

文章编号:1007-6735(2012)04-0347-04

灰色支持向量机在机场吞吐量预测中的应用

徐勤兰, 樊重俊, 张 鹏
(上海理工大学 管理学院, 上海 200093)

摘要: 在分析传统预测方法不足的基础上,利用灰色支持向量机组合分析模型,以实际值与灰色模型预测值的比值序列作为支持向量机模型的输入,选取径向基函数为核函数,并通过交叉验证法选取最优参数,利用支持向量机模型分析预测比值序列,最后通过灰色模型还原为货邮吞吐量的预测值.以上海机场货邮吞吐量为例,对灰色支持向量机模型进行了实证分析,并与灰色模型、支持向量机模型进行了对比.

关键词: 机场货邮吞吐量;灰色模型;支持向量机模型;灰色支持向量机组合模型

中图分类号: N 945 **文献标志码:** A

Grey Support Vector Machine Combination Model Applied in Predicting the Throughput of Airport

XU Qin-lan, FAN Chong-jun, ZHANG Peng
(Business School, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: The deficiency of current methods for predicting the airport total cargo was analysed and the grey support vector machine combination model was introduced to predict the airport total cargo. The ratio sequence of the actual values to the results predicted by the gray model was regarded as an input of the support vector machine model, and the radial basis function was selected as a kernel function. By use of cross-validation method the optimal parameter was searched. The support vector machine model was then applied to predict the real ratio sequence, and finally the result was restored to the value of the airport total cargo by using again the gray model. The total cargo of Shanghai airport was taken as an example to illustrate the feasibility of the model, and its results were compared with those predicted by the gray model and the support vector machine model.

Key words: airport total cargo; gray model; support vector regression; grey support vector machine combination model

货邮吞吐量是机场吞吐量的重要组成部分,是 机场运营管理的重要依据之一,是实现机场资源有

收稿日期: 2012-04-13
基金项目: 国家自然科学基金资助项目(71171136)
作者简介: 徐勤兰(1986-),女,硕士研究生.研究方向:信息系统工程. E-mail: xuqinlan1986@163.com
樊重俊(联系人),男,教授.研究方向:信息系统工程. E-mail: cjfan@sh163.net

效配置的基本依据.随着我国航空业的迅速发展,机场货邮吞吐量也增长迅猛,为实现机场及航空公司的有效管理,提高货邮吞吐量的预测精度越来越成为人们关注的重点.传统的预测方法按性质大致可分为两类^[1-2]:定性预测法和定量预测法.定性预测法注重事物发展性质方面的预测,具体方法主要有德尔非法、主观概率法、专家会议法,这些方法有较大的灵活性、操作简单、可靠性好,但受主观因素影响较大.定量预测法关注事物在数量上的变化,主要包括时间序列法、趋势外推法、计量经济法、重力模型法、灰色预测法、神经网络法等,但神经网络预测法具有较强的经验色彩,受网络结构复杂度和样本复杂度的影响较大,会出现泛化能力差和过学习等现象.这些传统的预测方法通常在处理线性数据时有较好的应用,但在处理非线性数据时常常不能满足人们的要求.随着人们对数据处理的精度要求越来越高,而单一的预测模型必然存在某方面的盲区,因此,利用组合预测模型进行数据处理,优势互补,提高预测精度,成为一种研究的趋势.本文重点关注灰色系统与支持向量回归机组合模型的研究以及在机场货邮吞吐量预测中的应用.

灰色系统理论是我国著名学者邓聚龙于1982年创立的一门学科,以“部分信息已知,部分信息未知”的“小样本”、“贫信息”不确定系统为研究对象.GM(1,1)模型是灰色系统理论的重要组成部分,具有建模所需数据少的特点^[3].支持向量机(support vector machines, SVM)是Vapnik等在统计学习理论(statistical learning theory, SLT)的基础之上提出的一种新的机器学习方法^[4].它基于结构风险最小化原则来提高泛化能力,有效解决了小样本、非线性、高维数、局部极小点等实际问题,具有良好的推广性和较好的分类精确性.

将灰色系统与支持向量机组合起来得到灰色支持向量机模型,此模型既具有灰色系统在处理“小样本”、“贫信息”数据方面的优势,又具有支持向量机在处理不规则、混沌等非线性数据方面的自适应能力.

1 灰色预测模型^[3]

灰色预测的实质是将“随机过程”当作“灰色过程”,“随机量”当作“灰色量”,并以灰色系统理论中的GM(1,1)模型为主进行数据分析处理.GM(1,1)模型的主要建模过程如下:

令非负原始序列为 $x^{(0)}$,对 $x^{(0)}$ 作一次累加生成得到1-AGO序列,即

$$\begin{cases} x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)) \\ x^{(1)}(k) = \sum_{r=1}^k x^{(0)}(r), k = 1, 2, \dots, n \end{cases}$$

则 $x^{(0)}$, $x^{(1)}$ 符合灰导数条件,将 $x^{(0)}$, $x^{(1)}$ 各时刻数据代入灰色微分方程

$$x^{(0)}(k) + az^{(1)}(k) = b$$

其中

$$\begin{cases} B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix} \\ y_N = [x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n)] \\ z^{(1)}(k) = 0.5[x^{(1)}(k) + x^{(1)}(k-1)] \\ \hat{a} = (a, b)^T \end{cases}$$

利用最小二乘法估计其参数为: $\hat{a} = (B^T B)^{-1} B^T y_N$,称 $dx^{(1)}/dt + ax^{(1)} = b$ 为灰微分方程 $x^{(0)}(k) + az^{(1)}(k) = b$ 的白化方程或影子方程,GM(1,1)影子方程的解为

$$\bar{x}^{(1)}(k+1) = [x^{(0)}(1) - \frac{b}{a}]e^{-ak} + \frac{b}{a}$$

将 $\bar{x}^{(1)}(k+1)$ 作累减生成得 $\bar{x}^{(0)}(k+1)$,即

$$\bar{x}^{(0)}(k+1) = \bar{x}^{(1)}(k+1) - \bar{x}^{(1)}(k)$$

2 支持向量机模型^[5-8]

支持向量机模型已在模式识别、回归分析、文章分类等领域得到了应用,并取得了较好的效果,其基本思想如下:

对于一个给定的样本集 $D = \{(x_i, y_i) | i = 1, 2, \dots, n\}$, $x_i \in \mathbf{R}^n$, $y_i \in \mathbf{R}$,使用一个非线性映射 $\varphi(\cdot)$ 将样本的输入 x_i 由输入空间 X 映射到一个高维特征空间 H ,并在特征空间 H 中构造线性支持向量回归机,即 $f(x) = (\omega \cdot \varphi(x)) + b$.其中映射函数 $\varphi(\cdot)$ 满足Mercer条件,映射的完成由 $\varphi(\cdot)$ 构成的核函数 $K(x_i, y_j) = \varphi(x_i) \cdot \varphi(y_j)$ 实现.

这时非线性回归问题的优化方程转化为

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

其中, $\xi_i, \xi_i^* \geq 0, i = 1, 2, \dots, n$,且约束于

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

式中, ξ_i, ξ_i^* 为松弛变量; C 为惩罚系数; ε 为不敏感损失误差. 这是一个二次规划问题, 引入Lagrange 函数

$$\begin{aligned} L(\omega, b, a, a^*) = & \frac{1}{2}|\omega|^2 + C\sum_{i=1}^n(\xi_i + \xi_i^*) - \\ & \sum_{i=1}^n a_i(\xi_i + \varepsilon - y_i + (\omega \cdot \varphi(x_i)) + b) - \\ & \sum_{i=1}^n a_i^*(\xi_i^* + \varepsilon - y_i - (\omega \cdot \varphi(x_i)) - b) - \\ & \sum_{i=1}^n \eta_i(\xi_i + \xi_i^*) \end{aligned}$$

其中, $a_i, a_i^* \geq 0, i = 1, 2, \dots, n$, 所以原问题的 Lagrange 对偶问题

$$\begin{aligned} \min_{a_i, a_i^*} \quad & \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (a_i - a_i^*)(a_j - a_j^*) K(x_i, x_j) - \\ & \sum_{i=1}^n y_i(a_i - a_i^*) + \varepsilon \sum_{i=1}^n (a_i + a_i^*) \\ \text{约束于} \quad & \sum_{i=1}^n (a_i - a_i^*) = 0, 0 \leq a_i, a_i^* \leq C, \\ & i = 1, 2, \dots, n \end{aligned}$$

求解此对偶问题, 可得到回归决策函数为

$$f(x) = (\omega \cdot \varphi(x)) + b = \sum_{i=1}^n (a_i - a_i^*) \cdot K(x_i \cdot x) + b$$

3 灰色支持向量机建模过程

机场货邮吞吐量的产生受多种因素的影响, 而许多影响因素在做数据预测时不易定量处理或有些因素尚不明确, 因此本文尝试结合灰色模型与支持向量机模型各自的优点, 建立灰色支持向量机模型, 基本思想如下:

- a. 设原始序列为
$$X^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$$
其中, $x^{(0)}(k) \geq 0, k = 1, 2, \dots, n$, 利用上述 $G(1, 1)$ 模型对原始序列进行分析预测, 得到预测值
$$\hat{X}^{(0)} = (\hat{X}^{(0)}(1), \hat{X}^{(0)}(2), \dots, \hat{X}^{(0)}(n))$$
其中, $\hat{X}^{(0)}(k) \geq 0, k = 1, 2, \dots, n$
- b. 设原始序列值 $X^{(0)}(k)$ 与 $G(1, 1)$ 模型预测值 $\hat{X}^{(0)}(k)$ 的比值为 $d^{(0)}(k), k = 1, 2, \dots, n$. 从比值序列中选取模型的训练样本集和测试样本集.
- c. 选择适当的核函数 $K(x_i, x)$ 及参数.
- d. 用支持向量机方法求解比值序列优化问题

- 的 Lagrange 对偶问题, 并构造回归决策函数 $f(x)$.
- e. 利用回归决策函数 $f(x)$, 计算出比值序列 $d^{(0)}(k)$ 的预测值 $\hat{d}^{(0)}(k)$.
 - f. 将预测值 $\hat{d}^{(0)}(k)$ 乘以 $\hat{X}^{(0)}(k)$ 即得灰色支持向量机的预测结果.

4 实证研究

以上海浦东机场和上海虹桥机场 2009 年 9 月~2010 年 12 月货邮吞吐量数据为研究对象, 数据见表 1.

表 1 2009 年 9 月~2010 年 12 月年上海机场货邮吞吐量
Tab.1 Cargo and mail throughput of the Shanghai airport in 2009.9~2010.12 万 t

年-月	编号	吞吐量	年月	编号	吞吐量
2009-09	01	28.45	2010-05	09	33.76
2009-10	02	27.92	2010-06	10	30.99
2009-11	03	29.72	2010-07	11	33.05
2009-12	04	29.60	2010-08	12	31.14
2010-01	05	28.38	2010-09	13	30.05
2010-02	06	22.90	2010-10	14	32.79
2010-03	07	28.70	2010-11	15	33.25
2010-04	08	32.36	2010-12	16	31.60

首先, 利用表中 01~12 的数据建立灰色 $G(1, 1)$ 模型, 并利用此模型对 2010 年 9 月~2010 年 12 月的吞吐量进行分析预测, 其预测结果及误差见表 2.

表 2 模型预测值及误差
Tab.2 Predictive value and error of the model 万 t

年-月	吞吐量 实际值	预测值		
		灰色模型	支持向量机	灰色支持向量机
2010-09	30.05	32.94	32.72	31.14
2010-10	32.79	33.49	30.20	31.85
2010-11	33.25	34.05	32.29	33.23
2010-12	31.6	34.62	31.28	32.77
平均相对误差(%)		5.93	5.18	2.56

- 其次, 利用支持向量机与灰色支持向量机理论建立分析预测模型, 其间需要做如下工作:
- a. 选取训练样本与测试样本 支持向量机选取原始数据中 01~12 的数据作为训练样本, 13~16 的数据作为测试样本. 灰色支持向量机选取比值序列中 01~12 的数据作为训练样本, 并利用均值生成

法处理编号为 06 的异常数据,13~16 的数据作为测试样本。

b. 选择核函数 由于径向基函数具有较好的普适性,且在处理时间序列问题时表现一般比其它核函数更好,因此,选用径向基函数 $K(x_i, x_j) = \exp\left(-\frac{|x - x_i|^2}{2\sigma^2}\right)$ 作为灰色支持向量机模型与支持向量机模型的核函数。

c. 选择最优参数 需要确定的参数有不敏感损失误差 ϵ 、惩罚系数 C 以及核函数参数 σ 。然而,对这 3 个参数的选择,目前尚没有较好的方法,也没有较成熟地选取理论与指导原则^[9],其主要的选择算法有留一法、交叉验证法、试凑法、网格搜索法、梯度下降法和免疫算法等。本文选取计算精度较高的交叉验证法作为参数的选择算法,其基本原理为:将训练数据集分成 k 份相等的子集,每次将其中 $k-1$ 份数据作为训练数据,而将另外一份数据作为测试数据。这样重复 k 次,根据 k 次迭代后得到的均方误差平均值来估计期望泛化误差,最后选择一组最优的参数^[10]。

经计算得支持向量机的最优参数为 $C = 4.0$ 、 $\sigma = 8.0$ 、 $\epsilon = 0.001$;灰色支持向量机的最优参数为 $C = 256.0$ 、 $\sigma = 0.001$ 、 $\epsilon = 0.001$ 。

最后,分别利用支持向量机与灰色支持向量机对数据进行分析预测,预测及对比结果分别见表 2 和图 1。

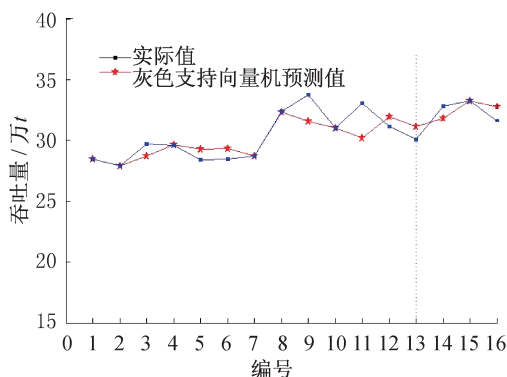


图 1 灰色支持向量机分析结果与实际值的比较

Fig.1 Compare gray support vector machine prediction results with the actual value

5 结束语

从结果及误差可以看出灰色支持向量机模型在分析预测机场货邮吞吐量方面明显优于灰色及支持向量机模型。灰色模型与支持向量机模型的融合弥补了单一模型在数据处理方面的缺陷,尽管组合模型中有个别数据拟合误差偏大,但从整体预测效果看,仍然优于单一的灰色模型或支持向量机模型。

由于机场货邮吞吐量的产生受多种因素影响,包括宏观经济发展、贸易发展、资源状况等因素等,因此,预测机场货邮吞吐量时,不应仅仅考虑时间因素,也应综合考虑社会经济等综合因素,以建立更准确地预测模型,这也是以后要进一步深入研究的方向。

参考文献:

- [1] 王芳. 基于支持向量机的航段运量预测研究[D]. 江苏:南京航空航天大学,2007.
- [2] 张慧,王喆. 机场吞吐量预测方法探讨[J]. 中国民用航空,2008,10(94):67-68.
- [3] 邓聚龙. 灰理论基础[M]. 武汉:华中科技大学出版社,2002.
- [4] Vapnik V. The nature of statistical learning theory [M]. New York:Springer,1999.
- [5] 董雁萍. 支持向量机预测模型的构建及其应用[D]. 西安:西安理工大学,2010.
- [6] 梁路宏,艾海舟,张钊. 基于模板匹配与支持向量机的人脸检测[J]. 计算机学报. 2002,25(1):22-29.
- [7] Thissen U, Brakel R, Weijer A P, et al. Using support vector machines for time series prediction [J]. Chemometrics and Intelligent Laboratory System, 2003,69(1):35-49.
- [8] Gavrishchaka V V, Ganguli S B. Volatility forecasting from multiscale and high-dimensional market data[J]. Neurocomputing,2003,55(1):285-305.
- [9] 朱家元,张喜斌. 多参数装备费用的支持向量机预测[J]. 系统工程与电子技术,2003,25(6):701-703.
- [10] 奉国和. SVM 分类核函数及参数选择比较[J]. 计算机工程与应用,2011,47(3):123-128.