

基于混沌自适应遗传 ν -SVR 的城市客运量预测

康海贵*, 李明伟, 周鹏飞, 赵泽辉

(大连理工大学 建设工程学部, 辽宁 大连 116024)

摘要: 针对城市客运量预测问题本身所存在的小样本、高维数和非线性等特点, 将 ν 支持向量回归机 (ν -support vector regression, ν -SVR) 应用于城市客运量预测. 为了提高 ν -SVR 模型的预测精度和泛化性能, 利用基于混沌理论和自适应机制的混沌自适应遗传算法 (chaos adaptive genetic algorithm, CAGA) 优选 ν -SVR 模型参数, 建立了基于 CAGA 进行参数优选的 CAGA- ν -SVR 城市客运量预测模型. 结合 1978~2008 年统计数据进行了仿真预测, 结果表明该模型的预测性能优于 RBF 神经网络模型、GA-SVR 模型和 GA- ν -SVR 模型, 平均绝对相对误差控制在 2.3% 以内, 可有效应用于城市客运量预测.

关键词: ν 支持向量回归机; 遗传算法; 混沌映射; 自适应机制; 客运量预测

中图分类号: U116.5 **文献标志码:** A

0 引言

支持向量回归机 (以下简称 SVR) 以结构风险最小化取代经验风险最小化^[1], 在模型的复杂性和学习能力之间寻求最佳折中^[2], 将实际优化问题转化为求解凸二次规划 (QP) 问题, 能很好地解决小样本、非线性和高维数等问题^[3,4]. 文献 [5] 建立了一种基于支持向量机的空中交通流量组合预测模型, 并对空中交通流量进行了成功预测. 然而传统支持向量回归机的不敏感损失函数 ϵ 的取值区间并不明确, 搜索 SVR 模型的最优参数组合难度较大, ν 支持向量回归机^[6] (ν -SVR) 的提出克服了上述不足, 但 ν -SVR 本身没有给出参数的选择方法, 而现有的搜索方法^[7] 当输入空间维数较大时, ν -SVR 模型的训练时间较长, 限制了 ν -SVR 模型的推广应用. 为了提高对 ν -SVR 模型参数的优选效率, 本文利用混沌自适应遗传算法 (CAGA) 对 ν -SVR 模型的参数进行优选, 提出基于混沌自适应遗传算法和 ν -SVR 进行城市客运量预测的新方法, 并结合统计数据进行了仿真预测和对比实验.

1 ν -SVR 算法

设已知训练集 $T = \{(x_1, y_1), \dots, (x_l, y_l)\}$, 其中 $x_i \in \mathbf{R}^d$ 为训练影响因子, $y_i \in \mathbf{R}$ 为训练目标因子, $i = 1, \dots, l$, 此时拟合函数为

$$f(\mathbf{x}) = \boldsymbol{\omega} \cdot \mathbf{x} + b \quad (1)$$

式中: b 为拟合函数的偏置量; $\boldsymbol{\omega}$ 为拟合函数的权矢量. 所求的拟合函数 $f(\mathbf{x})$ 使下面的目标函数 (结构风险) 最小:

$$F = \min \left[\frac{1}{2} \|\boldsymbol{\omega}\|^2 + C \cdot R_{\text{emp}} \right] \quad (2)$$

式中: $\frac{1}{2} \|\boldsymbol{\omega}\|^2$ 为决策函数复杂性的表达能力项;

C 为调节因子; $R_{\text{emp}} = \frac{1}{l} \sum_{i=1}^l |y_i - f(\mathbf{x}_i)|$, 为训练误差, 可用 ϵ -不敏感损失函数来度量.

$$|y - f(\mathbf{x})| = \begin{cases} 0; & |y - f(\mathbf{x})| < \epsilon \\ |y - f(\mathbf{x})| - \epsilon; & \text{其他} \end{cases} \quad (3)$$

为了克服 ϵ 选取困难的不足, 引进一个新的参数 ν 来自动选择 ϵ 的大小, 此时, 优化问题如下^[7]:

收稿日期: 2010-04-11; 修回日期: 2011-11-02.

基金项目: 高等学校博士学科点专项科研基金资助项目 (200801411105); 河南省交通厅科技计划资助项目 (200912).

作者简介: 康海贵* (1945-), 男, 教授, 博士生导师, E-mail: hgkang@dlut.edu.cn.

$$\begin{aligned} \min_{\omega, b, \zeta^*, \epsilon} \quad & \tau(\omega, \zeta^*, \epsilon) = \frac{1}{2} \|\omega\|^2 + C \cdot (\nu \cdot \epsilon + \\ & \frac{1}{l} \sum_{i=1}^l (\zeta_i + \zeta_i^*)) \\ \text{s. t.} \quad & (\omega \cdot x_i + b) - y_i \leq \epsilon + \zeta_i \\ & y_i - (\omega \cdot x_i + b) \leq \epsilon + \zeta_i^* \\ & \zeta_i^* \geq 0, \epsilon \geq 0 \end{aligned} \quad (4)$$

式中: ω 为 d 维列向量; $\zeta^{(*)} = \{\zeta_1, \zeta_1^*, \dots, \zeta_l, \zeta_l^*\}$, 为松弛变量; $\nu (\nu \in [0, 1])$ 为给定的已知参数; 而 ϵ 作为优化问题的变量出现, 其值作为解的一部分给出.

$$\begin{aligned} & \text{引入 Lagrange 函数} \\ & L(\omega, b, a^{(*)}, \beta, \zeta^{(*)}, \epsilon, \eta^{(*)}) = \\ & \frac{1}{2} \|\omega\|^2 + C \cdot \nu \cdot \epsilon + \frac{C}{l} \sum_{i=1}^l (\zeta_i + \zeta_i^*) - \beta \cdot \epsilon - \\ & \sum_{i=1}^l (\eta_i \zeta_i + \eta_i^* \zeta_i^*) - \sum_{i=1}^l a_i (\zeta_i + y_i - (\omega \cdot x_i) - \\ & b + \epsilon) - \sum_{i=1}^l a_i^* (\zeta_i^* + (\omega \cdot x_i) + b - y_i + \epsilon) \end{aligned} \quad (5)$$

其中乘子 $a^{(*)}, \beta, \eta^{(*)} \geq 0$. 为求它的极小, 分别对 $\omega, b, \zeta^{(*)}, \epsilon$ 计算偏导数, 并同时引入核函数 $K(x_i, y_i)$, 求解上述优化问题(4)的对偶问题为

$$\begin{aligned} \min_{a, a^*} \quad & \omega(a, a^*) = \frac{1}{2} \sum_{i,j=1}^l [(a_i^* - a_i)(a_j^* - a_j) \times \\ & K(x_i, y_j)] - \sum_{i=1}^l (a_i^* - a_i) y_i \end{aligned} \quad (6)$$

$$\begin{aligned} \text{s. t.} \quad & \sum_{i=1}^l (a_i^* - a_i) = 0 \\ & 0 \leq a_i^{(*)} \leq \frac{C}{l}; i = 1, \dots, l \\ & \sum_{i=1}^l (a_i^* + a_i) \leq C \cdot \nu \end{aligned}$$

由约束条件 $\sum_{i=1}^l (a_i^* + a_i) \leq C \cdot \nu$ 知 $\frac{1}{C} \sum_{i=1}^l (a_i^* + a_i) \leq \nu$, 由约束条件 $0 \leq a_i^{(*)} \leq \frac{C}{l}$ 知 $0 \leq \frac{a_i^{(*)}}{C} \leq \frac{1}{l}$, 而错误样本个体对应的 $a_i^{(*)}$ ($i = 1, 2, \dots, q$) 均满足 $a_i = 0, a_i^* = \frac{C}{l}$ 或 $a_i^* = 0, a_i = \frac{C}{l}$, 故有下列不等式成立: $\frac{q}{l} = \frac{1}{C} \cdot \frac{C}{l} \cdot q = \frac{1}{C} \sum_{i=1}^q (a_i + a_i^*) \leq \frac{1}{C} \sum_{i=1}^l (a_i + a_i^*) \leq \nu$, 所以 ν 是

错误样本个体数所占总样本数份额的上界.

由上述 Lagrange 函数 L 对 ϵ 计算偏导数, 由 $\nabla_{\epsilon} L = 0$ 得

$$C \cdot \nu - \sum_{i=1}^l (a_i + a_i^*) - \beta = 0 \quad (7)$$

由式(4)的 KKT(Karush-Kuhn-Tucker) 条件知, 当 $\epsilon^* > 0$ 时, $\beta = 0$, 若记支持向量的个数为 p , 由于支持向量对应于 $0 \leq a_i \leq \frac{C}{l}, a_i^* = 0$ 或 $0 \leq a_i^* \leq \frac{C}{l}, a_i = 0$, 则由约束条件 $\sum_{i=1}^l (a_i^* + a_i) \leq C \cdot \nu$ 和式(7)知: $\nu = \frac{l}{C} \sum_{i=1}^p (a_i + a_i^*) \leq \frac{1}{C} \cdot \frac{C}{l} \cdot p = \frac{p}{l}$, 所以 ν 是错误样本个体数所占总样本数份额的下界, 从而为 ν 值的选取提供了一个依据, 从这个意义上讲 ν -SVR 优于 ϵ -SVR.

2 混沌自适应遗传算法

遗传算法^[8](genetic algorithms, GA)应用于模型参数的优选, 具有全局寻优、鲁棒性和自适应性强等优点^[9], 然而用于小样本、高维数和非线性问题的参数寻优时, 却难于保证寻优精度. 混沌自适应遗传算法的提出在一定程度上克服了标准遗传算法的上述不足, 应用于模型参数的优选具有高效、精准等特点, 因此, 为了提高 ν -SVR 模型的预测精度和泛化性能, 本文利用混沌自适应遗传算法对 ν -SVR 模型的参数 C, ν 和 δ 进行优选.

2.1 Logistic 映射初始化父代群体

混沌是自然界中一种普遍的非线性现象, 其行为复杂且类似随机, 但存在精致的内在结构, 具有随机性、遍历性和规律性等特点. 基于混沌系统对初值非常敏感^[10]这一特点, 利用 Logistic 混沌映射^[11]初始化父代群体.

设 Logistic 映射产生的混沌变量为 $\{x(i)\}$, $i = 1, 2, \dots, n$. 则有

$$x(i+1) = u \cdot x(i) \cdot (1 - x(i)); x(i) \in [0, 1] \quad (8)$$

其中 $x(i)$ 是变量 x 在第 i 次的迭代值, u 是控制参数. 当 $u = 4$ 时系统完全处于混沌状态^[12], $x(0)$ 可设为 $(0, 1)$ 区间内除 $0.25, 0.50, 0.75$ 以外的任何初值, 然后根据式(8)得到混沌变量, 此时混沌变量可以遍历到整个搜索区域^[13].

2.2 混沌自适应遗传操作

为了在保证遗传算法搜索精度的同时, 加快

收敛速度,本文采用下式分别对交叉概率 p_c 和变异概率 p_m 进行计算:

$$p_c = \begin{cases} k_{c-\max}; & f'_i \leq f_{\text{avg}} \\ k_{c-\max} - (k_{c-\max} - k_{c-\min}) \frac{f'_i - f_{\text{avg}}}{f_{\text{best}} - f_{\text{avg}}}; & f'_i > f_{\text{avg}} \end{cases} \quad (9)$$

$$p_m = \begin{cases} k_{m-\max}; & f_i \leq f_{\text{avg}} \\ k_{m-\max} - (k_{m-\max} - k_{m-\min}) \frac{f_i - f_{\text{avg}}}{f_{\text{best}} - f_{\text{avg}}}; & f_i > f_{\text{avg}} \end{cases} \quad (10)$$

式中: $k_{c-\max}$ 和 $k_{c-\min}$ 分别为交叉概率的最大值和最小值; f_{avg} 为种群的平均适应度值; f_{best} 为种群最优适应度值; f'_i 为要交叉的两个个体适应度值的平均值; f_i 为要变异的个体适应度值; $k_{m-\max}$ 和 $k_{m-\min}$ 分别为变异概率的最大值和最小值。

在标准遗传算法进化过程中交叉和变异操作通过以下两式进行:

$$\begin{cases} y_2(j, i) = u_1 y(j, i_1) + (1 - u_1) y(j, i_2); \\ u_3 < p_c(i) \\ y_2(j, i) = u_2 y(j, i_1) + (1 - u_2) y(j, i_2); \\ u_3 \geq p_c(i) \end{cases} \quad (11)$$

$$\begin{cases} y_3(j, i) = u(j); & u_4 < p_m(i) \\ y_3(j, i) = y(j, i); & u_4 \geq p_m(i) \end{cases} \quad (12)$$

式中: $y_2(j, i)$ 和 $y_3(j, i)$ 表示通过交叉和变异产生的子代群体; $y(j, i)$ 表示父代群体; u_1 、 u_2 、 u_3 、 u_4 和 $u(j)$ 均为随机数,然而随机分布不能保证序列的遍历性,因此,在混沌自适应遗传算法中应用混沌映射来生成 u_1 、 u_2 、 u_3 、 u_4 和 $u(j)$,用于提高交叉和变异的效率。

2.3 引入加速遗传算子

为了提高标准遗传算法的稳健性与计算效率,引进加速遗传算子,利用进化迭代过程中产生的优秀个体所包含的变化特性的信息,来调整参数的初始区间,形成加速遗传,加快模型的收敛速度。

3 CAGA 优选 ν -SVR 城市客运量预测模型的计算步骤

本文以训练结束后的回归序列方差的倒数作为进化个体的适应度值 f 进行遗传操作。

$$f = \frac{1}{\frac{1}{n} \sum_{t=1}^n (Y^1(t) - Y(t))^2} \quad (13)$$

其中 n 为输入样本个数, $Y^1(t)$ 为回归序列, $Y(t)$ 为实际序列。具体步骤如下:

步骤 1 数据归一化。按下式对已知统计数据 and 模型参数的初始区间进行归一化处理。

$$x^*(i) = (x(i) - x_{\min}) / (x_{\max} - x_{\min}) \quad (14)$$

其中 $x^*(i)$ 为序列中归一化后的第 i 个值, $x(i)$ 为序列中第 i 个原始值, x_{\min} 为序列中的最小值, x_{\max} 为序列中的最大值。

步骤 2 混沌初始化。设种群规模为 m ,按式(8)生成父代群体 $y(j, i)$ ($j = 1, 2, 3; i = 1, 2, \dots, m$),其中 $y(1, i)$ 为第 i 个染色体上的 C 值, $y(2, i)$ 为第 i 个染色体上的 ν 值, $y(3, i)$ 为第 i 个染色体上的 δ 值。本文取 $m = 100$, $u = 4.0$, $y(j, 0) = 0.4$ 。

步骤 3 模型训练。以生成的每个染色体上的基因 C 、 ν 和 δ 作为 ν -SVR 模型参数,训练 ν -SVR 模型,输入已知数据,计算相应的回归序列,并根据回归序列按式(13)计算染色体的适应度值。

步骤 4 进化终止判断。判断种群中最优个体是否达到进化终止条件,若达到条件转入步骤 8,否则转入步骤 5,进化代数 $G = G + 1$ 。进化终止条件采用最大进化代数 G_{\max} 与相邻种群中最优个体适应度值相对变化率 E 相结合。

步骤 5 选择、混沌自适应交叉和混沌自适应变异。根据步骤 3 计算所得的染色体的适应度值,进行选择、混沌自适应交叉和混沌自适应变异操作,其中交叉概率 p_c 和变异概率 p_m 按式(9)和(10)计算,混沌自适应交叉和混沌自适应变异操作按式(11)和(12)进行。

步骤 6 演化迭代。计算得到新个体的适应度值 f ,选取最优秀的 m 个子代个体作为下一轮演化迭代的父代群体。转入步骤 3,继续进行演化迭代过程。

步骤 7 加速遗传。当演化迭代次数达到 N_{\max} ,将进化得到的子代个体,按适应度值进行从大到小排序,取排在前面的 m 个优秀个体,将其浮动区间作为新的初始化区间,转入步骤 2,继续进行进化。

步骤 8 目标值预测。基于优选得到的最优参数 C 、 ν 和 δ ,训练 ν -SVR 模型,计算目标预测值。

计算流程见图 1。

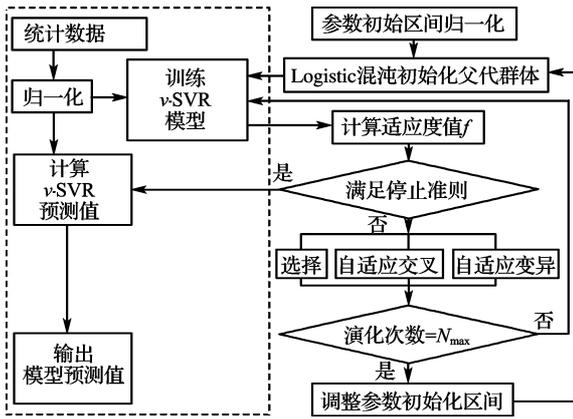


图1 CAGA-v-SVR流程图
Fig.1 Process of CAGA-v-SVR

4 CAGA-v-SVR 城市客运量预测模型仿真预测

为了验证本文提出模型的预测效果,以某市1978~2008年实际数据进行建模和对比预测.将城市客运量与总人口、能源总量、人均收入、公路里程数、消费品总额、汽车拥有量和生产总值7个可能影响因素进行相关性分析,分析结果如表1所示.

表1 城市客运量与影响因素相关性分析
Tab.1 Correlation analysis between passenger traffic volume and influence factors

影响因素	相关系数	影响因素	相关系数
总人口	0.957 8	消费品总额	0.915 8
能源总量	0.880 7	汽车拥有量	0.881 9
人均收入	0.930 4	生产总值	0.899 4
公路里程数	0.443 4		

由表1可以看出,在可能影响城市客运量Q的7个影响因素中,除公路里程数的相关系数低于0.5外,其他各影响因素的相关系数均超过了0.8,所以本文将历年的统计数据中除公路里程数以外的影响因素作为模型的影响因子集,选取1982年等6年的数据作为模型预测性能测试集,其余为训练集进行模型参数优选和目标值预测.

训练过程中C的取值范围为[1,1 000], ν 的取值范围为[0,1], δ 的取值范围为[0,20],最大进化代数 $G_{max}=100$, $k_{c-max}=0.99$, $k_{c-min}=0.40$, $k_{m-max}=0.15$, $k_{m-min}=0.01$,最大演化次数 $N_{max}=3$,邻代最优个体适应度值相对变化率 $E=0.000 1$,经过训练得到全局参数最优组合为 $C=278$, $\nu=0.43$ 和 $\delta=0.36$.

基于最优参数组合 $C=278$, $\nu=0.43$ 和 $\delta=$

0.36,利用CAGA-v-SVR预测模型对历史数据进行拟合,模型拟合均方根误差为2.78%,拟合平均绝对相对误差为1.91%.图2为CAGA-v-SVR预测模型的拟合曲线,由图可知CAGA-v-SVR预测模型对历史序列拟合效果较好.

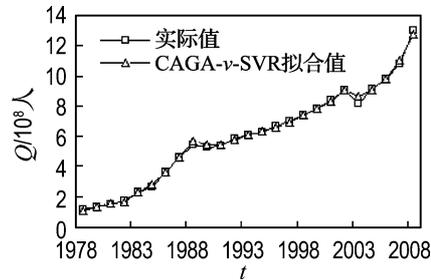


图2 CAGA-v-SVR模型训练结果
Fig.2 Training result of CAGA-v-SVR

根据得到的CAGA-v-SVR城市客运量预测模型对目标预测值进行计算.模型预测平均绝对相对误差为2.24%,仿真预测均方根误差为3.19%,将模型预测值与实际值进行对比,如图3所示,由图可知模型的预测精度达到了预期效果.

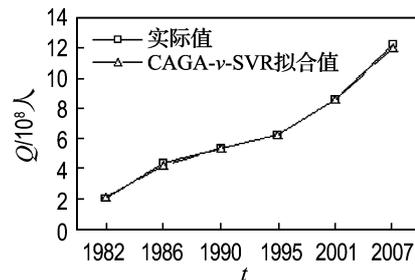


图3 CAGA-v-SVR模型仿真预测结果
Fig.3 Simulation forecasting result of CAGA-v-SVR

为了测试CAGA对v-SVR模型参数的寻优效率,本文同时利用标准遗传算法GA在同一台计算机和相同编程环境下对v-SVR模型参数进行优选,CAGA与GA寻优性能结果如表2所示.

表2 GA与CAGA寻优性能对比
Tab.2 Seeking the optimal parameters performance comparison of GA and CAGA

模型	寻优时间/s	迭代次数	适应度值	
			最大值	最小值
GA	3 000	812	0.005 7	0.002 1
CAGA	3 000	544	0.007 6	0.004 3

由表2可以看出,在一定的运行时间(3 000 s)内,由于CAGA算法采用混沌序列初始化代

群体,算法对最优解搜索的遍历性增强,从而使搜索到染色体的最大适应度值明显大于 GA 算法.加速遗传操作和自适应遗传操作的引入加速了算法对最优解的搜索,从而使 CAGA 搜索到的解更能集中于最优解的附近.然而随着混沌初始化父代群体和自适应遗传操作的引入,其每迭代一次的运行时间必然高于 GA 的运行时间,导致在规定的时间内迭代次数减少,但 CAGA 算法给出最终解的效率要明显高于 GA 算法.因此 CAGA 用于 ν -SVR 模型参数 C 、 ν 和 δ 的优选是可行的.

为了检测 CAGA- ν -SVR 模型的预测性能,本文选用 RBF 神经网络模型、标准遗传算法优选 SVR 模型(GA-SVR)、标准遗传算法优选 ν -SVR 模型(GA- ν -SVR)3 个模型分别进行建模和仿真预测,为了使对比模型具有可比性,4 个模型均在同一台微机上计算,均使用 Matlab 7.1 进行编程.考虑到寻优次数的增加也会提高优化效果,对比模型参数的选取应保证每个模型的最大寻优次数相同,对 RBF 神经网络模型优选次数的设定,应保证与 CAGA 算法具有相同的最大优选次数.同时采用以下 3 个评价指标对模型的预测效果进行分析:

(1)平均绝对相对误差:

$$e_{mar} = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y^1(t) - Y(t)}{Y(t)} \right|$$

(2)最大绝对相对误差

$$e_{mxar} = \max \left| \frac{Y^1(t) - Y(t)}{Y(t)} \right|$$

(3)均方根误差

$$e_{rms} = \sqrt{\frac{1}{n} \sum_{t=1}^n \left(\frac{Y^1(t) - Y(t)}{Y(t)} \right)^2}$$

结合统计数据,利用上述对比模型,分别对客运量进行预测,4 种模型的预测结果如表 3 所示;同时进行误差指标处理,处理结果如表 4 所示.

由表 4 可以看出,从基于不同建模结构的预测模型来看,SVR 预测模型的 e_{mar} 、 e_{mxar} 和 e_{rms} 三个指标值均小于 RBF 神经网络模型,说明 SVR 在逼近原始序列的能力方面较强,预测精度优于 RBF 神经网络模型;从是否引入参数 ν 的 SVR 模型的预测性能方面来看, ν -SVR 模型通过引入参数 ν ,明确了参数 ν 的优选区间,提高了最优解的搜索效率,所以在规定的搜索时间内 GA- ν -SVR 模型的预测精度略优于 GA-SVR 模型;对 GA- ν -SVR 模型和 CAGA- ν -SVR 模型预测精度的比较

表 3 CAGA- ν -SVR 与其他 3 个模型的仿真预测比较

Tab. 3 Comparison of simulation forecasting results between CAGA- ν -SVR and three different models

年份	真实值/ 10 ⁴ 人	模型预测值/10 ⁴ 人			
		RBF 神经网络	GA-SVR	GA- ν -SVR	CAGA- ν -SVR
1982	20 129	17 763	18 452	18 618	21 354
1986	43 590	37 651	39 241	39 621	41 708
1990	53 567	56 823	51 438	51 978	53 334
1995	61 964	66 018	59 223	59 894	61 923
2001	85 412	810 21	88 517	88 962	85 560
2007	122 557	114 592	116 424	119 403	119 682

表 4 预测误差对比

Tab. 4 Comparison of forecasting error indexes %

模型	e_{mar}	e_{mxar}	e_{rms}
RBF 神经网络	8.27	13.62	8.87
GA-SVR	5.89	9.98	6.36
GA- ν -SVR	4.75	9.10	5.39
CAGA- ν -SVR	2.24	6.09	3.19

方面,因混沌初始化父代群体和自适应遗传操作的引入,增强了最优解搜索的遍历性,从而能够搜索到更优的模型参数组合,因此 CAGA- ν -SVR 模型的预测精度优于 GA- ν -SVR 模型.总体来说,CAGA- ν -SVR 模型通过引入参数 ν ,明确了参数的取值区间,父代群体的混沌初始化和混沌自适应操作增强了对最优参数搜索的遍历性,提高了模型的预测精度,但同时也增加了模型的计算步骤,减慢了模型的搜索速度,而加速遗传操作的引入又提高了最优参数的搜索速度,最终使得 CAGA- ν -SVR 模型在保证搜索速度的前提下取得了较高的预测精度.

5 结 论

(1)本文利用混沌自适应遗传算法对 ν -SVR 模型参数 C 、 ν 和 δ 进行优选,弥补了 ν -SVR 模型参数优选难度大、效率低的不足,增强了 ν -SVR 模型的广泛应用能力和鲁棒性,提高了 ν -SVR 模型的预测精度.

(2)将建立的 CAGA- ν -SVR 城市客运量预测模型应用于城市客运量的预测,对比结果表明:本文提出模型的预测精度优于对比模型,模型预测平均绝对相对误差小于 2.3%,取得了较好的预测效果,具有一定的有效性和实用性.

参考文献:

- [1] VAPNIK V, LEVIN E, LECUN Y. Measuring the VC-dimension of a learning machine [J]. **Neural Computation**, 1994, **6**(5):851-876
- [2] VAPNIK V. **The Nature of Statistical Learning Theory** [M]. New York:Springer, 1995
- [3] GUNN S R. Support vector machines for classification and regression [R]. Southampton: Southampton University, 1998
- [4] CAO J R, CAI A N. A robust shot transition detection method based on support vector machine in compressed domain [J]. **Pattern Recognition Letters**, 2007, **28**(12):1534-1540
- [5] 耿睿, 崔德光, 徐冰. 应用支持向量机的空中交通流量组合预测模型[J]. 清华大学学报, 2008, **48**(7): 1205-1208
- [6] SCHOLKOPF B, SMOLA A J, WILLIAMSON R C, *et al.* New support vector algorithms [J]. **Neural Computation**, 2000, **12**(5):1207-1245
- [7] SCHOÖLKOPF B, BURGESS C J C, SMOLA A J. **Advances in Kernel Methods — Support Vector Learning** [M]. Cambridge:MIT Press, 1999
- [8] 孙月峰, 张胜红, 王晓玲, 等. 基于混合遗传算法的区域大系统多目标水资源优化配置模型[J]. 系统工程理论与实践, 2009, **29**(1):139-142
- [9] MIN Sung-hwan, LEE Ju-min, HAN Ingoo. Hybrid genetic algorithms and support vector machines for bankruptcy prediction [J]. **Expert Systems with Applications**, 2006, **31**(3):652-660
- [10] RETTEMEIR K, FALKENHAGEN B, KONGETER J. Risk assessment-new trends in Germany [C] // **Proceedings of the 21st International Congress on Large Dams**. Beijing: The International Commission on Large Dams, 2000:625-641
- [11] XIANG Tao, LIAO Xiao-feng, WONG Kwok-wo. An improved particle swarm optimization algorithm combined with piecewise linear chaotic map [J]. **Applied Mathematics and Computation**, 2007, **190**(2):1637-1645
- [12] LÜ Qing-zhang, SHEN Guo-li, YU Ru-qin. A chaotic approach to maintain the population diversity of genetic algorithm in network training [J]. **Computational Biology and Chemistry**, 2003, **27**(3): 363-371
- [13] LIU Bo, WANG Ling, JIN Yi-hui, *et al.* Improved particle swarm optimization combined with chaos [J]. **Chaos, Solitons and Fractals**, 2005, **25**(5):1261-1271

Prediction of passenger traffic volume using ν -support vector regression optimized by chaos adaptive genetic algorithm

KANG Hai-gui*, LI Ming-wei, ZHOU Peng-fei, ZHAO Ze-hui

(Faculty of Infrastructure Engineering, Dalian University of Technology, Dalian 116024, China)

Abstract: Aiming at the prediction of passenger traffic volume with small samples, multi-dimension and nonlinearity, ν -support vector regression (ν -SVR) is introduced to forecast passenger traffic volume. To seek the optimal forecast accuracy and generalization performance of ν -SVR, chaos adaptive genetic algorithm (CAGA) is used to optimize the parameter, which is based on chaos mapping and adaptive mechanism. Then, a new passenger traffic volume forecasting model of ν -SVR named by CAGA- ν -SVR is proposed. The model is applied to forecasting passenger traffic volume with data of 1978-2008. Compared with RBF neural network model, GA-SVR model and GA- ν -SVR model, it is concluded that CAGA- ν -SVR prediction model has higher prediction precision, and can effectively predict passenger traffic volume with less than 2.3% of mean absolute relative error.

Key words: ν -support vector regression; genetic algorithm; chaos mapping; adaptive mechanism; passenger traffic volume prediction