

基于 PSO-SVR 模型的煤炭铁路物流需求预测

陈培友¹, 刘 璐²

(黑龙江科技大学 1. 经济管理系; 2. 研究生学院, 哈尔滨 150027)

摘要:在煤炭铁路物流需求预测中,存在历史样本量较小和非线性强的特点,从而致使预测精度较低。将支持向量回归机 (support vector regression, SVR)与粒子群算法 (Particle Swarm Optimization, PSO)相结合,提出适用于小样本量学习的 PSO-SVR 模型。选取 1995—2011 年的煤炭铁路货运量及其影响因素作为学习样本,利用粒子群算法对支持向量机参数进行优化,通过训练、测试得到具有良好学习与推广能力的煤炭铁路货运量预测模型。建立 BP 神经网络模型,并将二者的预测值进行对比,结果表明在解决我国煤炭铁路物流需求预测这种小样本,非线性及高维模式识别问题中 PSO-SVR 模型预测精度优于 BP 神经网络模型。

关键词:煤炭铁路物流需求预测;支持向量回归机;粒子群算法

中图分类号:F252.8;U294.1+3 **文献标志码:**A **文章编号:**1671—1807(2013)12—0051—06

煤炭是我国重要的基础能源,在一次商品能源消费中约占 70%,并在未来相当长的时期内仍是支撑我国经济高速增长的主要能源保障。我国煤炭资源西多东少,北富南贫的不均匀分布决定了中国煤炭物流的基本格局,即“北煤南运,西煤东送”。长期以来,由于煤炭运能与实际运量的突出矛盾,使得运输瓶颈一直都是煤炭产业发展中最为严重的制约因素。本文通过结合我国煤炭物流发展现状,将支持向量回归机引入煤炭铁路物流需求预测中,对我国煤炭铁路货运量的影响因素进行分析,以便于政府合理规划、建设铁路煤炭通道物流基础设施,改进煤炭物流供给系统。

1 煤炭物流及运输现状

煤炭物流是一个系统物流,它存在于煤炭产品的开发准备、生产过程和销售活动的全过程之中,是生产煤炭产品以及组织煤炭销售等一系列物料实体的运送搬运等动态流转过程。如图 1 所示的煤炭物流网络体系结构模型。由图 1^[1]可知,煤炭物流是集生产地、中转地和消费地为一体,物流、信息流、资金流等贯穿其中,多主体、多结点、多通道、跨地区的动态性复杂网络体系。煤炭集散、转运和配送中心作为三类重要物流节点,是实施煤炭物流管理的重要平台。由于煤炭属于散装货物具有体积大,污染环境等特

性,以及我国煤炭的分布与消费的地理差距的制约,所以在集散、转运和配送过程中的主要技术在运输方面。

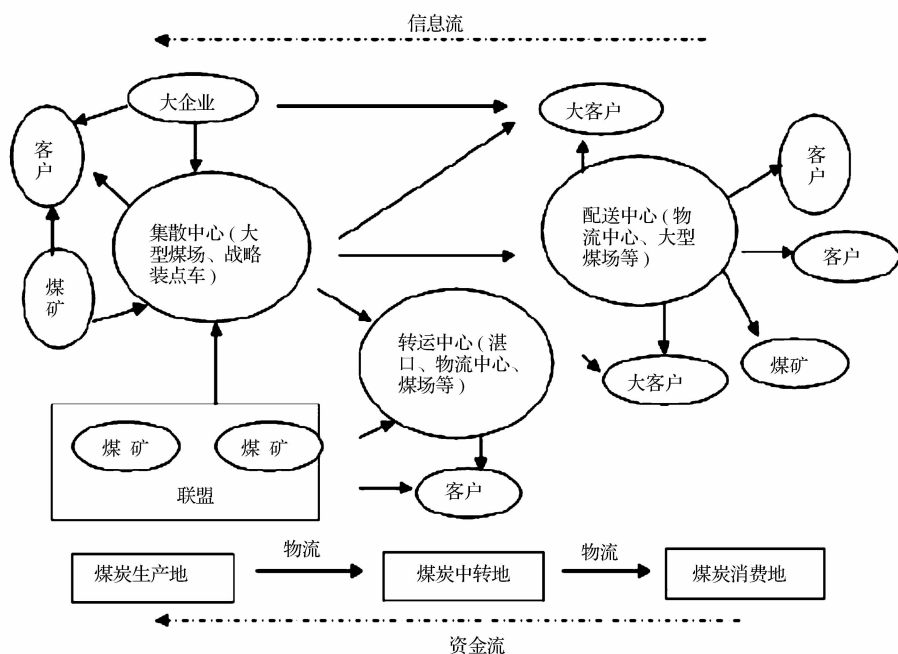
煤炭的运输可分为铁路运输、公路运输、水路运输三种主要运输方式。三者的区别见表 1。由表 1 知,我国的煤炭 60%以上依靠铁路运送。虽然我国在铁路运输的生产效率指标上处于领先地位,但由于缺乏及时的沟通交流和明确的统筹规划,缺乏相应的投入与建设,使得煤炭运输体系存在很多的问题,如铁路运输通道建设严重滞后,煤炭需求与铁路运能之间存在着结构性矛盾,需求大的东南地区缺少运力,而运力比较充足的西北地区相对缺少需求。通过对煤炭铁路货运量主要影响因素的有效把握,借助于定性和定量的分析手段预测比较精确的需求量,为政府提供制定铁路煤炭通道发展规划的重要依据。

表 1 铁路、水路和公路运输的区别^[2]

铁路	水路	公路	
运输比例	60%	30%	10%
运量	多	较多	少
成本	低	较低	高
灵活程度	差	较好	好
运输长度	长	长	短

收稿日期:2013—06—21

作者简介:陈培友(1967—),男,黑龙江鸡西人,黑龙江科技大学教授,管理学博士,硕士研究生导师,研究方向:系统工程、电子商务建模与优化。



2 煤炭铁路物流需求预测指标选取

目前,关于物流需求一般认为包括质和量两个方面,即物流需求质量指标(物流时间,物流成本,物流效率等)和物流需求规模指标,在定量预测中基本上没有使用物流需求质量指标。多数学者用货运量、社会物流总费用、物流成本占 GDP 比例三种数据中的一种来衡量物流需求规模^[3]。我国由于现阶段并没有物流的具体统计数据和专门的统计渠道,因此在实际预测过程中,一般把货运量作为进行物流量分析的类比指标^[4]来进行物流需求的预测与分析。虽然煤炭货运量只是煤炭物流需求量中的一部分,但煤炭的运输贯穿于煤炭物流活动的始终,是联系其他物流服务功能的纽带。据统计,每年的货运量占物流总成本的一半以上,因此,煤炭铁路货运量在一定程度上也是能够反映煤炭铁路物流需求规模变化的。

煤炭铁路货运量不仅受国民经济因素,如国民经济发展水平、产业结构等方面的影响,还受煤炭需求因素,如煤炭消费量、能源结构,此外,与铁路自身供给因素,如铁路货车拥有量,货车平均静载重等因素也有关系。各因素对煤炭铁路货运量影响程度不同并且随时间发生变化,这种错综复杂的关系决定煤炭铁路货运量与各影响因素之间存在着复杂的非线性关系,难以用精确的数学模型进行描述。

目前,对于预测问题可以建立回归模型或时间序列模型予以解决,这是两类使用广泛的定量预测方

法,以数学理论和假设为基础,通过演绎推理建立数学模型,但是很难全面、本质地揭示数据的内在结构和复杂特性。近年来神经网络作为一种非线性系统模型,具有良好的自学习、自适应能力和泛化容错能力,能够较好地解释因变量与影响因素之间复杂的非线性关系,因而被用于能源需求预测^[5]和货运量预测中^[6]。由于神经网络本身固有的缺陷,实际应用中常常发生收敛速度慢、训练时间长、极易陷入局部极小值、过拟合和“维数灾难”等问题。支持向量机(SVM)以统计学中的 VC 维理论为基础,综合考虑经验风险和置信风险,遵循结构风险最小化原则,适用于小样本、非线性、高维数、局部极小等问题,在有限数据样本情况下,其泛化能力的提高,能取得较好的预测效果^[7-8]。因此,本文选定支持向量回归机模型对煤炭铁路货运量进行预测。

3 基本原理介绍

3.1 支持向量回归机算法

设定数据训练集为 $T = \{(x_1, y_1), \dots, (x_i, y_i)\} \in (X \times Y)^l$, $x_i \in X = R^n$, $y_i \in Y = R$, $i = 1, 2, \dots, l$ 。 N 维的系统输入向量, 其中 x_i 为输入矢量, y_i 为相应的目标值。其基本思想就是通过非线性映射 ϕ , 将输入空间 X 的样本数据 x_i 映射到某一高维空间 (Hilbert 空间) 中, 从而将输入空间 X 中的非线性函数估计问题, 转化为高维空间的线性函数估计问题^[9]。

标准 ϵ -SVR 算法的原始最优化问题为:

$$\min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (1)$$

约束条件为

$$s. t. \begin{cases} \omega \cdot \phi(x) + b - y_i \leq \xi_i + \epsilon, i = 1, 2, \dots, n \\ y_i - \omega \cdot \phi(x) - b \leq \xi_i^* + \epsilon, i = 1, 2, \dots, n \\ \xi_i, \xi_i^* \geq 0, i = 1, 2, \dots, n \end{cases} \quad (2)$$

引入 Lagrange 算子 α 和 α^* , 分别对 ω, b, ξ, ξ^* 求偏导, 并令其为 0, 将(4)的优化问题转化为对偶形式进行求解:

$$\max \omega(\alpha_i, \alpha_i^*) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) y_i - \frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) \langle \phi(x_i) \cdot \phi(x_j) \rangle - \sum_{i=1}^n (\alpha_i + \alpha_i^*) \epsilon \quad (3)$$

约束条件为

$$\sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0, \alpha_i, \alpha_i^* \in [0, C]$$

求解规划(3)最终确定回归函数为:

$$f(x) = \omega \cdot \phi(x) + b = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x_i, x) + b \quad (4)$$

$K(x_i, x)$ 称为核函数, 本文选取 $K(x_i, x) = \exp\{-\|x_k - x_i\|^2 / 2\sigma^2\}$ 作为 SVM 模型的核函数。将其带入(4)经过等价代换得:

$$f(x) = \sum_{k=1}^n \alpha_k \exp\{-\|x_k - x_i\|^2 / 2\sigma^2\} + b \quad (5)$$

式(5)中: α_k 是支持向量所对应的参数值, x_k 是训练年份输入向量, x_i 是预测年份输入向量, $f(x)$ 为输出向量集合, 由式(5)计算得出 α_k 和 b , 即可得煤炭铁路货运量预测模型^[8]。

3.2 标准微粒群算法

PSO 是由 Kenney 等人^[10]于 1995 年提出的一种全局优化进化算法, 假设在一个 D 维的目标搜索空间中, 有 N 个粒子组成一个群落, 每个粒子代表解空间的一个候选解, 解的优劣程度由适应度函数决定, 其中适应度函数根据优化目标定义。每个粒子还有一个速度决定其飞翔的方向和距离, 然后粒子们就追随当前的最优粒子在解空间中搜索个体极值和全局极值。在找到这两个最优值时, 粒子根据如下的式(6)和式(7)来更新自己的速度和位置^[11]:

$$v_{id}^{k+1} = \omega \times v_{id}^k + c_1 r_1 (p_{id}^k - x_{id}^k) + c_2 r_2 (p_{gd}^k - x_{id}^k) \quad (6)$$

$$x_{id}^{k+1} = x_{id}^k + v_{id}^{k+1} \quad (7)$$

式中, v_{id}^{k+1} 是第 i 个粒子在 $k+1$ 代的飞行速度, $v_{id} \in [-v_{\max}, v_{\max}]$, $v_{\max}$ 是常数。 ω 为惯性权重, $i = 1, 2, \dots, N$, $d = 1, 2, \dots, D$; c_1, c_2 为非负常数, 称为加速因子, 根据经验, 通常 $c_1 = c_2 = 2$, 后来也有理论说明 c_1 和 c_2 不应该相等, 所以在编程实现时设定为不相等。 r_1 和 r_2 是介于 $[0, 1]$ 之间的随机数。

4 PSO-SVR 模型

在目前对于 SVR 参数的选择中, 没有一个统一的方法, 主要有经验确定和网格搜索。但经验确定法要求使用者有深厚的 SVR 理论基础; 而网格搜索法的计算量较大, 因此两种方法并不能保证找到全局最优解。并且相较于其他一些成熟的智能优化算法如遗传算法, PSO 在编码和优化策略上更加简单、高效, 因此本文采用 PSO 对 SVR 参数进行优化。

需要对 SVM 的惩罚系数 C 、高斯径向基核参数 σ 、不敏感损失系数 ϵ 进行优化。首先设置参数 ϵ 较小的初始值, 利用 PSO 对模型参数进行优选, 寻求均方误差最小的参数对, 逐步提高参数 ϵ 值, 一直到满意为止。

PSO-SVR 模型的工作流程图如下:

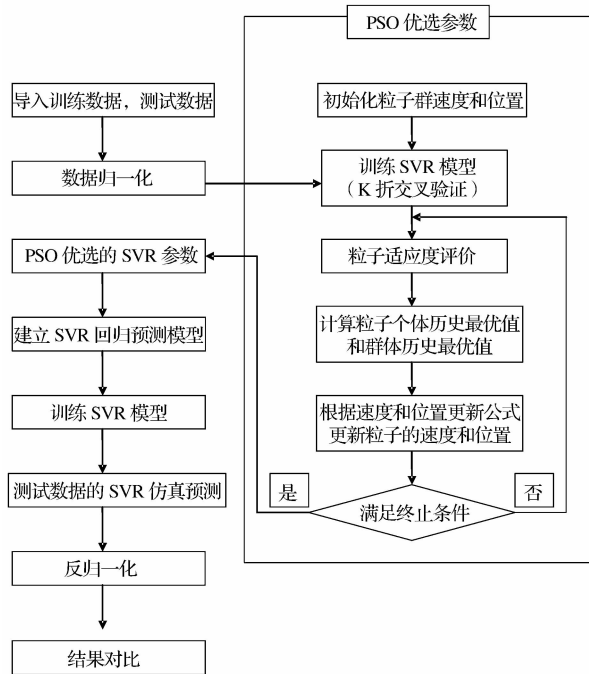


图 2 PSO-SVR 模型工作流程图

5 PSO-SVR 的煤炭铁路货运量预测模型的构建

5.1 影响因素的选取

选取 1995—2011 年我国煤炭铁路货运量为系统特

征行为序列,各影响因素为相关因素序列,为确定各因素对货运量的影响大小,本文通过灰色关联分析法^[12]确定煤炭铁路货运量 Y 及影响因素之间的关系,从而选取主要影响因素。各类影响因素的划分详见表 2。

表 2 各类影响因素一览表

类别划分	影响因素
国民经济因素	国内生产总值(GDP),第一产业生产总值,第二产业生产总值,第三产业生产总值,产业结构(工业产值占 GDP 的比重)
煤炭需求因素	煤炭消费总量、能源结构(煤炭的消费比重),城镇化率,单位 GDP 能耗
铁路供给因素	铁路货运周转量,铁路货车拥有量,货车平均静载重,铁路营业里程

煤炭铁路货运量及影响因素的数量级别差异较大,需要先通过灰色序列算子对各个数据序列进行适当预处理,将其转化为数量级别大体相近的无量纲数据。这里选择初值化算子进行数据预处理,分别得到

$$\gamma_{0i}(k)=\frac{\min_i\min_k\{|x_0(k)-x_i(k)|\}+\xi\max_i\max_k\{|x_0(k)-x_i(k)|\}}{|x_0(k)-x_i(k)|+\xi\max_i\max_k\{|x_0(k)-x_i(k)|\}}$$

(8)

$$\gamma_{0i}=\frac{1}{n}\sum_{k=1}^nr_{0i}(k)$$

(9)

计算关联度及关联序时,需确定分辨系数 ξ。取 ξ=0.5,铁路货运量及其影响因素灰色关联分析结果见表 3。

根据对 Y 关联度的大小,最终选定排名前 7 的影响因素,分别是:铁路货运周转量 X₁(亿吨公里),煤炭消费总量(万吨)X₂,城镇化率(%)X₃,铁路货车拥有量(量)X₄,货车平均静载重(吨)X₅,产业结构(%)X₆,能源结构(%)X₇。

5.2 煤炭铁路货运量预测模型的构建

将 1995—2011 年的煤炭铁路货运量影响因素和

煤炭铁路货运量及其影响因素初值像,然后根据关联系数计算式(8)和关联度计算式(9)计算煤炭铁路货运量(X₀)与各影响因素(X_i)之间的关联度 γ_{0i},并在此基础上得出灰色关联序^[13]。

货运量数据组成样本集,具体数据如表 4 所示:

表 3 影响因素的灰色关联分析结果

关联度	关联序	关联度	关联序
0.987 6	铁路货运周转量	0.900 9	第一产业生产总值
0.977 6	煤炭消费总量	0.897 6	铁路营业里程
0.948 8	城镇化率	0.855 9	第二产业生产总值
0.931 5	铁路货车拥有量	0.851 5	国内生产总值(GDP)
0.927 6	货车平均静载重	0.833	单位 GDP 能耗
0.926 9	产业结构	0.786 3	第三产业生产总值
0.919 3	能源结构		

表 4 煤炭铁路货运量与影响因素数据

年份	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	Y
1995	13 049.5	97 857.3	29.04	432 731	56.8	41.04	74.6	67 357
1996	13 106.2	99 366.1	30.48	443 893	57.1	41.37	73.5	72 058
1997	13 269.9	97 039.0	31.91	437 686	57.3	41.69	71.4	70 345
1998	12 560.1	96 554.5	33.35	439 326	57.6	40.31	70.9	64 081
1999	12 910.3	99 241.7	34.78	436 236	57.7	39.99	70.6	64 917
2000	13 770.5	100 707.5	36.22	439 943	57.9	40.35	69.2	68 545
2001	14 694.1	102 727.3	37.66	449 921	58.1	39.74	68.3	76 625
2002	15 658.4	108 413.1	39.09	469 583	58.2	39.42	68.0	81 852
2003	17 246.7	128 286.8	40.53	501 478	58.3	40.45	69.8	88 132
2004	19 288.8	148 351.9	41.76	520 101	59.3	40.79	69.5	99 210
2005	20 726.0	167 085.9	42.99	541 824	60.1	41.76	70.8	107 082
2006	21 954.4	183 918.6	44.34	558 483	60.9	42.21	71.1	112 034
2007	23 797.0	199 441.2	45.89	579 209	61.3	41.58	71.1	122 081
2008	25 106.3	204 887.9	46.99	584 961	62.0	41.48	70.3	134 325
2009	25 239.2	215 879.5	48.34	594 388	62.6	39.67	70.4	132 720
2010	27 644.1	220 958.5	49.95	622 284	63.1	40.03	68.0	156 020
2011	29 465.8	238 033.4	51.27	649 273	63.5	39.99	68.8	172 126

注:数据来源于中国统计年鉴(1996—2012 年)

1) 选择合适的学习样本是 SVR 预测模型的基础。将学习样本分为训练样本和测试样本两部分,先用训练样本学习训练,再用测试样本验证模型的有效性。本文选取 1995—2008 年的煤炭铁路货运量及其影响因素为训练样本,2009—2011 年的数据为测试样本。

2) 对样本集中的数据按照公式进行归一化

$$x_i' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (10)$$

$$y_i' = \frac{y_i - y_{\min}}{y_{\max} - y_{\min}} \quad (11)$$

3) 用 PSO 对模型参数进行优选^[12]。初始化粒子群,群体数量取 30,学习因子 $c_1 = 1.2$, $c_2 = 1$,惯性权重均取 1,最大迭代次数取 200。随机产生一组参数 (C, σ) 作为粒子的初始位置和速度,选用 5 折交叉验证误差最小作为 SVR 参数选择的目标值,设置参数 ϵ 较小的初始值,按照式(9)和式(10)更新粒子最优位置,判断终止条件,逐步提高参数 ϵ 的值,最终得到最佳参数值分别为 $C = 97.1572$, $\sigma = 0.01$, $\epsilon = 0.01$,交叉验证均方误差为 0.016103。

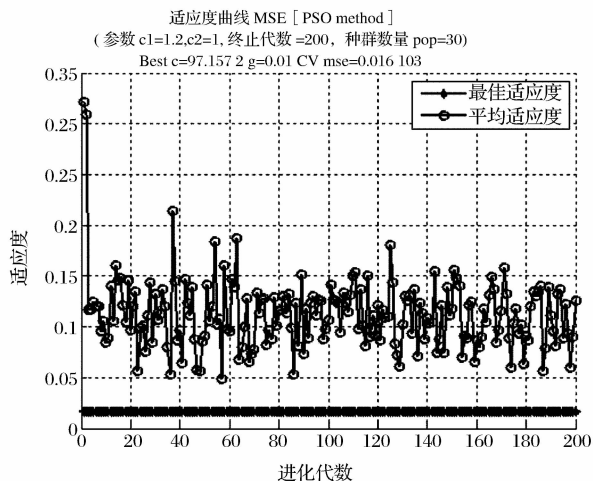


图3 适应度曲线图

4) 利用优化参数创建 SVR 模型,可得 $f(x)$ 表达式,即为煤炭铁路货运量预测模型。

$$f(x) = 6.5910K(x_2, x) - 8.3580K(x_5, x) + 0.5059K(x_8, x) - 22.9875K(x_{12}, x) + 24.2487K(x_{14}, x) - 0.5643$$

6 结果分析与比较

为了评价模型的预测性能,本文选择预测误差 $Error$ 、平均误差 E_{ME} 、均方根误差 $RMSE$ 、预测准确度 A 作为性能指标,来评价模型函数的拟合能力和误差^[14]。具体定义为:

$$1) \text{ 预测误差: } Error = \left| \frac{y_i - y_i^*}{y_i} \right| \quad (12)$$

$$2) \text{ 平均误差: } E_{ME} = \left| \frac{y_i - y_i^*}{y_i} \right| \quad (13)$$

$$3) \text{ 均方根误差: } RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i^* - y_i)^2} \times 100\% \quad (14)$$

$$4) \text{ 预测准确度: } A = 1 - RMSE = \left[1 - \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i^* - y_i)^2} \right] \times 100\% \quad (15)$$

其中 y_i^* 为预测值, y_i 为实际值, n 为样本数量。

为了更好的对比预测效果本文采用平均相对误差表示预测有效性,另外建立 BP 神经网络预测模型,输入层取 7 个神经元,隐含层取 10 个,输出层取 1 个,同样采用 PSO-SVR 模型中划分的学习样本进行预测分析。各预测模型的拟合结果见图 4。

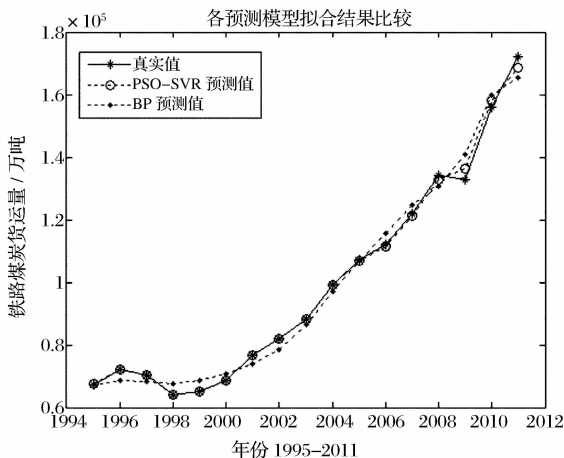


图4 各预测模型拟合结果

各模型预测值与真实值之间的比较如表 5 所示。

通过预测结果对比表明 PSO-SVR 煤炭铁路货运量预测模型的预测精度要优于 BP 神经网络,其预测准确率可达 97.937%,能够有效的预测煤炭铁路货运量的未来值。

7 结语

本文针对煤炭物流的复杂网络特性,分析运输现状,选取煤炭铁路物流需求预测指标,将支持向量机回归(SVR)与粒子群算法(PSO)相结合,提出了适用于小样本量学习的 PSO-SVR 预测模型。选取了铁路货运周转量,煤炭消费总量,城镇化率等七个影响因素以及 1995—2011 年的煤炭铁路货运量作为学习样本,建立“影响因素—煤炭铁路货运量”的 SVR 模型,结合粒子群算法优化选择模型参数;对煤炭铁

路货运量进行预测,实例结果表明:在解决我国煤炭铁路物流需求预测这种小样本、非线性及高维模式识别问题中,PSO-SVR 煤炭铁路货运量预测模型的预

测精度和误差相比 BP 神经网络预测都有一定的改进,是一个很有价值的煤炭铁路物流需求预测模型。

表 5 预测结果及误差对比

年份	实际值	BP 预测值	预测误差	PSO-SVR 预测值	预测误差
2009	132 720	141 050	6.382%	136 230	2.645%
2010	156 020	159 680	2.346%	157 990	1.263%
2011	172 126	165 360	3.931%	168 610	2.043%
平均误差 EMA		4.220%		1.984%	
均方根误差 RMSE		4.535%		2.063%	
预测精度 A		95.465%		97.937%	

参考文献

[1] 李琰,李红霞. 我国煤炭物流产业评价体系的构建[J]. 统计与决策,2012(15):182—183.

[2] 陈培友,等. 联盟运输调度的煤炭运输运作模式的研究[J]. 物流技术,2012,31(3):17—19.

[3] 孙有望,周福东. 我国宏观物流市场预测与分析方法研究[J]. 同济大学学报:自然科学版,2005,33(1):47—51.

[4] 王晓原,李军. 灰色 GM(1,1)模型在区域物流规模预测中的应用[J]. 武汉理工大学学报:交通科学与工程版,2005,29(3):613—615.

[5] 胡雪锦,赵国浩. 基于 Matlab 的 BP 神经网络煤炭需求预测模型[J]. 中国管理科学,2008,10(16):521—525.

[6] 吴晓玲,符卓. 基于多因素的铁路货运量 BP 神经网络预测研究[J]. 铁道货运,2009(10):33—36.

[7] 黄虎,严余松,等. 基于支持向量回归机的公路货运量预测模型[J]. 计算机应用研究,2008,25(2):632—634.

[8] 孙涵,杨普容,成金华. 基于 Matlab 支持向量回归机的能源需求预测模型[J]. 系统工程理论与实践,2011,31(10):2001—2007.

[9] 白鹏,张喜斌,张斌. 支持向量机理论及工程应用实例[M]. 西安:西安电子科技大学出版社,2008.

[10] KENNEDY J, EBERHART R. Particle swarm optimization//Proceedings of the 4th IEEE International Conference on Neural Networks[C]. Piscataway: IEEE Service Center,1995:1942—1948.

[11] LANGDON W B, POLI R. Evolving problems to learn about particle swarm and other optimizers//Proc. CEC-2005[C]. 2005:81—88.

[12] 刘思峰,党耀国,方志耕,谢乃明. 灰色系统理论及其应用[M]. 北京:科学出版社,2010:5.

[13] 熊伟丽,徐保国. 基于 PSO 的 SVR 参数优化选择方法研究[J]. 系统仿真学报,2006,18(9):2442—2445.

[14] 胡燕祝,吕宏义. 基于支持向量回归机的物流需求预测模型研究[J]. 物流技术,2008,27(5):66—68.

Coal Railway Logistics Demand Forecasting Based on PSO-SVR

CHEN Pei-you¹, LIU Lu²

(1. College of Economic Management; 2. College of Graduate Studies, Heilongjiang University of Science & Technology, Harbin 150027, China)

Abstract: Combined with the support vector regression machine (support vector regression, the SVR) and Particle Swarm Optimization algorithm, (Particle Swarm Optimization, PSO), the paper proposes PSO-SVR model which is suitable for the learning of small samples. The paper selects the data of coal railway freight volumes and influence factors from 1995 to 2011 as the learning samples and then uses the particle swarm algorithm to optimize model parameters. The SVR model with good learning and generalization ability is established through training and testing. Setting up the BP neural network model and comparing the predicted value of both, the results show that the prediction accuracy of PSO-SVR model is superior to the BP neural network model.

Key words: coal railway logistics demand forecasting; support vector regression machine; particle swarm algorithm