1}^n e_t^2 = \sum_{t=1}^n (y(t) - \hat{y}(t))^2 \\ s. t. & \sum_{i=1}^m \omega_i = 1 \quad \omega_i \geq 0 (i = 1, \dots, m) \end{cases} \quad (2)$$

用该方法确定出的权重, 较之等权重算法, 可信度有所提高。但是单一的误差指标并不能全面评价各独立预测方法的优劣, 权重系数的算法有待完善。

#### 1.3 基于各种智能算法的组合预测法

由于智能算法是目前的研究热点, 因此基于智能算法的各种组合预测法有很多。文献[11]提出了基于贝叶斯与神经网络的组合预测方法, 文献[12]提出了基于粒子群的模糊神经网络交通流组合预测方法, 文献[13-14]提出了基于遗传算法与神经网络的组合预测方法, 文献[15]研究了基于混沌的短时交通流预测模型, 文献[16-18]利用支持向量机算法分别与小波变换、卡尔曼滤波和组合核函数相结合的方法建立了组合预测模型。这些研究主要是将两种或两种以上智能算法进行组合优化, 从而建立组合预测模型。

以基于遗传算法的组合预测法为例<sup>[19]</sup>,在确定权重系数时,以某种误差最小为目标,需要进行初始化、编码表示,遗传算法循环执行计算适应值、选择复制和应用杂交和变异算子的步骤,直到算法找到一个能接受的解,或迭代了预置的次数。

该种组合预测法能充分发挥各独立预测方法的优势,预测精度高,但有两大大劣势:一是算法复杂,计算周期长,不能满足实时预测的需要;二是需要大量的训练样本,在交通流量信息不完全时不适用。由于信息熵理论和模型可以充分利用样本数据的显性、隐性信息,解决小样本或信息匮乏的问题,很好地弥补组合预测方法的缺点,因此,我们采用基于熵权法的组合预测模型。

## 2 基于熵权法的组合预测模型

### 2.1 熵权法基本原理

Shannon 在 1948 年把熵概念引入到信息理论,用信息熵来对系统的不确定性进行量度。定义如下:

设系统可能处于几种不同的状态,分别为  $x_1, x_2, \dots, x_n$ , 即  $X = \{x_i\} | i = 1, 2, \dots, n$ , 且这些状态是互不相容的, 命每种状态出现的概率为分别为  $p_1, p_2, \dots, p_n$ , 且  $\sum_{i=1}^n p_i = 1$ , 则该系统信息熵的公式为:  $H(X) = - \sum_{i=1}^n p_i \ln p_i$ 。

熵权法的基本思想是依据信息熵的概念和性质,把各个独立预测模型的多种误差所包含的信息进行量化,利用此信息来确定各独立模型在组合预测中所占的权重,进而建立基于熵权法的组合预测模型<sup>[20]</sup>。原理如下:

设有  $m$  种独立子模型预测方法,  $n$  个误差指标, 则有指标矩阵:

$$\mathbf{R}' = \begin{bmatrix} r'_{11} & r'_{12} & \cdots & r'_{1n} \\ r'_{21} & r'_{22} & \cdots & r'_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ r'_{m1} & r'_{m2} & \cdots & r'_{mn} \end{bmatrix} \quad (3)$$

对  $\mathbf{R}'$  做标准化, 记  $r_{ij}$  为第  $i$  种预测法在第  $j$  个误差指标上的值, 将  $\mathbf{R}'$  正则化为:  $\mathbf{R} = (r_{ij})_{m \times n}$ , 其中  $r_{ij} = \frac{r'_{ij}}{\sum_{i=1}^m r'_{ij}} \in [0, 1]$ , 第  $j$  个指标的信息熵为:

$$E_j = -K \sum_{i=1}^m r_{ij} \ln r_{ij} \quad (j = 1, 2, \dots, n) \quad (4)$$

式中:  $K = \frac{1}{\ln n}$ , 且规定当  $r_{ij} = 0$  时,  $r_{ij} \ln r_{ij} = 0$ 。

定义第  $j$  个指标的权重为:

$$\theta_j = \frac{1 - E_j}{\sum_{j=1}^n (1 - E_j)} \quad (5)$$

因为信息熵以原始数据为基础,客观性比较强,所以采用信息熵来确定组合预测模型中各个子模型的权重,不易受主观因素的影响。

### 2.2 算法步骤

由于各独立预测方法的误差本身就隐含着该方法的诸多信息,因此,借用各个独立预测方法的误差指标信息熵,确定各个指标的权重,从而最终确定组合预测模型中各独立预测方法的权重信息。在熵权法原理基础上,设计了针对交通流参数的组合预测模型,算法流程如图 1 所示。

其中, 指标矩阵按公式(3)的计算方法, 仿真实验中选取绝对百分比误差作为误差指标。与理想方案的贴近度计算公式如下:

$$d_{ij} = D_{ij} / \sum_{i=1}^m \sum_{j=1}^n D_{ij} \quad (6)$$

式中,  $D_{ij}$  反映的是评价指标。常用评价体系中的指标类型有成本型、效益型和固定型三种。本研究采用的是成本型指标, 也称为反向指标, 即指评价价值

越小越好的指标, 因此取  $D_{ij} = \frac{r_j^*}{r_{ij}} = \frac{\min_i |r_{ij}|}{r_{ij}}$ 。

熵值  $E$  表示第  $j$  个误差指标对各预测方法的相对重要性的不确定度, 根据 Jaynes 最大熵原理<sup>[21]</sup>, 确定的指标综合权系数应使 Shannon 信息熵取极大值。对于离散随机变量, 当其可能的取值等概率分布时, 其熵达到最大值。因此,

$$E = - \sum_{i=1}^m \frac{d_{ij}}{d_j} \ln \frac{d_{ij}}{d_j} \quad (7)$$

式中,  $d_j = \sum_{i=1}^m d_{ij}$ 。对  $E$  进行归一化处理, 得到表征误差指标  $r_j$  的评价决策重要性的熵值公式:

$$e_j = - \frac{1}{\ln m} \sum_{i=1}^m \frac{d_{ij}}{d_j} \ln \frac{d_{ij}}{d_j} \quad (8)$$

则第  $j$  个指标的客观权重:

$$\theta_j = \frac{1 - e_j}{\sum_{j=1}^n (1 - e_j)} \quad (9)$$

各个独立预测方法的权重计算:

$$\omega_i = \sum_{j=1}^n \theta_j D_{ij} / \sum_{i=1}^m \sum_{j=1}^n \theta_j D_{ij} \quad (10)$$

最后, 根据  $\omega_i$  及式(1), 计算得到最终的组合预测值。

### 3 实证仿真实验

算法采用 MATLAB 编程实现。为便于比较, 采用文献[19]中的样本数据进行仿真, 即烟台市 1990~1999 年市内公路交通综合流量。文中给出了实际的交通流量值, 以及三种独立预测法的预测值和误差, 算法分别是遗传参数算法、最小二乘法和神经网络法, 分别记为  $f_1$ 、 $f_2$ 、 $f_3$ , 数据值见表 1 所示。

在  $f_1$ 、 $f_2$ 、 $f_3$  三种独立预测结果的基础上, 用本研究提出的基于熵权法的组合预测模型, 计算得到各模型所占的权重、组合预测值及误差, 如表 2 所示。其中  $\omega_1$ 、 $\omega_2$ 、 $\omega_3$  分别表示  $f_1$ 、 $f_2$ 、 $f_3$  在组合预测模型中所占的权重。

将表 2 中计算的误差按年份分别与表 1 中的独立预测法的值对比可知, 基于熵权法的组合预测模型得到的预测值更接近实际值, 误差较小, 说明该方法切实可行。

另外, 为了与其他种类的组合预测模型比较, 将基于熵权法的组合预测模型记为  $f_5$ , 文献[19]中基于遗传算法的组合预测模型记为  $f_4$ , 用两种不同的组合模型分别计算预测值和误差, 结果如表 3 所示。

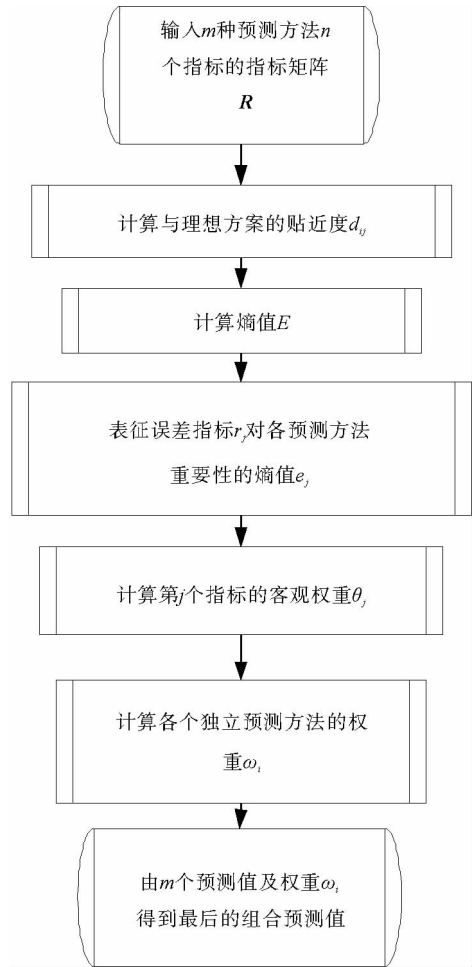


图 1 基于熵权法的组合预测模型算法流程  
Fig. 1 Algorithm steps of combination prediction model based on EWM

表 1 单一算法的预测值和误差

Tab. 1 Predicted value and errors of individual prediction algorithms

年份	实际值	$f_1$		$f_2$		$f_3$	
		预测值	误差/%	预测值	误差/%	预测值	误差/%
1990	3 217.1	3 534.4	9.9	2 871.4	10.7	3 386.1	5.1
1991	3 519.8	3 718.9	5.7	3 408.2	3.2	35 29.5	2.7
1992	3 727.9	3 913.1	5.0	3 690.1	1.0	3 683.8	1.2
1993	3 785.0	4 117.4	8.8	3 523.5	6.9	3 844.9	1.6
1994	3 724.7	4 332.4	16.3	3 539.5	5.0	4 013.0	7.7
1995	3 612.6	4 558.7	26.2	3 395.8	6.0	4 188.5	16.0
1996	3 861.1	4 796.7	24.2	3 326.5	13.8	4 371.7	13.2
1997	4 088.6	5 047.2	23.4	3 057.7	25.2	4 562.9	11.6
1998	4 728.5	5 310.7	12.3	3 821.3	19.2	4 762.4	0.7
1999	5 392.3	5 588.1	3.6	5 376.7	0.3	4 970.6	7.8
平均误差		13.54		9.13		6.76	

表 2 基于熵权法的组合预测结果

Tab. 2 Result by combination prediction model based on EWM

年份	权重	组合预测值	误差/%
1990	$w_1 = 0.258 6; w_2 = 0.239 3; w_3 = 0.502 1$	3 301.3	2.62
1991	$w_1 = 0.204 4; w_2 = 0.364 1; w_3 = 0.431 5$	3 524.0	0.12
1992	$w_1 = 0.098 4; w_2 = 0.419 8; w_3 = 0.409 8$	3 709.5	0.49
1993	$w_1 = 0.128 6; w_2 = 0.164 0; w_3 = 0.707 4$	3 827.2	1.11
1994	$w_1 = 0.156 8; w_2 = 0.511 2; w_3 = 0.332 0$	3 821.0	2.59
1995	$w_1 = 0.142 8; w_2 = 0.623 4; w_3 = 0.233 8$	3 747.2	3.73
1996	$w_1 = 0.218 0; w_2 = 0.382 3; w_3 = 0.399 7$	4 064.8	5.28
1997	$w_1 = 0.253 4; w_2 = 0.235 3; w_3 = 0.511 2$	4 331.0	5.93
1998	$w_1 = 0.052 1; w_2 = 0.033 3; w_3 = 0.914 6$	4 759.6	0.66
1999	$w_1 = 0.074 3; w_2 = 0.891 4; w_3 = 0.034 3$	5 378.5	0.26

表 3 两种组合预测算法结果对比

Tab. 3 Effect comparison of 2 combined prediction models

年份	$f_4$		$f_5$	
	组合预测值	误差/%	组合预测值	误差/%
1990	3 294.8	2.4	3 301.3	2.6
1991	3 557.1	1.0	3 524.0	0.1
1992	3 758.4	0.8	3 709.5	0.5
1993	3 845.1	1.6	3 827.2	1.1
1994	3 987.2	7.0	3 821.0	2.6
1995	4 093.0	13.2	3 747.2	3.7
1996	4 225.7	9.4	4 064.8	5.3
1997	4 312.0	5.5	4 331.0	5.9
1998	4 683.6	0.9	4 759.6	0.7
1999	5 276.2	2.1	5 378.5	0.3
平均误差		5.7		2.3
平均用时		0.65 s		0.01 s

由三个表格数据对比可知,新模型  $f_5$  的预测值平均误差分别比单一模型  $f_1$ 、 $f_2$ 、 $f_3$  预测误差低 11.24%、6.83%、4.46%;与基于遗传算法的组合预测相比,虽然预测精度只提高了 3.4%,但在算法的运行时间上具有明显优势,说明新的模型较基于智能算法的组合模型计算更快。特别是在大量数据的预测上,能极大节省运算时间。

将 5 种算法的预测值与实际的综合交通流量相比较,作交通流量曲线如图 2 所示。

图 2 直观地给出了五种预测算法的效果。用实线表示的为交通流量的实际值,从预测值与实际值曲线的贴附度来看,算法  $f_5$ ,即基于熵权法的组合预测模型效果最优,与实际值的误差最小。

通过理论分析及实际仿真结果的验证表明,基于熵权法的组合预测模型能发挥各独立模型的优势,在较短时间内给出较好的预测结果。

### 4 结论

本研究提出的基于熵权法的交通流组合预测模型,充分利用各个独立预测方法所包含的显性信息和隐性信息,相比其他几种组合模型,尤其是复杂的智能算法组合预测,基于熵权法的交通流组合预测模型客观性优点突出,且计算简单、实时性和实用性强,实证分析表明了该模型的有效性。

#### 参考文献:

[1]CHIEN S I,KUCHIPUDI C M. Dynamic travel time prediction with real-time and historic data [J]. Journal of Transportation Engineering,2003,129(6):608-616.

[2]DOUGHERTY M. A review of neural networks applied to transport [J]. Transportation Research Part C,1995,3(4):247-260.

[3]OKUTANI I,STEPHANEDES Y J. Dynamic prediction of traffic volume through Kalman filtering theory [J]. Transportation Research Part B,1984,18(1):1-11.

[4]博贵. 城市智能交通动态预测模型的研究及应用[D]. 广州:华南理工大学,2014.

[5]杨春霞,符义琴,鲍铁男. 基于相似性的短时交通流预测[J]. 公路交通科技,2015,32(10):124-128.  
YANG Chunxia,FU Yiqin,BAO Tienan. Short-term traffic flow prediction based on similarity[J]. Journal of Highway and Transportation Research and Development,2015,32(10):124-128.

[6]SMITH B L,DEMETSKY M J. Traffic flow forecasting:Comparison of modeling approaches[J]. Journal of Transportation Engineering,1997,123(4),261-266.

[7]SMITH B L,WILLIAMS B M,OSWALD R K. Comparison of parametric and nonparametric models for traffic flow forecasting [J]. Transportation Research Part C,2002,10(4):303-321.

[8]BATES J M,GRANGER C W J. The combination of forecasts [J]. Operational Research,1969,20(4):451-468.

[9]康军,段宗涛,唐蕾,等. 高斯过程回归短时交通流预测方法[J]. 交通运输系统工程与信息,2015,15(4):51-56.  
KANG Jun,DUAN Zongtao,TANG Lei,et al. A short term traffic flow prediction method based on Gaussian processes regression[J]. Journal of Transportation Systems Engineering and Information Technology (Nature Science),2015,15(4):51-56.

[10]高宏岩. 融合移动信号流的高速公路交通拥挤预警与调控[D]. 青岛:山东科技大学,2011.

[11]ZHENG W Z,LEE D H. Short-term freeway traffic flow prediction;Bayesian combined neural network approach[J]. Journal of Transportation Engineering,2006,132(2):114-121.

[12]陈俊洵,程龙生. 基于粒子群的模糊神经网络的短时交通流量组合预测[J]. 数学的实践与认识,2014,44(22):177-184.

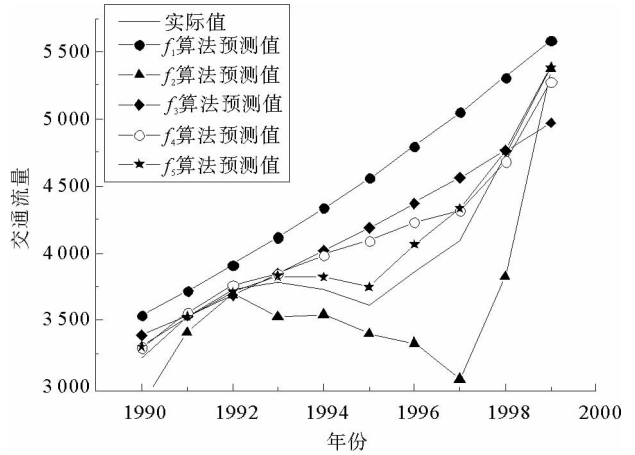


图 2 5 种不同算法的预测值与实际值曲线

Fig. 2 Curve comparison of actual value with predicted value by 5 different algorithms

- CHEN Junxun, CHENG Longsheng. Combined forecasting model of short-term traffic flow based on modified particle swarm optimizer and fuzzy neural network[J]. Mathematics in Practice and Theory, 2014, 44(22): 177-184.
- [13] 卢建中, 程浩. 改进 GA 优化 BP 神经网络的短时交通流预测[J]. 合肥工业大学学报(自然科学版), 2015, 38(1): 127-131.  
LU Jianzhong, CHENG Hao. Short-term traffic flow forecast based on modified GA optimized BP neural network[J]. Journal of Hefei University of Technology(Nature Science), 2015, 38(1): 127-131.
- [14] 张琛, 徐国丽. 基于云遗传的 RBF 神经网络的交通流量预测[J]. 计算机工程与应用, 2014, 50(16): 216-220.  
ZHANG Chen, XU Guoli. Prediction for traffic flow of RBF neural network based on cloud genetic algorithm[J]. Computer Engineering and Applications, 2014, 50(16): 216-220.
- [15] 刘力军, 李松, 解永乐. 短时交通流混沌预测模型比较研究[J]. 数学的实践与认识, 2011, 41(17): 106-114.  
LIU Lijun, LI Song, XIE Yongle. Comparative study on chaotic prediction model in short-term traffic flow[J]. Mathematics in Practice and Theory, 2011, 41(17): 106-114.
- [16] 陈旭梅, 龚辉波, 王景楠. 基于 SVM 和 Kalman 滤波的 BRT 行程时间预测模型研究[J]. 交通运输系统工程与信息, 2012, 12(4): 29-34.  
CHEN Xumei, GONG Huibo, WANG Jingnan. BRT vehicle travel time prediction based on SVM and Kalman filter[J]. Journal of Transportation Systems Engineering and Information Technology, 2012, 12(4): 29-34.
- [17] 徐娜, 钱超. 基于小波-SVR 模型的短时交通量预测研究[J]. 公路交通技术, 2015(4): 141-145.  
XU Na, QIAN Chao. Research on forecast of short-term traffic volume based on wavelet-SVR model [J]. Technology of Highway and Transport, 2015(4): 141-145.
- [18] BING Q, GONG B, YANG Z, et al. Short-term traffic flow prediction based on combined kernel function relevance vector machine model [J]. Mathematical Problems in Engineering, 2015(9): 1-9.
- [19] 宋曰聪, 胡伟, 张涛. 基于遗传算法的交通流量组合预测研究[J]. 微计算机信息, 2007, 29: 55-56.  
SONG Yuecong, HU Wei, ZHANG Tao. Combined prediction research of city traffic flow based on genetic algorithm[J]. Microcomputer Information, 2007, 29: 55-56.
- [20] 任敏, 胡彧. 熵权法和云模型下的物联网鲁棒性评估方法[J]. 自动化仪表, 2015, 36(5): 60-63.  
REN Min, HU Yu. Evaluation method based on entropy weight and cloud model for robustness of IOT[J]. Process Automation Instrumentation, 2015, 36(5): 60-63.
- [21] JAYNES E T. Information theory and statistical mechanics [J]. Physical Review, 1957, 106(4): 620-630.

(责任编辑: 李 磊)