

基于支持向量回归机的区域 物流需求预测模型及其应用*

黄 虎^{1a}, 蒋葛夫^{1a}, 严余松², 廖百胜³, 夏国恩^{1b}

(1. 西南交通大学 a. 物流学院; b. 经济管理学院, 成都 610031; 2. 四川师范大学, 成都 610068; 3. 西南科技大学 土木工程与建筑学院, 四川 绵阳 621010)

摘 要: 为了提高区域物流需求预测的能力,从区域经济等影响因素指标与区域物流需求之间的内在关系的角度,应用基于结构风险最小化准则的支持向量回归机(SVR) 方法,建立“影响因素—区域物流需求”SVR 预测模型来研究预测区域物流需求问题。在选择适当的参数和核函数的基础上,对上海市物流需求量进行预测,发现该方法能获得较小的训练相对误差和测试相对误差。

关键词: 区域物流需求; 支持向量回归机; 预测

中图分类号: U492. 313; TP391 文献标志码: A 文章编号: 1001-3695(2008) 09-2738-03

Model of regional logistics demand forecasts based on support vector regression and its application

HUANG Hu^{1a}, JIANG Ge-fu^{1a}, YAN Yu-song², LIAO Bai-sheng³, XIA Guo-en^{1b}

(1. a. College of Logistics, b. College of Economics & Management, Southwest Jiaotong University, Chengdu 610031, China; 2. Sichuan normal University, Chengdu 610068, China; 3. College of Civil Engineering & Construction, Southwest Science & Technology University, Mianyang Sichuan 621010, China)

Abstract: To improve the forecast ability of regional logistics demand, from the aspect of the intrinsic relations, between the factors of the impact (regional economic and so on) and regional logistics demand, support vector regression (SVR) based on structural risk minimization is applied, by establishing the factors of impact- regional logistics demand SVR forecast model to forecasting regional logistics demand. By selecting appropriate parameters and kernel function, the proposed approach is used for forecasting logistics demand of Shanghai city, experimental results show that the above method obtained less training relative error and testing relative error is lesser.

Key words: regional logistics demand; support vector regression(SVR) ; forecast

物流需求是社会经济生活在物质资料(生产、流通、消费领域的原材料、成品和半成品、商品以及废旧物品、废旧材料) 空间位移、时间效用等物流方面的有支付能力的需要, 涉及运输、库存、包装、装卸搬运、流通加工以及与之相关的信息需求等物流活动的诸方面。区域物流需求预测是区域物流规划、物流资源合理配置过程中的重要环节, 同时它也为政府制定物流产业发展政策、物流基础设施建设提供了必要的理论依据。国内外学者建立的物流需求预测模型大部分利用物流需求历史数据本身来进行物流需求预测^[1~6]。而利用区域物流影响因素来预测区域物流需求的研究目前还不多, 学者陈黎于 2006 年在对物流需求变化的影响因素分析以及对物流需求指标选取的基础上, 采用回归模型、灰色预测方法、加权组合模型对湖北物流需求进行预测^[7]; 初良勇等人通过对物流需求影响因素的分析, 首先建立了回归分析、灰色系统及神经网络方法的物流需求单项预测模型, 然后建立了物流需求组合预测模型^[8]; 后锐等人建立了“区域经济—物流需求”神经网络预测

模型^[9]。但是, 这些传统的物流需求预测方法(如线性回归分析、灰色 GM(1, 1) 模型、组合预测法等) 在处理大规模、高维度、含有非线性关系、非正态分布的影响因素指标数据时, 其效果不理想, 且其指标的主观随意性较大, 而且需要事先知道各种参数, 以及参数在什么情况下应如何修正。人工神经网络主要依靠的是经验风险最小化原则, 容易导致泛化能力的下降且模型结构难以确定。在学习样本数量有限时, 学习过程误差易收敛于局部极小点, 学习精度难以保证; 学习样本数量很多时, 又陷入“维数灾难”, 泛化性能不高。而支持向量回归机^[10]则很好地解决了上述方法面临的问题, 并在实际应用中有很好的性能。本文通过建立“影响因素—区域物流需求”SVR 预测模型, 采用区域经济等影响因素指标来预测区域物流需求。

1 支持向量回归机

考虑 l 个独立同分布的学习样本 $(x_i, y_i), x_i \in R^d, y_i \in R$,

收稿日期: 2007-08-26; 修回日期: 2007-11-20 基金项目: 国家自然科学基金资助项目(60674007)

作者简介: 黄虎(1976-), 男, 四川资中人, 博士研究生, 主要研究方向为物流工程(hh1115@ 126. com); 蒋葛夫(1954-), 男, 四川南充人, 教授, 博导, 博士, 主要研究方向为优化理论及应用、交通运输规划与管理; 严余松(1963-), 男, 四川简阳人, 教授, 博导, 博士, 主要研究方向为交通运输规划及系统优化、现代物流信息与技术、系统仿真; 廖百胜(1980-), 男, 重庆开县人, 讲师, 硕士, 主要研究方向为建筑环境; 夏国恩(1977-), 男, 四川内江人, 博士, 主要研究方向为管理信息系统。

$i = 1, 2, \dots, l$ 。函数回归问题就是寻找一个最优的函数 $f(x) = wx + b$ 。其中 $w, x \in R^n, b \in R$ 使得下面的泛函最小^[11]：

$$R(f) = (1/2) \|w\|^2 + C \sum_{i=1}^l L^\epsilon(x_i, y_i, f) \tag{1}$$

其中： $\|w\|^2$ 为结构风险；参数 C 为松弛因子，可控制 w 的大小； $L^\epsilon(x, y, f)$ 为 ϵ -不敏感损失函数，定义为

$$L^\epsilon(x, y, f) = |y - f(x)|_\epsilon = \max(0, |y - f(x)| - \epsilon) \tag{2}$$

这里 f 是域 X 上的实值函数， $x \in X$ 并且 $y \in R$ 。

通过引进正的松弛变量 ξ, ξ^* ，式(1)的最优化问题如下：

$$\begin{aligned} \min \Phi &= (1/2) \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \\ \text{s. t. } &\begin{cases} (w \times x_i + b) - y_i \leq \epsilon + \xi_i \\ y_i - (w \times x_i + b) \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0; i = 1, \dots, l \end{cases} \end{aligned} \tag{3}$$

式(3)为一个凸二次优化问题；引入 Lagrange 函数，相应的对偶问题可用标准方法导出：

$$\begin{aligned} \max W &= \sum_{i=1}^l (\alpha_i^* - \alpha_i) y_i - \epsilon \sum_{i=1}^l (\alpha_i^* + \alpha_i) - \\ &\quad (1/2) \sum_{i,j=1}^l (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) x_i \times x_j \\ \text{s. t. } &\begin{cases} 0 \leq \alpha_i, \alpha_i^* \leq C; i = 1, \dots, l \\ \sum_{i=1}^l (\alpha_i^* - \alpha_i) = 0; i = 1, \dots, l \end{cases} \end{aligned} \tag{4}$$

它对应的 KKT 互补条件是

$$\begin{cases} \alpha_i (w \times x_i + b - y_i - \epsilon - \xi_i) = 0; i = 1, \dots, l \\ \alpha_i^* (y_i - w \times x_i - b - \epsilon - \xi_i^*) = 0; i = 1, \dots, l \\ \xi_i \xi_i^* = 0, \alpha_i \alpha_i^* = 0; i = 1, \dots, l \\ (\alpha_i - C) \xi_i = 0, (\alpha_i^* - C) \xi_i^* = 0; i = 1, \dots, l \end{cases}$$

求解以上的二次规划问题，得到最优的 Lagrange 乘子 α, α^* 以及阈值 b 。

非线性时，先使用一个非线性映射将样本映射到一个高维特征空间 H ，将 x 作变换： $\Phi: R^d \rightarrow H$ 为 $\Phi(x)$ 再在高维特征空间进行回归，关键问题是核函数的采用。

通过核函数，回归决策函数 $f(x)$ 可直接表示为

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i, x) + b \tag{5}$$

根据上述分析，求回归函数 $f(x)$ 实际上可归结为求 α 和 α^* 。可通过条件极小化 $f(x)$ 确定 α 和 α^* 。

2 区域物流需求预测模型

2.1 影响区域物流需求的因素分析

影响区域物流需求的因素很多，主要有经济因素(主要包括区域经济整体水平和规模、区域经济产业结构和产品结构、区域经济的空间布局等)、物流本行业因素(主要包括物流设施和服务、物流费用的变动等)、环境因素(主要包括技术进步、经济政策和经济体制等)、其他因素(主要有区域地理位置、突发因素等)。这些因素影响着区域物流需求的增长或减少。影响因素与物流需求之间存在着一种内在的、隐含的映射关系，由于各影响因素对物流需求的影响角度和程度不尽相同，这种复杂的内在关系决定了各影响因素与物流需求之间的多元非线性映射关系。

2.2 “影响因素—区域物流需求”预测模型建立

为了定量地研究区域物流需求，需要建立预测模型来反映影响因素与物流需求之间的多元非线性映射关系。其数学表

达式表示如下：

$$Q = f(X, Y, Z, O) \tag{6}$$

其中： Q 表示区域物流需求； $X = (x_1, x_2, x_3, \dots)$ 表示区域经济各影响因素； $Y = (y_1, y_2, y_3, \dots)$ 表示物流本行业各影响因素； $Z = (z_1, z_2, z_3, \dots)$ 表示各环境影响因素； $O = (o_1, o_2, o_3, \dots)$ 表示其他各影响因素。

用 $Q_i = f(X_i, Y_i, Z_i, O_i)$ 表示该区域第 i 年的物流需求量。其中： $X_i = (x_{i1}, x_{i2}, x_{i3}, \dots)$ 表示该区域第 i 年经济各影响因素向量； $Y_i = (y_{i1}, y_{i2}, y_{i3}, \dots)$ 表示该区域第 i 年物流本行业各影响因素向量； $Z_i = (z_{i1}, z_{i2}, z_{i3}, \dots)$ 表示该区域第 i 年各环境影响因素向量； $O_i = (o_{i1}, o_{i2}, o_{i3}, \dots)$ 表示该区域第 i 年其他各影响因素向量。考虑到突发因素的不可预知性，故在实际预测区域物流需求时通常不考虑它，另外，由于统计数据的可得性问题，加之物流需求是经济发展带来的派生需求，经济因素是影响区域物流需求的决定性因素，在实际预测区域物流需求中通常只考虑区域经济的各影响因素。因为经济影响因素在短时期内很少出现大幅变动，所以可以用第 $i-1$ 年的经济影响因素来预测第 i 年的物流需求量。将 $X_{i-1} = (x_{(i-1)1}, x_{(i-1)2}, x_{(i-1)3}, \dots)$ 作为支持向量回归机的输入变量，将 Q_i 作为其输出变量， (X_{i-1}, Q_i) 组成一个训练对。基于 ϵ -SVR 的回归预测实际上是一个数据泛化拟合问题，即先根据输入/输出样本进行训练，然后对不在训练样本集中的输入数据计算出相应的输出值。由于 ϵ -SVR 回归函数类似于一个神经网络，其输出是若干中间节点的线性组合，每个中间节点对应于输入样本与一个支持向量的内积^[12]，可以得到如图 1 所示的区域物流需求 ϵ -SVR 网络预测模型。

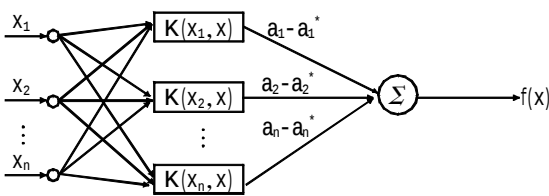


图 1 支持向量回归机结构模型

3 预测模型的应用实例

3.1 数据来源

以上海市物流需求预测为例，考虑到统计数据可获得性，本文选取货物运输量作为物流需求规模指标。根据对区域物流需求的影响因素分析，考虑到行业因素、环境因素等指标获得很难，加之经济因素是影响区域物流最直接最根本的因素，本着可操作性原则，选取各经济影响因素指标用于物流需求规模预测，具体选择的经济指标为：第一产业产值、第二产业产值、第三产业产值、区域零售总额、区域外贸总额、人均消费水平等。其中，三大产业产值不仅考虑了区域经济总量，还考虑了上海市经济结构对物流需求规模影响；由于商业流通也是区域物流需求的重要组成部分，设立了区域零售总额与人均消费水平两个经济指标。此外，考虑到上海的全国对外贸易港口地位，具有外向性的港口物流将占整个物流需求的较大部分，因此将区域外贸总额考虑入内。上海市物流需求规模及经济指标统计数据(1978—2004 年)见参考文献[9]。

3.2 数据处理

为了准确地选择 SVR 中各参数和减少计算复杂度，对每一个指标都采用 $x'_{ij} = x_{ij}/x_{\max j}$ 对收集到的数据进行归一化处理。其中： x'_{ij} 为第 i 个指标的第 j 个数据归一化后的数值； x_{ij} 为

第 i 个指标的第 j 个数据的实际值; x_{imax} 为第 i 个指标所有数据中的最大值(如对第 2 个指标, $x_{imax} = 2\ 977.61$)。对于最大值的归一化结果, 考虑到 SVR 能够获得较好的收敛性, 其取值为 0.999 9。结果如表 1 所示。将 1978—1995 年的样本作为训练样本, 1996—2003 年的样本作为测试样本。

表 1 上海市物流需求规模及经济指标统计数据							
年 份	指 标						
	x_1 第 一产业产 值 /亿元	x_2 第 二产业产 值 /亿元	x_3 第 三产业产 值 /亿元	x_4 区 域 零 售 总 额 / 亿 元	x_5 区 域 外 贸 总 额 / 亿 美 元	x_6 居 民 消 费 水 平 /元	y 货 物 运 输 量 / 万 吨
1978	0.121 9	0.070 9	0.017 1	0.024 4	0.027 5	0.029 7	
1979	0.126 2	0.074 3	0.018 1	0.030 8	0.035 3	0.034 4	0.270 1
1980	0.111 9	0.079 3	0.022 1	0.036 2	0.041	0.037 2	0.276
1981	0.117 2	0.082 1	0.023 5	0.04	0.037 8	0.039 3	0.277 5
1982	0.147 4	0.083 7	0.025	0.040 4	0.035 4	0.038 7	0.291 3
1983	0.149 8	0.085 7	0.027 9	0.045 3	0.037 7	0.041 4	0.297 4
1984	0.191 2	0.092 5	0.033	0.055 7	0.04	0.048 8	0.318 4
1985	0.216 4	0.109 4	0.040 9	0.078 1	0.047 1	0.069 3	0.333 9
1986	0.218 1	0.112 8	0.045 4	0.088 7	0.047 4	0.08	0.367 3
1987	0.239 3	0.122 4	0.053 6	0.101 4	0.054 6	0.087 3	0.375 2
1988	0.303 1	0.145 4	0.063 1	0.133 2	0.065 9	0.113	0.383 3
1989	0.328 2	0.156 6	0.067 4	0.149 2	0.071 4	0.062 4	0.381
1990	0.361 1	0.162 1	0.081	0.150 4	0.067 6	0.135 1	0.368 8
1991	0.369 6	0.185 2	0.103 8	0.172 1	0.073 2	0.162 8	0.379 6
1992	0.378 4	0.227 5	0.135 3	0.209 3	0.088 8	0.191 2	0.407 4
1993	0.423 3	0.302 4	0.192 5	0.281 2	0.115 9	0.279 9	0.417 2
1994	0.538 3	0.383 9	0.262 1	0.347 1	0.144 4	0.359 4	0.393 7
1995	0.683 3	0.473 5	0.333	0.436 9	0.173 2	0.451 5	0.379 7
1996	0.793 0	0.531 5	0.431 4	0.523	0.202 6	0.520 8	0.631 1
1997	0.839 7	0.589 2	0.514 1	0.596 8	0.225 4	0.585 1	0.632 7
1998	0.869 6	0.620 4	0.592 2	0.662 5	0.285 3	0.619	0.636 7
1999	0.886 2	0.656 2	0.672 3	0.716 3	0.351 4	0.694 7	0.666 6
2000	0.921 7	0.726 7	0.774 2	0.775 7	0.498	0.776 6	0.719
2001	0.947 2	0.791 1	0.843 3	0.838 3	0.554 3	0.845	0.747 7
2002	0.977 5	0.861 3	0.925 9	0.916 6	0.661 4	0.961 5	0.780 3
2003	0.999 9	0.999 9	0.999 9	0.999 9	0.999 9	0.999 9	0.841 2
2004							0.999 9

3.3 仿真结果及分析

针对上述的“影响因素—区域物流需求 ”SVR 预测模型, 在 MATLAB 6.5 环境下进行训练与测试的仿真实验。SVR 的核函数和参数采用折交叉验证方法进行优化选择, 为 $K(x_i, x_j) = ((x_i \times x_j) + 1)^2$, $\varepsilon = 0.12$, $C = 0.05$ 。测试样本的仿真结果如表 2 所示。总体上说, 预测结果能满足预测要求, 误差比较小, 拟合程度较高。

本文阐述的“影响因素—区域物流需求 ”SVR 预测模型, 是建立在对区域物流需求与其影响因素(主要是区域经济) 之间内在关系研究的基础上提出的。该模型在一定程度上反映了区域物流需求与其影响因素之间的复杂映射关系, 它不同于以往简单的物流需求线性回归或时间序列预测模型。

表 2 上海市物流需求规模预测结果				
年 份	真 实 值	预 测 值	误 差	误 差 率 /%
1997	45 938	40 515.07	- 5 422.93	- 11.804 9
1998	46 230	44 249	- 1 981	- 4.285 1
1999	48 398	47 208.33	- 1 189.67	- 2.458 1
2000	52 206	50 460.39	- 1 745.61	- 3.343 7
2001	54 049	55 232.57	1 183.57	2.189 8
2002	56 652	59 512.87	2 860.87	5.049 9
2003	61 073	65 187.73	4 114.73	6.737 4
2004	72 605	72 330.12	- 274.88	- 0.378 6

4 结束语

本文将 SVR 方法应用于区域物流需求预测研究, 从区域经济等影响因素指标与区域物流需求之间的内在关系的角度, 建立了“影响因素—区域物流需求 ”SVR 预测模型, 采用 ε -insensitive 损失函数, 选用多项式核函数, 其实际结果有相当理想的精度, 能满足预测要求。“影响因素—区域物流需求 ”SVR 预测模型有很强的学习性、自适应性, 且收敛快、准确性高, 说明该模型是可信的。但对核函数的选择, 参数和损失函数的选取没有确定的方法, 主要靠使用者的经验, 难以找到全局最优解, 只能找到满意解, 对此问题还需要进一步研究。

参考文献:

[1] ARRIDO R A, MAHMASSANI H S. Forecasting freight transportation demand with the space-time multinomial probity model[J]. Trans on Research Part B, 2000, 34(5) : 403-418.

[2] FITE J, TAYLOR G, USHER J, *et al.* Forecasting freight demand using economic indices[J]. International Journal of Physical Distribution & Logistics Management, 2001, 31(4) : 299.

[3] 张拥军, 叶怀珍, 任民. 神经网络模型预测运输货运量[J]. 西南交通大学学报, 1999, 34(5) : 602-605.

[4] 王隆基, 张仲鹏, 孙晓霞. 基于 BP 神经网络的物流预测方法[J]. 起重运输机械, 2005(5) : 30-32.

[5] 汪宇瀚. 预测物流需求的一元线性回归分析法[J]. 商场现代化, 2006(34) : 136.

[6] 陈森, 周峰. 基于灰色系统理论的物流需求预测模型[J]. 统计与决策, 2006(3) : 59-60.

[7] 初良勇, 田质广, 谢新连. 组合预测模型在物流需求预测中的应用[J]. 大连海事大学学报, 2004, 30(4) : 43-46.

[8] 陈黎. 我国区域物流发展预测[J]. 统计与决策, 2006(12) : 127-129.

[9] 后锐, 张毕西. 基于 MLP 神经网络的区域物流需求预测方法及其应用[J]. 系统工程理论与实践, 2005(12) : 43-47.

[10] VAPNIK V. An overview of statistical learning theory[J]. IEEE Trans on NN, 1999, 10(3) : 988-999.

[11] CRISTIANINI N, SHAWE-TAYLOR J. 支持向量机导论[M]. 李国正, 王猛, 曾华军, 译. 北京: 电子工业出版社, 2004.

[12] 夏国恩, 曾绍华, 金炜东. 支持向量回归机在铁路客运量时间序列预测中的应用[J]. 计算机应用研究, 2006, 23(10) : 180-182.

[16] 何庆, 汤庸, 黄永钊. 基于本体的法律知识库的研究与实现[J]. 计算机科学, 2007, 34(2) : 175-177.

[17] NEWCOMER E, LOMOW G. Understanding SOA with Web services [M]. [S. l.] : Addison-Wesley Professional, 2004.

[18] McINTOSH, ROGER L. Open-source tools for distributed device control within a service-oriented architecture[J]. Journal of the Association for Laboratory Automation, 2004, 9(6) : 404-410.

[19] KNIGHT, W. Security: built-in or bolted-on to the SOA? [J]. Infosecurity Today, 2005, 2(2) : 38-40.

[20] 廖述梅. 基于本体的语义标注原型评述[J]. 计算机工程与科学, 2006, 28(9) : 123-125, 128.

(上接第 2737 页)

[13] IU Xiao-li, WU Guo-qing, YANG Min. Building knowledge base from semi-structured data[C] //Proc of International Conference on Machine Learning and Cybemetics. Hong Kong: IEEE, 2007: 839-843.

[14] XU Bin-feng, LUO Xiao-gang, PENG Cheng-lin, *et al.* Based on ontology: construction and application of medical knowledge base[C] //Proc of IEEE/ICME International Conference on Complex Medical Engineering. Beijing: IEEE, 2007: 1586-1589.

[15] 范黎林, 王晓东, 屈喜龙. 基于 Ontology 知识库系统建模[J]. 计算机应用研究, 2006, 23(5) : 134-136.