

基于混合人工鱼群优化 SVR 的交通流量预测

姚卫红*, 方仁孝, 张旭东

(大连理工大学 计算机科学与技术学院, 辽宁 大连 116024)

摘要: 实时准确的交通流量预测是智能交通系统(ITS)中的重要内容. 支持向量回归(SVR)能够用于解决交通流量预测问题, 针对 SVR 中存在的参数选择困难, 提出一种混合人工鱼群(AFS)算法. 该算法利用粒子群优化(PSO)算法公式改进 AFS 算法, 减小 AFS 算法中步长因子的影响, 并引入混沌初始化 AFS 机制, 选取最优 SVR 参数, 建立了基于混沌 PSO-AFS 优化 SVR 的交通流量预测模型. 仿真结果表明, 该交通流量预测模型具有更优的预测性能, 证明了其可行性和有效性.

关键词: 交通流量预测; 支持向量回归(SVR); 人工鱼群(AFS)算法; 粒子群优化(PSO); 混沌机制

中图分类号: TP391

文献标识码: A

doi:10.7511/dllgxb201506011

0 引言

准确、可靠的道路交通流量信息是智能交通系统中交通控制和信息管理的重要组成部分. 在每天的车辆出行高峰时段, 不准确的交通流量信息会增加出行者时间, 造成道路拥堵、能源消耗和环境污染, 所以高峰时段准确的交通流量预测就显得尤为重要^[1-3].

传统的交通流量预测模型有基于数学测定方法的自回归移动平均方法(ARIMA)模型、卡尔曼滤波模型和非线性回归模型^[4]. 这些模型有计算简单、速度快的优点, 但是它们不能处理交通流量信息的不确定性和非线性的复杂特征. 目前常用的交通流量预测模型有基于知识发现的智能模型, 如模糊理论、小波理论和 ANN 人工神经网络^[5]. ANN 模型由于其无先验性地解决高复杂度问题的良好性能在交通流量预测领域取得了一定的成果. 但 ANN 是基于经验风险最小化的, 存在收敛速度慢和训练误差非凸的问题, 难以求得最优解^[3].

支持向量机(SVM)最早是为了解决模式识

别和分类问题的, 随着 Vapnik 的 ϵ 不敏感损失函数的引入, SVM 开始用于解决非线性回归估值问题, 称之为支持向量回归(SVR). SVR 能很好地解决小样本、非线性和高维的问题, 已经成功地用于交通系统领域的预测^[2-3]. 然而, SVR 模型不能得到准确的预测结果是因为 SVR 模型中参数组合 C 、 ϵ 和 σ 的值设置得不准确. 传统的人工选择和网格遍历的方法存在主观经验判断和寻优时间长的缺点, 并不能得到好的效果, 但是运用智能的选择进化算法(如 ACO、PSO 和 AFS 等)优化 SVR 参数能选择到合适的回归参数^[6-7]. 每种智能进化算法都有各自的优缺点. 研究表明, 采用混合优化算法比采用单一优化算法能取得更好的预测结果^[8-9].

本文针对 AFS 算法的初始参数设置较多和步长因子的设置对寻优性能的影响的缺点, 提出采用 PSO 算法公式改进 AFS 算法, 减小步长因子对 AFS 算法的影响, 并且引入混沌机制初始化鱼群位置, 来进行 SVR 的参数选择, 建立混沌粒子群人工鱼群混合优化 SVR 的 CPSOAFS-SVR 交通流量预测模型.

1 SVR 模型

SVR 的核心思想是基于结构风险最小化原则,通过定义一个非线性映射将输入数据集映射到高维特征空间,在这个高维特征空间存在一个线性函数 f 能够明确地表示输入数据集和输出数据集之间的非线性关系. SVR 函数定义如下:

$$f(\mathbf{x}) = \mathbf{w}^T \cdot \varphi(\mathbf{x}) + b \quad (1)$$

式中: f 决定预测值, $\varphi(\cdot)$ 为映射函数, $\mathbf{x}$ 为输入数据向量, $\mathbf{w}$ 和 b 分别为高维特征空间中的向量和常数,两者的值可以通过下式求得:

$$\begin{aligned} \min \quad & J = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\zeta_i^+ + \zeta_i^-) \\ \text{s. t.} \quad & y_i - \mathbf{w}^T \varphi(\mathbf{x}_i) - b \leq \epsilon + \zeta_i^+ \\ & \mathbf{w}^T \varphi(\mathbf{x}_i) + b - y_i \leq \epsilon + \zeta_i^- \\ & \zeta_i^+, \zeta_i^- \geq 0, i = 1, 2, \dots, n \end{aligned} \quad (2)$$

式中: C 为惩罚系数, ϵ 为不敏感损失函数的最大误差系数, ζ_i^+ 和 ζ_i^- 为松弛系数, n 为输入数据的样本个数. SVR 致力于求解高维特征空间的最优超平面,使得输入输出数据集之间的整体误差最小.

通过求解上述带不等式约束的二次优化问题,可以得到式(1)中的参数向量 $\mathbf{w}$ 的值:

$$\mathbf{w} = \sum_{i=1}^n (\beta_i^* - \beta_i) \varphi(\mathbf{x}_i) \quad (3)$$

其中 β_i^* 和 β_i 分别为运用拉格朗日乘子法求解二次优化问题的拉格朗日系数. 最后 SVR 问题就得到如下方程:

$$f(\mathbf{x}) = \sum_{i=1}^n (\beta_i^* - \beta_i) K(\mathbf{x}_i, \mathbf{x}_j) + b \quad (4)$$

其中 $K(\cdot)$ 为计算高维特征空间中的两个输入向量内积的核函数. 本文采用 Gauss 函数 $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma^2)$ 作为模型的核函数.

SVR 模型中参数组合 C 、 ϵ 和 σ 的选择影响预测的精度,所以建立优化的交通流量预测模型就需要选取最优的参数组合来最小化实际值和预测值之间的误差^[7].

2 CPSOAFS-SVR 预测模型

2.1 AFS 算法

人工鱼群(AFS)算法是由李晓磊于 2003 年

提出的通过模拟自然界中鱼群的行为得到的一种群智能优化算法^[10]. 每条人工鱼代表 D 维空间中的一个点 $\mathbf{x}_i$, 人工鱼个体所在位置的适应度为 F_i , 本文的交通流量预测致力于 F_i 最小化. 两条人工鱼之间的距离为 $\|\mathbf{x}_i - \mathbf{x}_j\|$, 其他的参数包含鱼群个数 n 、视野 v 、步长 s 和拥挤度 δ . 每条人工鱼通过模拟鱼的觅食、聚群、追尾和随机行为寻找最优适应度来进行搜索优化过程.

(1) 觅食行为

觅食行为是人工鱼趋向较优适应度的基本行为. 其计算公式为

$$\begin{aligned} \mathbf{x}_{i+1} &= \mathbf{x}_i + \mathbf{r}^1 \cdot s \cdot \frac{\mathbf{x}_j - \mathbf{x}_i}{\|\mathbf{x}_j - \mathbf{x}_i\|}; F_j < F_i \\ \mathbf{x}_{i+1} &= \mathbf{x}_i + \mathbf{r}^2 \cdot s \end{aligned} \quad (5)$$

式中: $\mathbf{r}^1$ 和 $\mathbf{r}^2$ 分别代表随机变量区间 $[0, 1]$ 和 $[-1, 1]$, $\mathbf{x}_i$ 代表第 i 条人工鱼位置信息, F_j 为 $\mathbf{x}_j$ 在视野范围内随机移动到另一位置 $\mathbf{x}_j$ 的适应度. 若反复尝试若干次后,仍不满足前进条件,则随机移动一步.

(2) 聚群行为

聚群行为是人工鱼个体在游动过程中为了趋利避害,自然地聚集成群向邻居中心移动的行为. 其计算公式为

$$\begin{aligned} \mathbf{x}_{i+1} &= \mathbf{x}_i + \mathbf{r} \cdot s \cdot \frac{\mathbf{x}_c - \mathbf{x}_i}{\|\mathbf{x}_c - \mathbf{x}_i\|}; \\ F_c &< F_i, \frac{n_s}{n} < \delta \end{aligned} \quad (6)$$

其中 n_s 为当前视野范围内的邻居数目, $\mathbf{x}_c$ 代表所有邻居的聚集群体的中心位置. 当 $\mathbf{x}_c$ 的适应度优于当前 $\mathbf{x}_i$ 位置的适应度,并且鱼群不太拥挤时, $\mathbf{x}_i$ 就向中心位置移动. 否则,执行觅食行为.

(3) 追尾行为

追尾行为是一种向着附近的有更优适应度的人工鱼移动的行为. 其计算公式为

$$\begin{aligned} \mathbf{x}_{i+1} &= \mathbf{x}_i + \mathbf{r} \cdot s \cdot \frac{\mathbf{x}_{\min} - \mathbf{x}_i}{\|\mathbf{x}_{\min} - \mathbf{x}_i\|}; \\ F_{\min} &< F_i, \frac{n_s}{n} < \delta \end{aligned} \quad (7)$$

其中 $\mathbf{x}_{\min}$ 代表视野范围内的最优适应度的人工鱼所在位置. 当 $\mathbf{x}_{\min}$ 的适应度优于当前 $\mathbf{x}_i$ 位置的适应度,并且鱼群不太拥挤时, $\mathbf{x}_i$ 就向最优适应度

位置移动. 否则, 执行觅食行为.

另外, AFS 算法设置了公告牌来记录最优状态和当前迭代的最优人工鱼位置信息.

2.2 PSO 算法

PSO 算法是一种广泛用于解决数值优化问题的群智能优化算法. PSO 算法假设解空间内的每个位置为一个粒子, 每个粒子具有由被优化函数确定的适应度, 同时每个粒子按照一定的搜索规则和较优适应度方向对自身的速度及位置进行调整来求得最优解. PSO 算法的速度和位置更新公式为

$$\begin{aligned} v_{j+1} &= \omega v_j + c_1 r_1 (p_{bj} - x_j) + c_2 r_2 (g_{bj} - x_j); \\ x_{j+1} &= x_j + v_{j+1} \end{aligned} \tag{8}$$

式中: v_j 为粒子当前速度, v_{j+1} 为更新后的粒子速度, ω 为速度惯性权重, c_1 和 c_2 为加速系数, r_1 和 r_2 为区间 $(0, 1)$ 内的随机数, x_j 和 x_{j+1} 分别为粒子的当前位置和更新后的位置, p_{bj} 为粒子当前自身最优位置, g_{bj} 为粒子当前全局最优位置.

2.3 PSO 算法公式改进 AFS 算法

AFS 算法初始参数的设置对算法性能有一定的影响. AFS 算法的步长因子限制了人工鱼个体的移动步长, 如果步长太小就将个体鱼限制在较小搜索空间, 增大了寻优时间; 如果步长太大, 则可能找不到最优值. 并且步长因子依赖于人工鱼之间距离的计算, 对算法性能影响较大^[10-11].

针对上述步长因子对 AFS 算法性能影响的问题, 本文使用 PSO 算法公式改进 AFS 算法, 以减小步长因子对 AFS 算法性能的影响. AFS 算法中鱼的觅食、聚群和追尾行为的式 (5)、(6) 和 (7) 修改为如下新的公式:

$$\begin{aligned} x_{i+1} &= x_i + c_1 r_1 (x_j - x_i) + c_2 r_2 (x_b - x_i); \\ F_j &< F_i \end{aligned} \tag{9}$$

$$\begin{aligned} x_{i+1} &= x_i + r_3 \cdot v \\ x_{i+1} &= x_i + c_1 r_1 (x_c - x_j) + c_2 r_2 (x_b - x_j); \\ F_c &< F_i, \frac{n_s}{n} < \delta \end{aligned} \tag{10}$$

$$\begin{aligned} x_{i+1} &= x_i + c_1 r_1 (x_{\min} - x_j) + c_2 r_2 (x_b - x_j); \\ F_{\min} &< F_i, \frac{n_s}{n} < \delta \end{aligned} \tag{11}$$

这样 AFS 算法中的人工鱼个体就可以像 PSO

算法中的粒子一样, 只依赖于视野因子而不依赖于步长因子进行游动行为的选择, 寻求最优解.

2.4 混沌机制

混沌 (chaotic) 是非线性系统中普遍存在的现象, 具有确定性和遍历性的特点. 在群智能算法中, 混沌机制的运用能够提高种群的多样性, 进而避免算法出现过早收敛现象, 增大搜索到全局最优的性能^[7,12]. 本文选取 Logistic 映射方法混沌初始化人工鱼群位置, 从而寻求 SVR 参数全局最优组合. 选取的混沌初始化映射方法公式为

$$Z_{i+1} = \mu Z_i (1 - Z_i); i = 1, 2, \dots, n \tag{12}$$

其中 μ 为控制参数, 对系统状态起关键作用. 当 $\mu = 4$ 时, 系统处于完全混沌状态.

2.5 CPSOAFS-SVR 模型

本文运用 PSO 公式机制优化 AFS 的混合优化算法优化 SVR, 建立了混沌粒子群人工鱼群混合优化 SVR 的 CPSOAFS-SVR 交通流量预测模型. 模型具体流程如图 1 所示. 具体实现步骤如下:

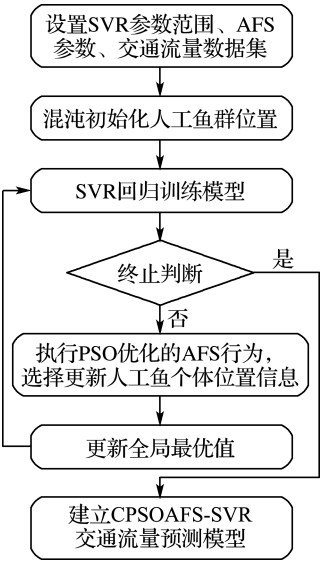


图 1 CPSOAFS-SVR 交通流量预测模型

Fig. 1 CPSOAFS-SVR traffic flow forecasting model

步骤 1 数据集设置和参数初始化. 将交通流量数据分为训练集和测试集, 进行数据预处理. 设置 SVR 的参数 C 、 ϵ 和 σ 的取值范围以及 PSO 和 AFS 算法的初始参数. 混沌初始化人工鱼群位置.

步骤 2 计算鱼群的初始适应度. 根据训练集训练 SVR 模型, 选出当前最优并记录初始全局最优的人工鱼的状态, 即参数组合 (C, ϵ, σ) .

步骤 3 人工鱼行为选择. 每条人工鱼分别执行混合鱼群算法的觅食、聚群、追尾和随机行为, 更新人工鱼个体的位置信息, 保存全局最优值以及对应的 SVR 参数组合.

步骤 4 算法的终止判断. 判断是否达到鱼群的最大迭代次数, 若不满足则迭代次数加 1, 跳转至步骤 3 继续执行; 若满足则输出鱼群的最优适应度和最优值对应的 SVR 参数组合, 建立 CPSOAFS-SVR 交通流量预测模型, 并用于测试集的回归测试, 检验测试性能.

3 仿真实验及结果

本文分别采用混沌 PSO 优化 SVR (CPSO-SVR) 模型和混沌 AFS 优化 SVR (CAFS-SVR) 模型作为参考模型, 进行对比仿真实验.

3.1 数据集准备

本文采用加利福尼亚高速公路性能评估系统 (PeMS) 采集的实际交通流量数据作为实验数据来源, 数据样本时间间隔为 5 min. 实验选取 1 d 中的早 6:00~10:00 的交通流量作为实验数据集. 实验数据采用工作日情境下的 5 个站点数据, 每个站点的训练集有 576 个数据, 预测第 25 d 的 48 个数据.

实验采用历史时间序列的交通流量数据预测未来时刻的交通流量. 设 $x_i(t)$ 为路段 i 上 t 时刻的交通流量, 实验采用前 5 个时间段的交通流量 $x_i(t-5)$ 、 $x_i(t-4)$ 、 $x_i(t-3)$ 、 $x_i(t-2)$ 和 $x_i(t-1)$ 作为输入, $x_i(t)$ 作为模型的预测输出^[3].

3.2 评价指标和参数设置

为综合比较各模型的仿真实验结果和预测性能, 实验采用均方误差 e_{ms} 、平均绝对百分误差 e_{map} 和均方根误差 e_{rms} 共 3 个评价指标:

$$e_{\text{ms}} = \frac{1}{n} \sum_{i=1}^n (y_i - y_i^*)^2 \tag{13}$$

$$e_{\text{map}} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - y_i^*}{y_i} \right| \times 100\% \tag{14}$$

$$e_{\text{rms}} = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - y_i^*}{y_i} \right)^2} \times 100\% \tag{15}$$

其中 y_i 表示实际值, y_i^* 表示预测值, n 为测试样本个数. 评价指标值越小, 预测性能越好.

实验的初始参数设置如表 1 所示. 混合 PSO-AFS 优化算法中参数与 PSO 和 AFS 算法中的相同.

表 1 初始参数设置

Tab. 1 Initial parameter settings

算法	参数	值
SVR	C	[10, 100]
	ϵ	[0.01, 0.50]
	σ	[0.01, 10.00]
PSO	ω	1.0
	c_1, c_2	1.7, 1.7
	n, n_{max}	20, 30
AFS	v	0.3
	s	0.2
	n_{try}	10
	δ	0.7

3.3 实验结果和分析

本实验利用 libsvm-3.17 Matlab 工具箱^[13]实现了基于上述各算法优化支持向量回归的交通流量预测模型. 表 2 给出了站点 1 各个模型下选取的 SVR 的参数 C 、 ϵ 和 σ 的最优组合.

表 2 站点 1 各个模型的参数优化结果

Tab. 2 Parameter optimization results for each model on site 1

算法	C	ϵ	σ
CPSO-SVR	54.030	0.010	0.043
CAFS-SVR	94.380	0.136	0.015
CPSOAFS-SVR	45.490	0.154	0.103

图 2 显示了站点 1 的测试样本在 CPSO-SVR、CAFS-SVR 和 CPSOAFS-SVR 模型下的预测值与实际值, 表 3 给出了 3 个对比模型的评价结果. 如图 2 和表 3 所示, CPSO-SVR 和 CAFS-SVR 模型预测结果相近, 而本文提出的 CPSOAFS-SVR 模型比 CPSO-SVR 和 CAFS-SVR 模型的 e_{map} 分别减小了 8.9% 和 8.1%, 提高了预测精度, 证明了 CPSOAFS-SVR 模型的准确性.

更一般的,分别对 5 个站点的数据集进行实验,因为 e_{map} 最能反映预测性能,所以表 4 给出了 5 个站点的 e_{map} . 结果显示,在不同站点的测试数据集情况下,CPSOAFS-SVR 模型都有最优的预测结果,其 e_{map} 平均值为 4.108%,比 CPSO-SVR 模型(e_{map} 平均值为 4.312%)和 CAFS-SVR 模型(e_{map} 平均值为 4.274%)分别减小了 5.0%和 4.0%,从而证明了 CPSOAFS-SVR 模型的健壮性.

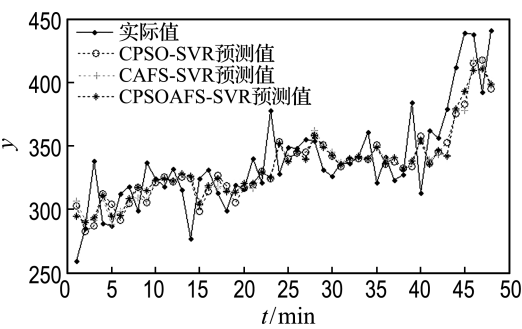


图 2 站点 1 各个模型的实际值与预测值

Fig.2 Actual and predicted values for each model on site 1

表 3 站点 1 各个模型的评价结果

Tab.3 Evaluation results for each model on site 1

算法	e_{ms}	$e_{\text{map}}/\%$	$e_{\text{rms}}/\%$
CPSO-SVR	682.51	6.22	7.59
CAFS-SVR	699.71	6.17	7.62
CPSOAFS-SVR	632.16	5.71	7.30

表 4 5 个站点的 e_{map} 结果对比

Tab.4 Comparison of 5 sites' e_{map} results

站点	$e_{\text{map}}/\%$		
	CPSO-SVR	CAFS-SVR	CPSOAFS-SVR
1	6.220	6.170	5.710
2	2.310	2.260	2.210
3	4.520	4.520	4.480
4	3.850	3.810	3.780
5	4.660	4.610	4.360

4 结 语

准确的交通流量预测对交通控制系统是至关重要的,特别是在针对车流高峰期避免碰撞和提高交通效率方面.本文提出将基于混沌初始化机

制的混合人工鱼群算法优化 SVR 的预测方法,应用到交通流量预测领域来研究模型的可行性.仿真结果表明,提出的 CPSOAFS-SVR 交通流量预测模型比单一的 PSO 和 AFS 优化 SVR 模型有更好的预测性能.模型良好的预测结果归功于用 PSO 算法减小了 AFS 算法中步长因子的影响,从而选出更优的 SVR 参数组合,并且混沌初始化机制的运用能够更好地搜索解空间,取得最优预测值.未来将研究一些其他的混合群智能优化算法来更有效地解决交通流量预测问题.

参考文献:

[1] Vlahogianni E I, Karlaftis M G, Golias J C. Short-term traffic forecasting: Where we are and where we're going [J]. **Transportation Research Part C: Emerging Technologies**, 2014, **43**:3-19.

[2] 姚智胜,邵春福,高永亮. 基于支持向量回归机的交通状态短时预测方法研究[J]. 北京交通大学学报, 2006, **30**(3):19-22.

YAO Zhi-sheng, SHAO Chun-fu, GAO Yong-liang. Research on methods of short-term traffic forecasting based on support vector regression [J]. **Journal of Beijing Jiaotong University**, 2006, **30**(3): 19-22. (in Chinese)

[3] Castro-Neto M, Jeong Y S, Jeong M K, et al. Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions [J]. **Expert Systems with Applications**, 2009, **36**(3 Part 2):6164-6173.

[4] Smith B L, Williams B M, Keith Oswald R. Comparison of parametric and nonparametric models for traffic flow forecasting [J]. **Transportation Research Part C: Emerging Technologies**, 2002, **10**(4):303-321.

[5] Vlahogianni E I, Karlaftis M G, Golias J C. Optimized and meta-optimized neural networks for short-term traffic flow prediction: A genetic approach [J]. **Transportation Research Part C: Emerging Technologies**, 2005, **13**(3):211-234.

[6] Hong W C, Dong Y, Zheng F, et al. Forecasting urban traffic flow by SVR with continuous ACO [J]. **Applied Mathematical Modelling**, 2011,

35(3):1282-1291.

[7] Hong W C. Traffic flow forecasting by seasonal SVR with chaotic simulated annealing algorithm [J]. **Neurocomputing**, 2011, **74**(12-13):2096-2107.

[8] Fang N, Zhou J, Zhang R, *et al.* A hybrid of real coded genetic algorithm and artificial fish swarm algorithm for short-term optimal hydrothermal scheduling [J]. **International Journal of Electrical Power and Energy Systems**, 2014, **62**:617-629.

[9] 姚祥光,周永权,李咏梅. 人工鱼群与微粒群混合优化算法 [J]. 计算机应用研究, 2010, **27**(6):2084-2086, 2102.

YAO Xiang-guang, ZHOU Yong-quan, LI Yong-mei. Hybrid algorithm with artificial fish swarm algorithm and PSO [J]. **Application Research of Computers**, 2010, **27**(6): 2084-2086, 2102. (in Chinese)

[10] 李晓磊. 一种新型的智能优化方法——人工鱼群算法[D]. 杭州:浙江大学, 2003.

LI Xiao-lei. A new intelligent optimization method—artificial fish school algorithm [D]. Hangzhou:Zhejiang University, 2003. (in Chinese)

[11] Tsai H C, Lin Y H. Modification of the fish swarm algorithm with particle swarm optimization formulation and communication behavior [J]. **Applied Soft Computing Journal**, 2011, **11**(8):5367-5374.

[12] LI Ming-wei, Hong Wei-chiang, KANG Hai-gui. Urban traffic flow forecasting using Gauss-SVR with cat mapping, cloud model and PSO hybrid algorithm [J]. **Neurocomputing**, 2013, **99**:230-240.

[13] Chang C C, Lin C J. LIBSVM: A library for support vector machines [J]. **ACM Transactions on Intelligent Systems and Technology**, 2011, **2**(3):27.

Traffic flow forecasting based on optimized SVR
with hybrid artificial fish swarm algorithm

YAO Wei-hong*, FANG Ren-xiao, ZHANG Xu-dong

(School of Computer Science and Technology, Dalian University of Technology, Dalian 116024, China)

Abstract: Real-time and accurate traffic flow forecasting is the important content of intelligent transportation systems (ITS). Support vector regression (SVR) can be applied to traffic flow forecasting problem. For the disadvantage of parameter selection of SVR, a hybrid artificial fish swarm (AFS) algorithm is constructed. In order to minimize the impact of the step factor of AFS algorithm, the algorithm uses particle swarm optimization (PSO) formulation to improve AFS algorithm. In addition, chaotic initialization of AFS mechanism is introduced. Therefore, the optimal SVR parameters are selected. Then, an improved traffic flow forecasting model which combines the SVR with chaotic PSO-AFS (CPSOAFS-SVR) is constructed. Simulation results show that the traffic flow forecasting model has better prediction performance, and its feasibility and effectiveness are proved.

Key words: traffic flow forecasting; support vector regression (SVR); artificial fish swarm (AFS) algorithm; particle swarm optimization (PSO); chaotic mechanism