

基于因素分析和 SVM 模型的江海联运运量预测方法

杨丁红^{1,2,3}, 郑彭军^{1,2,3*}

(1. 宁波大学 海运学院, 浙江 宁波 315832; 2. 宁波港航物流服务体系协同创新中心, 浙江 宁波 315832;

3. 国家道路交通安全管理工程技术研究中心宁波大学分中心, 浙江 宁波 315832)

摘要: 鉴于江海联运运量受众多因素的影响, 为了解决江海联运运量预测问题, 先对江海联运运量影响因素进行分析, 再用灰色关联度分析筛选出其中的典型因素. 在此基础上应用粒子群算法优化的支持向量机建立江海联运运量预测模型, 应用于宁波港域江海联运量的预测. 结果表明, 该模型与传统时间序列预测方法相比具有较高的拟合度和预测精度, 为解决江海联运运量等多因素非线性系统预测提供了一条新的途径.

关键词: 江海联运运量; 因素分析; 灰色关联度分析; 粒子群算法; 支持向量机

中图分类号: U6-9

文献标志码: A

文章编号: 1001-5132 (2020) 02-0086-07

江海联运是发展港口和腹地经济的重要支撑, 其作为重要的集疏运方式之一, 能减少中转运输环节, 且能有效增强多种运输方式的深度融合, 具有成本、效率和环保优势. 更因能够打通内河运输和海洋运输两个独立的体系, 而成为连接沿海港口与内陆腹地的重要纽带, 降低了全社会物流成本, 是构建节约型社会, 实现绿色发展的重要运输方式. 对江海联运的运量作科学预测, 是港口航运发展科学决策的重要内容, 对统筹规划、经营开发、建设投资、资源配置等决策和规划有着不可或缺的作用.

目前, 关于江海联运运量预测的研究较少. 姜颖琿^[1]通过回归分析法与指数平滑法, 结合洋山港规划以及长江集装箱运输情况, 对洋山港江海联运运量进行预测, 并结合外部环境变化的各影响因素对预测结果进行调整. 董文娟等^[2]分析港口集装箱吞吐量与“水水中转”集装箱量的相关度, 并根据港口集装箱吞吐量预测值对“水水中转”需求进行预测. 赵亚鹏等^[3]采用趋势外推模型和外贸集装箱生成量关联分析方法对长江港口集装箱吞吐量(长江沿岸省市至洋山港江海联运运量)进行预

测. 但上述方法主要依托时序分析, 需要许多难以精确获取的数据, 受异常数据点的影响大; 且运量非均匀变化时, 预测精度较差, 具有一定局限性.

江海联运运量有许多影响因素, 传统的时间序列法也仅仅考虑了时序变化规律, 不一定适用于江海联运运量的预测. 本文先用灰色关联度分析(Grey Relation Analysis, GRA)确定影响江海联运运量的典型因素, 再用粒子群算法优化(Particle Swarm Optimization, PSO)的支持向量机(Support Vector Machine, SVM)进行预测, 建立基于因素分析和 SVM 模型的江海联运运量预测模型, 并通过实例验证本模型对江海联运运量预测的准确性和实用性, 为江海联运运量预测提供一种新的方法.

1 江海联运运量影响因素分析

1.1 江海联运运量的影响因素

影响江海联运运量的因素十分复杂, 查阅国内外学者关于江海联运影响因素的研究, 提取出影响江海联运运量的因素主要有下面几个部分: (1) 腹地经济因素, 而且, 腹地的集疏运能力、能源消耗水平对江海联运的运量的影响也不容忽视. (2)

收稿日期: 2019-08-09.

宁波大学学报(理工版) 网址: <http://journalg.nbu.edu.cn/>

基金项目: 国家自然科学基金(61074142); 国家重点研发计划政府间国际合作专项(2017YFE9134700); 欧盟地平线 2020 项目(690713); 浙江省自然科学基金(LY15E080013).

第一作者: 杨丁红(1994-), 女, 山东莘县人, 在读硕士研究生, 主要研究方向: 江海联运. E-mail: 952692327@qq.com

*通信作者: 郑彭军(1965-), 男, 浙江宁波人, 教授, 主要研究方向: 交通运输与港航管理. E-mail: zhengpengjun@nbu.edu.cn

港口本身因素, 其中包括港口本身的建港条件, 如港口的地理位置、自身的基础设施条件、港口周围体系竞争状况、管理与综合服务水平^[4]。(3)其他因素, 包括水文航道条件、天气环境、国家政策和国际环境、重大项目等等。

因国家政策、港口基础设施条件、管理与综合服务水平等因素很难量化, 在选择指标时不予考虑, 可以在预测后作为调整依据, 只对影响江海联运运量的部分因素进行分析和筛选。因此, 在分析江海联运的运量预测影响因素时, 只选取有较大影响的因素, 并且可以用一些具体指标来衡量这些因素。本文选取影响江海联运运量的因素和具体量化指标见表 1。

1.2 灰色关联度分析江海联运运量影响因素

选出江海联运运量的影响因素后, 采用灰色关联度分析哪些因素对江海联运运量影响大, 关系更为密切。灰色关联度分析方法是灰色系统分析的一种重要方法, 是对某一发展变化系统的动态过程、发展态势定量描述和比较的一种方法^[5]。通过灰色关联度分析计算得出各影响因素的关联度大小来判断其对江海联运运量的影响程度。关联度越大, 两者关联越大, 反之, 两者关联度越小。

将江海联运运量数据列记为: $X_0(t)=[X_0(1), X_0(2), \dots, X_0(n)]$, n 为数列的长度。江海联运影响因素指标数据列记为 $X_i(t)=[X_i(1), X_i(2), \dots, X_i(n)]$, $i=0, 1, \dots, l$, l 为指标数列。由于各指标原始数据物理意义和计量单位不一定相同, 从而导致量纲不同, 数量级差也悬殊, 对此, 本文用初值化对数据进行处理:

$$X'_i = X_i / X_i(1). \quad (1)$$

求江海联运运量数据列 $X_0(t)$ 与各江海联运运量影响因素数据列 $X_i(t)$ 的绝对差:

$$\Delta_i(K) = |X'_0(k) - X'_i(k)|, \quad (2)$$

然后再求最大绝对差与最小绝对差:

$$\Delta_{\max} = \max_i \max_k |X'_0(k) - X'_i(k)|, \quad (3)$$

$$\Delta_{\min} = \min_i \min_k |X'_0(k) - X'_i(k)|. \quad (4)$$

将上式代入式(5), 计算江海联运运量影响因素的关联系数:

$$\xi_i(k) = (\Delta_{\min} + \rho\Delta_{\max}) / (\Delta_i(k) + \rho\Delta_{\max}), \quad (5)$$

式中, ρ 为分辨系数, 一般取 $\rho=0.5$ 。

计算各个被比较数据列 $X_i(t)$ 对江海联运运量数据列 $X_0(t)$ 的关联度:

$$\rho'_i = \frac{1}{n} \sum_{k=1}^n \xi_i(k), \quad (6)$$

并将计算得到的关联度从小到大进行排序。

1.3 典型影响因素确立

通过分析宁波市的相关数据来选取影响江海联运运量的典型影响因素。宁波市港航局资料显示, 宁波江海联运的腹地城市有南京、扬州、镇江、泰州、南通、江阴、苏州、徐州、无锡、常州、淮安、盐城、安庆、池州、铜陵、芜湖、马鞍山、巢湖、亳州、九江、南昌、武汉、黄石、泸溪、张家界等。按运量大小, 本文主要分析 2004~2017 年南京、芜湖、马鞍山、九江、南通、镇江、苏州 7 个腹地城市的数据和宁波港域港口吞吐量、货船拥有量数据。

利用上述数据, 对 8 项指标数据用 Matlab 进行因素灰色关联度分析, 计算出各因素指标与宁波江海联运运量的灰色关联度(表 2)。

由上可知, 与江海联运运量关联度达到 0.8 以上的指标出现次数较多^[4], 有 X_7 、 X_6 、 X_2 、 X_5 , 分别是港口吞吐量、外贸进出口总额、能源消耗总量、水路集疏运货运量, 说明江海联运运量的典型影响因素为港口吞吐情况、腹地对外经济水平、腹地

表 1 江海联运运量预测影响因素及指标

影响因素	量化指标
腹地经济发展整体水平	地区生产总值(X_1)
腹地能源消耗水平	能源消耗总量(X_2)
腹地集疏运能力	集疏运货运量(公路 X_3 、铁路 X_4 、水路 X_5)
腹地对外经济水平	外贸进出口总额(X_6)
港口吞吐情况	港口吞吐量(X_7)
货运船舶拥有水平	货船拥有量(X_8)

表2 因素指标灰色关联度分析结果

腹地城市	关联度							
	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_6	ρ_7	ρ_8
南京	0.616	0.703	0.687	0.699	0.721	0.698	0.820	0.591
芜湖	0.639	0.533	0.512	0.580	0.661	0.814	0.844	0.633
马鞍山	0.911	0.893	0.843	0.832	0.860	0.941	0.962	0.892
九江	0.882	0.695	0.885	0.937	0.929	0.956	0.959	0.885
南通	0.639	0.655	0.582	0.557	0.560	0.628	0.823	0.596
镇江	0.786	0.829	0.745	0.724	0.813	0.805	0.906	0.756
苏州	0.772	0.820	0.536	0.791	0.553	0.827	0.892	0.726
全部	0.849	0.821	0.897	0.903	0.907	0.922	0.935	0.824

能源消耗水平、腹地集疏运能力。

2 SVM 模型

SVM 模型是 Vapnik 提出的一种新型统计学习方法^[6]。它可以弥补时间序列预测方法对时间序列的随机性与波动性考虑较少、回归预测需要数据量大、要求分布较典型、计算量大、人工神经网络过拟合、收敛速度慢、易陷入局部最优等不足^[7]，被越来越多的学者用来解决小样本、非线性、多变量等问题。支持向量机既可以用于分类，也可以用于回归；当用于分类时，支持向量机是寻找一个最优分类面将样本分开；当用于回归拟合时，是寻找一个最优分类面，并使得所有的训练样本离该最优分类面的误差最小。

假设在用于预测时，有 l 个训练样本的训练样本对为 $\{(x_i, y_i), i=1, 2, \dots, l\}$ ， x_i 是第 i 个训练样本的输入列向量， $y_i \in R$ 为对应的输出值。 $\varphi(x)$ 表示将 x 映射后的特征向量， ω 为权重向量， b 为阈值，于是在高维特征空间中，线性回归函数可表示为：

$$f(x) = \omega \varphi(x) + b. \quad (7)$$

引入 ε 线性不敏感函数，松弛变量后，有最小化函数：

$$\begin{cases} \min_{w, b} \frac{1}{2} \|\omega\|^2 + c \sum_{i=1}^l (\xi_i + \xi_i^*), \\ \text{s.t.} \begin{cases} y_i - \omega \varphi(x_i) - b \leq \varepsilon + \xi_i \\ -y_i + \omega \varphi(x_i) + b \leq \varepsilon + \xi_i^*, \\ \xi_i \geq 0, \xi_i^* \geq 0, i=1, 2, \dots, l, \end{cases} \end{cases} \quad (8)$$

其中， c 为惩罚参数，其值越大，则对训练误差大于 ε 的样本惩罚越大； ε 规定了回归函数的误差要求，其值越小，表示回归函数误差越小； ξ_i 和 ξ_i^* 是

松弛因子，预测存在误差时都大于零，否则等于零。构造 Lagrange 函数，将式(8)转换为对偶形式：

$$\begin{cases} \max_{\alpha, \alpha^*} \left[\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j) - \sum_{i=1}^l (\alpha_i + \alpha_i^*) \varepsilon + \sum_{i=1}^l (\alpha_i - \alpha_i^*) y_i \right], \\ \text{s.t.} \begin{cases} \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0, \\ 0 \leq \alpha_i \leq c, 0 \leq \alpha_i^* \leq c, \end{cases} \end{cases} \quad (9)$$

其中， α_i, α_i^* 是 Lagrange 的乘数。

对式(9)求解得到最终的回归函数：

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i, x) + b^*, \quad (10)$$

式中， $K(x_i, x)$ 为核函数^[8]，有 4 种常用的核函数：线性核函数、多项式核函数、径向基核函数(Radial Basis Function, RBF)和 Sigmoid 函数^[9]。而 RBF 具有广泛的收敛域，因此被广泛使用。

由于精确的 SVM 模型需要确立惩罚因子 c 和核参数 g ，惩罚因子 c 过大或者过小都会影响模型的泛化能力或训练误差，而核参数 g 过大或者过小可能会引发过拟合或者欠学习现象。因此，需使用粒子群算法对支持向量机的参数进行寻优，应用改进后的粒子群算法进行参数最优选择时，粒子自主调节初始状态，提高搜索精度，使支持向量机的预测更精准。

3 江海联运运量预测模型

将所有数据划分为训练数据集(4/5)和测试数据集(1/5)，先对数据进行归一化处理，使得所有数据的取值都在[0,1]范围内，建模之后，再对预测结

果进行反归一化处理.

对粒子群进行初始化. 设置种群规模 50, 最大迭代数 1 000, 加速常数 $c_1 = 1.5$, $c_2 = 1.7$, $wV = 1$, $wP = 1$, $v = 5$.

粒子群算法在更新粒子位置时, 粒子经常超出规定的速度边界^[10]. 当粒子跃过限定的最大速度 $V_{\max}$ 或者最小速度 $V_{\min}$ 时, 引入控系数 k 调节粒子飞行速度.

$$\begin{cases} V_{(c,g)\max} = k \cdot (c, g)_{\max}, \\ V_{(c,g)\min} = -V_{(c,g)\max}, \\ x_0 = ((c, g)_{\max} - (c, g)_{\min}) \cdot R_{\text{rand}} + (c, g)_{\min}, \\ v_0 = V_{(c,g)\max} \cdot R_{\text{rand}}, \end{cases} \quad (11)$$

式中, $V_{(c,g)\max}$ 为参数选择时的最大速度; $V_{(c,g)\min}$ 为参数选择时的最小速度; $(c, g)_{\max}$ 为参数 c 和 g 的最大值; $(c, g)_{\min}$ 为参数 c 和 g 的最小值; x_0 为粒子初始位置; v_0 为粒子初始速度, $k, R_{\text{rand}} \in (0, 1)$. 本文选择 $k = 0.7$, $c_{\max} = 100$, $c_{\min} = 0.1$, $g_{\max} = 1000$, $g_{\min} = 0.01$.

计算适应度值^[11], 其算式如下:

$$f(x) = \sqrt{\left(\sum_{i=1}^N (y_i - y_i^*)^2 \right) / N}, \quad (12)$$

式中, y_i 为 SVM 测试样本实际值; y_i^* 为 SVM 测试

样本预测值; N 为相应样本数.

若粒子当前值优于全局值, 则将全局值作为粒子位置, 更新粒子极值; 若粒子最好的适应值优于当前全局极值, 则将种群最优值作为该粒子的位置, 更新全局值. 对于每个粒子而言, 需将其适应值与其经过的最好位置作比较, 若当前适应值更高, 则将其作为当前最好的位置, 更新 P_i . 对于每个粒子而言, 将其适应值与群体经过的最好位置作比较, 若当前的适应值更高, 则将其作为当前最好的位置, 更新 P_g .

粒子速度和个数更新时, 可按如下公式来调整其速度与位置^[12]:

$$v_i^{t+1} = \omega v_i^t + c_1 r_1 (p_i^t - x_i^t) + c_2 r_2 (p_g^t - x_i^t), \quad (13)$$

$$x_i^{t+1} = x_i^t + v_i^{t+1}, \quad (14)$$

式中, t 为迭代次数 ($t \geq 0$); ω 为惯性权重; r_1 和 r_2 为 2 个在 $[0, 1]$ 范围内变化的随机值; p_i 为粒子个体历史最优位置; p_g 为群体历史最优位置; x_i^t 和 v_i^t 分别是第 i 个粒子在 t 次迭代的位置和速度. 若寻优达到最大迭代次数或者适应度值满足要求时, 则寻优结束, 否则继续进行速度和位置的更新. 直至寻优结束后, 输出最优解, 将其赋予给 SVM.

应用训练数据集对 SVM 进行训练, 其中核函数使用径向基核函数(RBF), 用 SMO(Sequential

表 3 各腹地 2004~2017 年经济因素基本数据

腹地城市	年份	地区生产总值 X_1 /(亿元)	能源消耗总量 X_2 /($\times 10^4$ t)	公路集疏运货 运量 X_3 /($\times 10^4$ t)	铁路集疏运货 运量 X_4 /($\times 10^4$ t)	水路集疏运货 运量 X_5 /($\times 10^4$ t)	外贸进出口总 额 X_6 /(亿美元)
南京	2004	1 576.33	1 430.34	9 741	986.8	6 206	1 430.34
	2005	2 411.11	1 460.54	10 530	1 056	6 483	1 460.54
	2006	2 773.78	1 580.06	11 249	1 096	6 042	1 580.06
	2007	3 283.73	1 553.22	12 686	1 080	6 077	1 553.22
	2008	3 814.62	1 466.94	13 650	964.2	9 485	1 466.94
	2009	4 230.26	1 460.71	12 686	1 080	6 077	1 460.71
	2010	5 012.64	2 213.67	14 983	1 450	9 561	2 213.67
	2011	6 145.52	2 777.04	17 683	1 593.55	11 292	2 777.04
	2012	7 201.57	2 861.32	19 820	1 820.49	14 090	2 861.32
	2013	8 011.78	3 517.22	22 020	1 805.83	15 090	3 517.22
	2014	8 820.75	3 623.73	23 738	1 701.55	15 556	3 623.73
	2015	9 720.77	3 670.95	12 365	1 266.45	13 322	3 670.95
	2016	10 502.02	3 822.21	13 341	1 299.38	13 812	3 822.21
	2017	11 715.10	3 784.38	14 974	1 393	14 840	3 784.38
	...	...	...	...	...	...	...

续表

腹地城市	年份	地区生产总值 X_1 /(亿元)	能源消耗总量 X_2 /($\times 10^4$ t)	公路集疏运货运 量 X_3 /($\times 10^4$ t)	铁路集疏运货 运量 X_4 /($\times 10^4$ t)	水路集疏运货 运量 X_5 /($\times 10^4$ t)	外贸进出口总 额 X_6 /(亿美元)
苏州	2004	3 450	752.12	6 941	968.5	1 892	1 032.01
	2005	4 138.21	635.46	7 762	887.19	2 631	1 405.89
	2006	4 900.63	426.33	8 891	791.51	2 208	1 742.64
	2007	5 850.11	472.71	9 958	763.65	1 876	2 117.96
	2008	7 078.09	492.59	10 863	622.11	1 768	2 285.26
	2009	7 740.2	622.99	10 820	571.06	550	2 014.46
	2010	9 228.91	671.87	12 768	491	649	2 740.76
	2011	10 716.99	699.84	14 798	499.34	804	3 008.63
	2012	12 011.65	771.21	16 441	405.21	884	3 056.92
	2013	13 015.7	1 137.61	10 799	279.32	987	3 093.48
	2014	13 760.89	1 110.35	11 855	221.84	1 098	3 113.06
	2015	14 504.07	1 087.09	11 814	178.4	1 191	3 053.50
	2016	15 475.09	1 135.35	12 287	185.97	1 221	2 737.58
	2017	17 319.51	1 109.15	13 614	188.5	1 323	3 160.79

注: (1)数据来源于各城市统计年鉴和统计公报; (2)能源消耗总量按标准煤计算。

Minimal Optimization)^[13]求解得到拉格朗日乘子 α_i 、 α_i^* 和阈值 b 。将参数代入式(10)中得到最终的回归函数式, 用训练完成的 SVM 进行江海联运运量预测。

4 实例分析

通过查阅历年统计年鉴及相关统计资料, 收集了 2004~2017 年共 14 年的统计数据, 鉴于数据过于繁多, 因此文中只列举部分(表 3 和表 4)。

按照腹地城市占宁波港域江海联运运量的比例作为权重, 相加得到全体腹地数据。将 2004~2015 年所确定的江海联运运量典型因素数据和江海联运运量作为训练样本, 2016~2017 年的数据作为测试样本检验所构建模型的预测精度, 用来预测宁波港域 2018~2022 年江海联运运量。

4.1 预测模型精度检验

为检验所提出模型的性能, 采用平均相对误差和均方根误差来对预测模型的测试结果进行精度检验, 相关计算公式如下。

平均相对误差:

$$MARE = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - y_i^*|}{y_i} \quad (15)$$

均方根误差:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - y_i^*)^2} \quad (16)$$

表 4 宁波港域 2004~2017 年基本数据

年份	港口吞吐量 X_7 /($\times 10^4$ t)	货船拥有量 X_8 /艘	江海联运运 量/($\times 10^4$ t)
2004	22 586	554	3 114.98
2005	26 881	641	3 332.42
2006	30 969	634	3 957.23
2007	34 519	628	3 485.34
2008	36 185	663	3 808.10
2009	38 385	658	4 281.66
2010	41 217	658	4 159.93
2011	43 339	657	4 213.12
2012	45 303	630	5 081.15
2013	49 592	627	5 679.16
2014	52 646	641	7 030.70
2015	51 004	563	6 334.06
2016	49 618	574	5 675.91
2017	55 151	581	5 933.23

注: 数据来源于宁波统计局和宁波港航管理局。

上式中, y_i 为 SVM 测试样本实际值; y_i^* 为 SVM 测试样本预测值; N 是相应样本数。

4.2 对比模型

为了检验本文 PSO-SVM 模型精度, 同时使用

ARIMA(Autoregressive Integrated Moving Average)模型^[14]和SVM模型对宁波港域江海联运运量进行了预测,并对它们的预测结果进行对比分析。

5 结果分析

对2016~2017年宁波港域江海联运运量预测,结果见表5。

表5 2016~2017年宁波港域江海联运运量 $\times 10^4$ t

年份	实际值	PSO-SVM 预测值	SVM 预测值	ARIMA 预测值
2016	5 675.9	5 756.9	5 855.2	6 585.6
2017	5 933.2	6 115.1	6 252.3	6 859.8

将各模型的预测值和实际值代入式(15)、(16),可得各模型预测值的平均相对误差和均方根误差值,结果如图1所示。ARIMA模型均方根误差为 918.2×10^4 t,平均相对误差为0.158;SVM模型均方根误差为 258.8×10^4 t,平均相对误差为0.043;PSO-SVM模型均方根误差为 140.8×10^4 t,平均相对误差为0.022。由表5、图1可得,无论是均方根误差还是平均相对误差,PSO-SVM的预测结果优于ARIMA、SVM。ARIMA预测误差最大,SVM预测结果优于ARIMA,而PSO-SVM预测误差最小,预测精度最高。

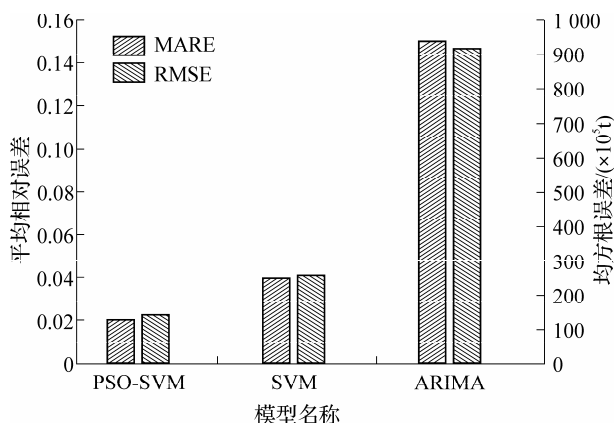


图1 不同模型的MARE和RMSE精度比较

由此可见,在样本数据较少,数据波动较大的情况下,仅依靠时序单一因素分析的ARIMA模型的预测精度较差,而建立在多种影响因素分析基础上的SVM模型可以在有限样本情况下获得较优解,具有较好的泛化能力。而PSO-SVM利用粒子群算法优化,在支持向量机模型基础上进一步提

高了精度,取得了满意的结果。

利用本文的PSO-SVM模型,进一步对宁波港域2004~2017年及2018~2022年江海联运运量进行了预测,结果如图2所示。由图可见,2018~2022年宁波港域江海联运运量有波动,但总体呈现平稳增长的趋势。

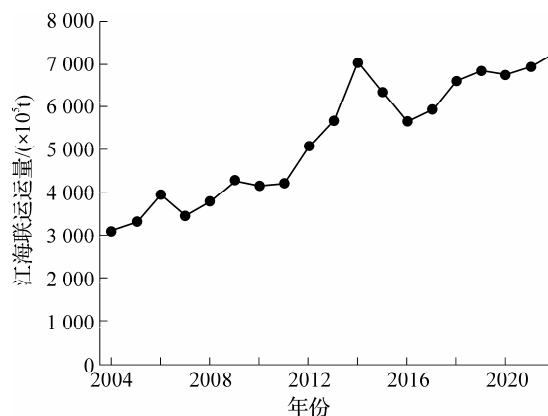


图2 2004~2017年及2018~2022年宁波港域江海联运运量预测

6 结论

本文提出一种基于因素分析和SVM模型的江海联运运量预测方法,使用灰色关联度分析江海联运运量的影响因素,提取出典型影响因素,再利用构建的粒子群算法优化的支持向量机来进行江海联运运量的预测,通过实例分析,验证了所提出的基于因素分析和SVM模型的江海联运运量预测方法预测精度优于ARIMA模型和传统SVM模型,可较为精确地预测未来江海联运运量的增长趋势,能为江海联运运量预测及港口航运的发展和规划提供一定的参考。

参考文献:

- [1] 姜颖琿. 外部环境变化对洋山港江海联运运量的影响[D]. 上海: 上海海事大学, 2005.
- [2] 董文娟, 徐莹. 宁波港“水水中转”发展策略研究[J]. 浙江万里学院学报, 2015, 28(5):13-17.
- [3] 赵亚鹏, 丁以中. 基于江海联运的长江沿线港口集装箱吞吐量分析预测[J]. 水运工程, 2006(11):1-3.
- [4] 刘枚莲, 朱美华, 黄键. 港口吞吐量预测影响因素筛选方法研究[J]. 水运工程, 2011(3):76-80.
- [5] 刘文文. 基于灰色关联铁水含硅量影响因素分析[J]. 中外企业家, 2013(27):228-229.
- [6] Vapnik V N. The Nature of Statistical Learning Theory:

- Statistics for Engineering and Information Science[M]. New York: Springer-Verlag, 2000.
- [7] 戴焱, 王锡淮, 肖健梅. 支持向量机在集装箱吞吐量预测上的应用[J]. 水运工程, 2005(8):18-21; 39.
- [8] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.
- [9] Elbisy M S. Sea wave parameters prediction by support vector machine using a genetic algorithm[J]. Journal of Coastal Research, 2015, 31(4):892-899.
- [10] 汪友明, 穆恒星, 徐国宁. 基于改进粒子群优化支持向量的光伏电池输出功率预测[J]. 电气自动化, 2019, 41(3):63-65; 91.
- [11] 杨小明, 崔雪, 周斌, 等. 基于粒子群优化支持向量机的短期负荷预测[J]. 武汉大学学报(工学版), 2018, 51(8):715-720.
- [12] Subasi A. Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders [J]. Computers in Biology and Medicine, 2013, 43(5): 576-586.
- [13] 白鹏. 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.
- [14] 吴琛. 基于时间序列模型的港口集装箱吞吐量预测[J]. 珠江水运, 2019(5):73-74.

Method for prediction of river-sea combined transport volume based on factor analysis and SVM model

YANG Dinghong^{1,2,3}, ZHENG Pengjun^{1,2,3*}

(1.Faculty of Maritime and Transportation, Ningbo University, Ningbo 315832, China;

2.Collaborative Innovation Center for Ningbo Port and Logistics Service System, Ningbo 315832, China;

3.Ningbo University Sub-centre, National Traffic Management Engineering & Technology Research Center, Ningbo 315832, China)

Abstract: The volume of river-sea combined transport is affected by many factors. In order to solve the problems found in forecasting this volume, its influencing factors are analyzed first, and then typical factors are selected by grey relation analysis. Based on the result, a support vector machine (SVM), optimized by particle swarm optimization (PSO) algorithm, is applied to establish a predictive model of the river-sea combined transport volume. The model is then applied to the prediction of the river-sea combined transport volume in Ningbo Port. The results show that the model is superior to the traditional time series forecasting method, and has a high degree of fitness and forecasting accuracy. It provides a new way to solve the multi-factor nonlinear system prediction problems, such as the one in river-sea combined transport volume prediction.

Key words: river-sea combined transport; factor analysis; grey relation analysis; particle swarm optimization; support vector machine

(责任编辑 章践立)