

基于二阶振荡微粒群最小二乘支持向量机的 物流需求预测*

耿立艳¹, 赵鹏², 张占福³

(1. 石家庄铁道大学经济管理学院, 石家庄 050043; 2. 河北科技师范学院欧美学院, 河北秦皇岛 066004;
3. 石家庄铁道大学四方学院, 石家庄 051132)

摘要: 为了提高物流需求的预测精度,在分析物流需求影响因素基础上,建立了物流需求的二阶振荡微粒群最小二乘支持向量机预测模型。利用最小二乘支持向量机(LSSVM)描述物流需求与其影响因素间的复杂非线性关系,并通过二阶振荡微粒群(TOOPSO)算法优化选择LSSVM参数。实例分析表明,模型具有较高的预测精度,TOOPSO算法搜索LSSVM最优参数时间明显少于传统交叉验证法,是一种有效的物流需求预测方法。

关键词: 物流需求预测; 最小二乘支持向量机; 二阶振荡微粒群算法

中图分类号: TP391 **文献标志码:** A **文章编号:** 1001-3695(2012)07-2558-03

doi:10.3969/j.issn.1001-3695.2012.07.042

Logistics demand forecasting based on LSSVM optimized by two-order oscillating PSO

GENG Li-yan¹, ZHAO Peng², ZHANG Zhan-fu³

(1. School of Economics & Management, Shijiazhuang Tiedao University, Shijiazhuang 050043, China; 2. E&A College, Hebei Normal University of Science & Technology, Qinhuangdao Hebei 066004, China; 3. Shijiazhuang Tiedao University Sifang College, Shijiazhuang 051132, China)

Abstract: Based on analyzing the factors of logistics demand, this paper proposed a new model named the two-order oscillating particle swarm least squares support vector machines (TOOPSO-LSSVM) model to improve the forecasting accuracy of logistics demand. The complex nonlinear relationship between logistics demand and its impact factors were explained through LSSVM. And then, it used TOOPSO algorithm to optimize the parameters of LSSVM model. An empirical analysis indicates that the forecasting performance of LSSVM is better than the other three models and the searching time for optimal parameters of LSSVM by TOOPSO is obviously less than cross validation method, which is an effective method for logistics demand forecasting.

Key words: logistics demand forecasting; LSSVM; two-order oscillating particle swarm optimization algorithm

物流需求预测在国家及区域物流发展规划中具有重要意义,对物流需求的准确预测不仅为各级政府制定宏观物流发展战略规划、确定物流基础设施建设规模提供决策依据,而且为物流企业合理配备物流设备、制定物流管理方案提供理论与方法指导。物流系统是一个庞大而复杂的开放系统,这就决定了物流需求预测的复杂性。物流需求受到如经济、政策、资源等众多因素的影响,这些因素有些可定量描述、有些无法定量描述,各影响因素的作用形式复杂且随时间的变化发生改变,致使物流需求本身具有较大的波动性和随机性特点,而且物流需求与其影响因素之间呈现高度非线性关系,难以建立确定的数学模型。此外,我国物流统计制度和物流统计体系还有待完善,有关物流业发展水平的详细统计数据相对缺乏,实际可利用的数据较少,这也增加了物流需求预测的难度。

传统的预测方法,如回归分析预测法、时间序列分析法通常以假设条件为前提建立模型,很难全面揭示数据内在的复杂性。神经网络作为一种人工智能方法,能够较好地刻画物流

需求与其影响因素间的非线性映射关系,通过网络本身的自学习、自适应能力能够实现较高精度的预测,因而被广泛应用于物流需求预测中^[1,2]。但神经网络所需数据量较大,在实际问题中经常遇到过学习、陷入局部极值等问题。支持向量机(SVM)是一种基于统计学习理论的机器学习算法,通过采用结构风险最小化原则克服了神经网络的局限性,适用于小样本、非线性、高维数、局部极值等实际问题^[3],在现有物流系统数据样本较少情况下,可以较好地描述物流需求与影响因素间的非线性关系,在物流需求预测中取得了较好的研究效果^[4]。最小二乘支持向量机(LSSVM)是对标准支持向量机的扩展^[5],将标准SVM中的线性不等式约束转换为线性等式约束,提高了求解效率,更适用于物流需求这种复杂系统的预测研究。LSSVM预测精度主要依赖于核函数的选择及参数的设定,传统的交叉验证法易于实现,但本身的局限性将限制LSSVM的推广应用。二阶振荡微粒群算法(TOOPSO)是在PSO算法^[6]基础上发展起来的优化算法,通过引入二阶振荡进化

收稿日期: 2011-12-08; **修回日期:** 2012-01-16 **基金项目:** 河北省社会科学基金资助项目(HB12YJ035); 国家软科学研究计划资助项目(2010GXQ5D320); 教育部人文社会科学研究青年基金资助项目(11YJC790048)

作者简介: 耿立艳(1979-),女,天津人,硕士,博士,主要研究方向为智能预测方法及应用(gengliyan_28117@yahoo.com.cn); 赵鹏(1977-),男,黑龙江齐齐哈尔人,助教,硕士,主要研究方向为物流工程; 张占福(1978-),男,天津人,助理工程师,主要研究方向为智能优化与控制。

方程增加了微粒的多样性,改善了算法的性能^[7]。

1 物流需求影响因素分析

物流系统是整个社会经济系统的一个子系统,物流需求受到诸如经济、社会、环境等多种因素的影响。综合现有研究成果,认为中国物流需求的影响因素有以下几方面:

a) 国民经济和产业结构。作为重要的经济活动,物流需求与国民经济具有密切的正相关关系,随着国民经济的增长,物流需求不断增长。产业结构的变动也将对物流需求产生重要影响,产业结构的调整与升级可以促进经济的快速增长,增加对原材料、成品及半成品的需求,带动物流需求的快速增长。三大产业产值之和构成国内生产总值(GDP),因此选取三大产业产值作为物流需求预测指标,可以同时兼顾国民经济和产业结构对物流需求的影响。

b) 固定资产投资总额。对固定资产进行投资可以加快经济的发展速度,进而加快物流需求的增长,特别是对铁路、公路和港口等基础设施的投资直接可以促进物流业的发展,增加对物流的需求。

c) 货物进出口总额。随着我国进出口贸易的增加,货物进出口总额对物流需求的影响越来越大,进出口贸易的发展促进了社会商品的流通,对物流需求起到推动作用。

d) 社会消费品零售总额。社会消费需求对物流需求有较大影响。在市场经济环境下,社会消费需求在拉动我国经济增长的同时,也带来了巨大的物流需求。

e) 居民消费水平。居民消费水平的变动可以带动物流产业结构的变动和升级,进而对物流需求产生影响。居民消费水平提高会增加对物流的需求;消费水平下降会抑制对物流的需求。另外,国家经济政策的倾向性也会影响到物流需求,经济政策的支持会加快国家经济发展,物流需求相应增加;反之物流需求相应减少。科学技术的发展,可以提升物流的服务质量和范围,从而增加对物流的需求,但国家经济政策和科学技术这两项指标难以量化。

综上所述,本着可操作性、可量化原则,选取第一产业产值、第二产业产值、第三产业产值、固定资产投资总额、货物进出口总额、社会消费品零售总额和居民消费水平七项因素作为影响物流需求的因素。

2 TOOPSO-LSSVM 预测模型

2.1 物流需求的 LSSVM 预测模型

设 $L = \{(x_k, y_k) | k = 1, 2, \dots, N\}$ 为物流需求的训练样本集,其中, $x_k \in \mathbb{R}^d$ 为输入变量, $y_k \in \mathbb{R}$ 为输出变量。本文中输入变量 $x = (x_1, x_2, x_3, x_4, x_5, x_6, x_7)$ 为七项影响因素组成的向量,输出变量为物流需求量 y 。LSSVM 通过某一非线性映射函数 $\varphi(x) = (\varphi(x_1), \varphi(x_2), \dots, \varphi(x_N))$ 实现输入变量到高维特征空间的映射,并在该空间中构造线性回归函数。根据结构风险最小化原则, LSSVM 回归算法的优化问题变为

$$\min_{\omega, b, e} J(\omega, b, e_k) = \frac{1}{2} \omega^T \omega + \frac{1}{2} \gamma \sum_{k=1}^N e_k^2$$

$$\text{s. t. } y_k(x) = \omega^T \varphi(x_k) + b + e_k (k=1, 2, \dots, N) \quad (1)$$

其中: ω 和 b 分别为权向量和阈值; γ 为正则化参数; e 为误差。通过建立 Lagrange 函数, LSSVM 将优化问题转换为求解线性方程组:

$$\begin{bmatrix} 0 & 1_N^T \\ 1_N & K + I_N/\gamma \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \quad (2)$$

其中: I_N 为单位矩阵; $y = [y_1, \dots, y_{N-1}]^T$, $\alpha = [\alpha_1, \dots, \alpha_{N-1}]^T$ 为拉格朗日乘子; $K(x_k, x_l) = \varphi(x_k)^T \varphi(x_l)$ ($l = 1, 2, \dots, N$) 为满足 Mercer 条件的核函数。由上式解出 α 和 b 可以得到物流需求的 LSSVM 预测模型为

$$y(x) = \sum_{k=1}^N \alpha_k K(x, x_k) + b \quad (3)$$

可选择的核函数主要有线性函数、多项式函数、径向基函数、Sigmoid 函数等。本文选取泛化能力较好的径向基函数为核函数:

$$K(x, x_k) = \exp(-\|x - x_k\|^2/\sigma^2) \quad (4)$$

其中: σ^2 为核函数的宽度。

2.2 模型参数优化

若选取径向基函数为核函数, LSSVM 需确定的参数为 (γ, σ^2) , 这两个参数对 LSSVM 的预测精度有重要影响。通常采用交叉验证法确定 (γ, σ^2) 的值,但交叉验证法存在人为选择的盲目性、计算工作量大、收敛速度慢等问题,故很难得到最优解。因此,利用交叉验证法选择 LSSVM 参数将在一定程度上影响物流需求预测的精度。

微粒群(PSO)算法是一种随机优化算法,通过微粒的自身信息和全局信息,动态地改变微粒的速度和位置搜索最优解。二阶振荡微粒群(TOOPSO)算法在进化方程中引入振荡环节,通过利用前一代微粒信息,增强算法的多样性和全局收敛性,本文采用 TOOPSO 算法选择 LSSVM 的最优参数 (γ, σ^2) , 这相当于将 LSSVM 预测模型的构造嵌入 TOOPSO 算法的寻优过程。TOOPSO-LSSVM 预测模型算法步骤具体如下:

a) 初始化微粒群。设置微粒的群体规模 m , 加速系数 c_1 和 c_2 , 最大、最小惯性权重 w_{\max} 和 w_{\min} , 最大迭代次数 t_{\max} 。随机产生一组 (γ, σ^2) 作为微粒的初始位置和速度,并将每个微粒的当前位置设为个体最优位置。

b) 定义适应度函数。LSSVM 的实质是在经验误差与泛化能力间找到最佳平衡点,使该点的结构误差最小化。为了有效平衡 LSSVM 的经验误差与泛化能力,提高预测精度,将数据样本集(共 N 组)分成两部分:前 m 组样本用于训练模型,后 n 组样本用于检验模型的预测能力。适应度函数定义如下:

$$F = \left\{ \frac{1}{m} \sum_{i=1}^m (y_{m,i} - f(x_{m,i}))^2 + \frac{1}{n} \sum_{j=1}^n (y_{n,j} - f(x_{n,j}))^2 \right\} \quad (5)$$

其中:第一部分与第二部分分别为训练误差与检验误差; $y_{m,i}$ 与 $f(x_{m,i})$ 分别为训练样本的实际值与预测值; $y_{n,j}$ 和 $f(x_{n,j})$ 分别为检验样本的实际值和预测值; $m + n = N$, 通常取 $m = (3/4)N$, 通过上式即可平衡 LSSVM 的经验误差与泛化能力。

c) 更新微粒最优位置。按式(5)计算各微粒的适应度值,根据适应度值确定初始全局最优位置。每个微粒按下式更新自己的速度和位置:

$$\begin{cases} V_i(t+1) = wV_i(t) + c_1r_1(P_i - (1 + \xi_1)X_i(t) + \xi_1 X_i(t-1)) + c_2r_2(P_g - (1 + \xi_2)X_i(t) + \xi_2 X_i(t-1)) \\ X_i(t+1) = X_i(t) + V_i(t+1) \end{cases} \quad (6)$$

其中: t 为迭代次数; V_i 为微粒 i 的速度; X_i 为微粒 i 的位置, $i = 1, 2, \dots, m$; P_i 和 P_g 分别为微粒的个体最优位置与全局最优位置; r_1 和 r_2 为 $[0, 1]$ 之间的随机数; ξ_1, ξ_2 为振荡参数,若 $t < t_{\max}$, ξ_1, ξ_2 分别取为

$$\xi_1 < \frac{2\sqrt{c_1 r_1} - 1}{c_1 r_1}, \xi_2 < \frac{2\sqrt{c_2 r_2} - 1}{c_2 r_2} \quad (7)$$

若 $t \geq t_{max}$, ξ_1, ξ_2 分别取为

$$\xi_1 \geq \frac{2\sqrt{c_1 r_1} - 1}{c_1 r_1}, \xi_2 \geq \frac{2\sqrt{c_2 r_2} - 1}{c_2 r_2} \quad (8)$$

更新微粒的个体最优位置 P_i 为 $\min(f(X_i), f(P_i))$ 所对应位置,更新微粒的全局最优位置 P_g 为 $\min(f(P_i), f(P_g))$ 所对应位置,惯性权重 w 按下式自动更新^[8]:

$$w = w_{max} - t(w_{max} - w_{min})/t_{max} \quad (9)$$

若所有微粒的迭代次数满足预先设定值,则停止计算,此时全局最优位置 P_g 即为 LSSVM 最优参数 (γ^*, σ^{2*}) ; 否则转步骤 b)。

d) 建立 LSSVM 预测模型。利用最优参数 (γ^*, σ^{2*}) 建立物流需求的 LSSVM 预测模型,并计算出物流需求的预测值。

3 实例分析

选取货运量作为物流需求的量化指标,利用本文模型对我国 1979—2010 年物流需求进行实例分析,所需数据取自《中国统计年鉴 2000》和《中国统计年鉴 2010》相关数据。将归一化到 $[0, 1]$ 区间的样本分成两组:1979—2004 年数据用于建立及训练模型,2005—2010 年数据用于检验模型的预测性能。TOOPSO 算法自身控制参数设置如下:a) 群体规模 m 设为 10; b) 最大迭代次数 t_{max} 设为 30; c) 加速系数设为 $c_1 = 0.2, c_2 = 1.8$; d) 最大、最小惯性权重分别取为 $w_{max} = 0.9, w_{min} = 0.1$ 。TOOPSO 算法优化选择 LSSVM 参数时,为减少随机性对参数选择的影响,选择仿真 10 次中的最优参数建立 LSSVM 模型,并向前一步预测 2005—2010 年货运量。

为比较 TOOPSO 算法优化参数的能力,同时利用线性递减微粒群(LDWPSO)算法、标准微粒群(PSO)算法和 5 折交叉验证法确定 LSSVM 参数,分别记为 LDWPSO-LSSVM、PSO-LSSVM 和 CV-LSSVM。其中,LDWPSO 算法的加速系数设为 $c_1 = c_2 = 2$, PSO 算法的惯性权重设为 $w = 0.5$ 。四种模型预测结果比较如表 1 和图 1 所示。

由表 1 可知,本文提出的 TOOPSO-LSSVM 模型较准确地预测了全国货运量,最大相对预测误差仅为 2.04%,明显小于 LDWPSO-LSSVM、PSO-LSSVM 和 CV-LSSVM 模型的最大相对预测误差(2.79%、-3.54% 和 -6.26%)。同时图 1 也显示 TOOPSO-LSSVM 较好地预测了货运量的变动趋势,预测值较其他模型更接近于实际值。

表 1 四种模型预测值与实际值比较

年份	货运量 /万吨	TOOPSO-LSSVM		LDWPSO-LSSVM		PSO-LSSVM		CV-LSSVM	
		预测值	相对误差	预测值	相对误差	预测值	相对误差	预测值	相对误差
2005	1862066	1876868	0.79%	1890746	1.54%	1847478	-0.78%	1850628	-0.61%
2006	2037060	2056617	0.96%	2070104	1.62%	2023266	-0.68%	1995001	-2.06%
2007	2275822	2322335	2.04%	2339346	2.79%	2280146	0.19%	2242107	-1.48%
2008	2585937	2607675	0.84%	2630853	1.74%	2554204	-1.23%	2524545	-2.37%
2009	2825222	2806911	-0.65%	2834793	0.34%	2745058	-2.84%	2712712	-3.98%
2010	3241807	3198075	-1.35%	3225458	-0.50%	3127200	-3.54%	3038947	-6.26%

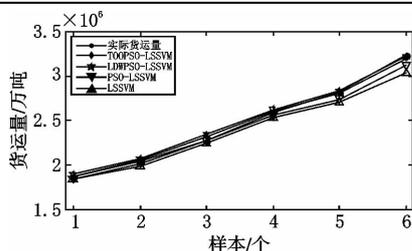


图 1 四种模型预测结果比较曲线

为验证本文模型的有效性,利用均方根误差(RMSE)、平均绝对误差(MAE)、平均相对误差(MPE)、西尔统计量(THEIL)和搜索参数花费时间(TIME)五项指标评价四种模型的预测性能。以上指标值越小,模型预测性能越好,结果如表 2 所示。

表 2 四种模型货运量预测性能比较

评价指标	TOOPSO	LDWPSO	PSO	CV
	-LSSVM	-LSSVM	-LSSVM	-LSSVM
RMSE(10^4)	3.023 7	3.725 2	5.914 6	10.051 2
MAE(10^4)	2.744 0	3.268 1	4.320 2	7.732 9
MPE/%	1.106 1	1.422 3	1.541 7	2.795 7
THEIL/%	0.600 6	0.736 9	1.186 8	2.032 7
TIME/s	5.203 1	5.187 5	4.468 8	107.468 8

通过比较表 2 的 RMSE、MAE、MPE 和 THEIL 值可知, TOOPSO-LSSVM、LDWPSO-LSSVM 和 PSO-LSSVM 模型的预测性能明显优于 CV-LSSVM 模型。其中,TOOPSO-LSSVM 模型在四种模型中的预测精度最高,主要原因是 TOOPSO 算法能够通过振荡参数动态地调整搜索力度,增强算法的全局最优性,进而提高 LSSVM 的预测精度。交叉验证法需经过多次试凑确定参数取值,其随机性和盲目性限制了 LSSVM 精度的提高。从 TIME 看出,TOOPSO 算法搜索 LSSVM 最优参数的时间比较短,仅为 5 s,而 CV-LSSVM 模型的 5 折交叉验证法由于计算工作量较大致使搜索时间增加到 107 s。因此,通过 TOOPSO 算法选择参数能够有效提高 LSSVM 预测精度,加快模型的收敛速度。

4 结束语

针对物流需求与其影响因素间的复杂关系,将二阶振荡微粒群算法与最小二乘支持向量机相结合,提出二阶振荡微粒群最小二乘支持向量机(TOOPSO-LSSVM)模型。该模型既能发挥 LSSVM 在预测领域的优点,又结合了 TOOPSO 算法全局最优性的特点。通过对我国货运量的实例分析证明,与其他算法相比,利用 TOOPSO 算法选择 LSSVM 参数有效提高了模型的预测精度,具有一定应用价值。

参考文献:

- [1] 尹艳玲. 基于自适应神经网络的物流需求预测研究[J]. 河南理工大学学报:自然科学版,2010,29(5): 700-704.
- [2] 耿勇,鞠颂东,陈娅娜. 基于 BP 神经网络的物流需求分析与预测[J]. 物流技术,2007,26(7): 35-37.
- [3] VAPNIK V N. An overview of statistical learning theory[J]. IEEE Trans on Neural Networks,1999,10(5): 988-999.
- [4] 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.
- [5] SUYKENS J T, Van GESTEL T. Least squares support vector machines[M]. Singapore: Singapore Word Scientific,2002: 13-15.
- [6] 师彪,李郁侠,于新花,等. 基于改进粒子群—模糊神经网络的短期电力负荷预测[J]. 系统工程理论与实践,2010,30(1): 157-166.
- [7] 胡建秀,曾建潮. 二阶微粒群算法[J]. 计算机研究与发展,2007,44(11): 1825-1831.
- [8] 黄少荣. 粒子群优化算法综述[J]. 计算机工程与设计,2009,30(8): 1977-1980.