

汽车销售混合预测方法研究

李 响¹, 宗 群², 童 玲²

(1. 天津大学管理学院, 天津 300072; 2. 天津大学自动化学院, 天津 300072)

摘 要:市场分析和预测已成为企业重要的决策依据和手段。就汽车销量问题提出了一种 ARMA 模型与 RBF 神经网络相结合的混合预测方法。采用 ARMA 模型对汽车销量趋势进行初步线性预测,利用 RBF 神经网络对线性预测的残差建模,得到非线性预测,两部分预测输出和为总的预测值。该方法既体现了销售量数据间的线性关系,又揭示了数据内部的非线性特征,克服了单一方法的局限性,提高了预测精度。仿真结构分析表明,该方法预测效果最佳。

关键词:混合预测; ARMA 模型; RBF 神经网络; 汽车销售预测

中图分类号:F713.3 **文献标志码:**A **文章编号:**1008-4339(2006)03-0175-04

在市场经济不断发展,竞争日趋激烈的今天,各大企业为了使其经济活动达到预期的目的,都需要通过分析历史数据,对某些经济变量进行预测,从而有利于制定今后的营销战略和生产计划。目前,我国汽车工业正处在成长阶段,新技术和新政策的大力支持以及人们生活水平的不断提高使得汽车市场容量不断扩大,销量不断上升,因而准确的销量预测则成为汽车市场竞争中至关重要的环节。

目前,在汽车销售领域中所选用的传统的预测方法都是以时间序列预测为核心,如线性回归法、指数平滑法、季节预测法等。线性回归法(如 AR、MA、ARMA 模型等)仅仅能体现销售数据随时间的直线形上升或下降趋势,只用作一般性的趋势观察来使用,预测结果误差较大。指数平滑法代表性地反映了销售情况的长期趋势,适合长期预测且误差较小,但短期预测效果并不理想。季节预测方法有效地反映了汽车销售随季节波动的特点,有一定的可靠性,但不能反映其他非线性因素带来的影响,预测效果欠佳。出于以上考虑,本文提出了一种基于 ARMA 模型和 RBF 神经网络相结合的汽车销售量混合预测方法。认定销量数据对应的时间序列是由线性元和非线性元两部分组成的,因此该方法采用 ARMA 模型实现销量数据中线性部分的预测模型拟合,同时利用 RBF 神经网络对非线性系统逼近效果

好的优势,解决销量数据中非线性部分预测问题。

一、汽车销售量混合预测模型基本原理

(一)汽车销量时间序列分析

汽车销售量是指单位时间内售出汽车数量随时间随机分布的时间序列,由于汽车销售活动是一个由人参与的、时变的、复杂的体系,它的显著特点之一就是具有高度的不确定性,往往很难确定该系统是严格的线性系统还是严格的非线性系统。基于这种情况,则认为汽车销量时间序列由线性和非线性两部分混合组成,对于线性部分采用被证明处理线性系统有明显优势的 ARMA 模型^[1]进行拟合,对于非线性部分则采用被广泛应用且证明有良好自适应性的神经网络进行预测问题描述^[2]。

假设汽车销量数据的数学描述为

$$X_t = X_{L_t} + X_{N_t} \quad (1)$$

式中: X_t 为 $t-1$ 到 t 这段时间内所售出的汽车数量; X_{L_t} 为销量数据中线性部分; X_{N_t} 为销量数据中非线性部分。

利用 ARMA 模型来构建线性部分 X_{L_t} ,那么线性模型的残差部分就是非线性关系,再用 RBF 神经

收稿日期:2004-08-30.

作者简介:李 响(1965—),男,博士研究生,高级工程师.

网络来处理。由于任何固定的线性模型或非线性模型都不能准确地预测出目标值,必然存在一定的残差,在该方法中视线性预测的残差为数据内部的非线性特性,即汽车销量序列是在线性序列的基础上又叠加了非线性随机序列。因此,令 e_t 表示 t 时刻线性模型的残差,则

$$e_t = x_t - \hat{x}_{L_t}$$

式中: x_t 为 t 时刻汽车销量的实际值; \hat{x}_{L_t} 为利用 AR-MA 模型在 t 时刻的线性估计。

利用 RBF 神经网络为残差 e_t 建模,可表示为

$$e_t = f(e_{t-1}, e_{t-2}, \dots, e_{t-n}) + \varepsilon_t \quad (2)$$

式中: f 为由 RBF 神经网络决定的非线性函数; ε_t 为随机误差。

由式(2)得到的预测值记为 \hat{x}_{N_t} , 则混合预测值可表示为

$$\hat{x}_t = \hat{x}_{L_t} + \hat{x}_{N_t} \quad (3)$$

汽车销售量混合预测方法的结构见图1。

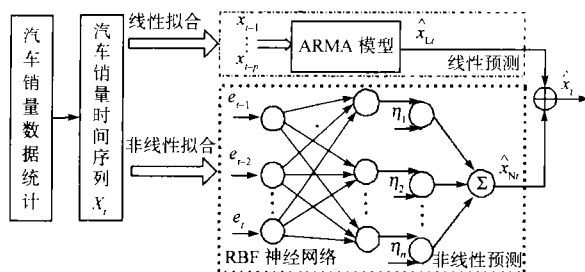


图1 汽车销售量混合预测方法结构图

(二) 汽车销售量混合模型的建立

1. ARAM 模型的线性预测

采用 ARAM 模型对汽车销量进行线性拟合。ARMA 模型是 MA 过程和 AR 过程联合构成的模型,包含 p 阶 AR 项和 q 阶 MA 项 (ARMA(p, q) 过程), 可写为

$$x_t - \alpha_1 x_{t-1} - \alpha_2 x_{t-2} - \dots - \alpha_p x_{t-p} = \beta_0 z_t + \beta_1 z_{t-1} + \dots + \beta_q z_{t-q} \quad (4)$$

式中: $\beta_i (i=0, 1, \dots, q)$ 与 $\alpha_i (i=1, 2, \dots, p)$ 分别为模型的参数和常数; $x_{t-i} (i=0, 1, 2, \dots, p)$ 为 $t-i$ 时刻的实际值; $z_{t-i} (i=0, 1, \dots, q)$ 为 $t-i$ 时刻的随机误差值。

(1) ARMA 模型阶数的确定^[3]。根据样本数据计算样本的自协方差基函数 $\hat{\gamma}_k$ 和自相关函数值 $\hat{\rho}_k$,

样本自协方差基函数

$$\hat{\gamma}_k = \frac{1}{N} \sum_{i=1}^{N-k} (x_i - \bar{x})(x_{i+k} - \bar{x}), k=0, 1, 2, \dots \quad (5)$$

式中: N 为样本的数量; $\hat{\gamma}_k$ 为第 k 个样本的自协方差基函数值

自相关函数值

$$\hat{\rho}_k = \hat{\gamma}_k / \hat{\gamma}_0, k=0, 1, 2, \dots \quad (6)$$

样本的偏相关函数值 $\hat{\phi}_{kk}$ 为

$$\begin{pmatrix} \hat{\phi}_{k1} \\ \hat{\phi}_{k2} \\ \dots \\ \hat{\phi}_{kk} \end{pmatrix} = \begin{pmatrix} 1 & \hat{\rho}_1 & \hat{\rho}_2 & \dots & \hat{\rho}_{k-1} \\ \hat{\rho}_1 & 1 & \hat{\rho}_1 & \dots & \hat{\rho}_{k-2} \\ \dots & \dots & \dots & \dots & \dots \\ \hat{\rho}_{k-1} & \hat{\rho}_{k-2} & \hat{\rho}_{k-3} & \dots & 1 \end{pmatrix} \begin{pmatrix} \hat{\rho}_1 \\ \hat{\rho}_2 \\ \dots \\ \hat{\rho}_k \end{pmatrix} \quad (7)$$

通常 $\hat{\phi}_{kk}$ 可利用下面的递推公式计算

$$\begin{cases} \hat{\phi}_{11} = \hat{\rho}_1 \\ \hat{\phi}_{k+1, k+1} = [\hat{\rho}_{k+1} - \sum_{j=1}^k \hat{\rho}_{k+1-j} \hat{\phi}_{kj}] [1 - \sum_{j=1}^k \hat{\rho}_j \hat{\phi}_{kj}]^{-1} \\ \hat{\phi}_{k+1, j} = \hat{\phi}_{kj} - \hat{\phi}_{k+1, k+1} \hat{\phi}_{k, k-(j-1)}, j=1, 2, \dots, k \end{cases} \quad (8)$$

若 ρ_k 在 $k > q$ 时为零, 则具有截尾性, 可以计算模型的阶数 q 。在实际中, 当 $|\hat{\rho}_k| < 6/\sqrt{N}$ 时, 认为在第 k 步具有截尾性, 则模型的阶数 $q = k$; 同理, 若 ϕ_{kk} 在 $k > p$ 时趋于零, 则具有截尾性, 可以计算模型的阶数 p 。在实际中, 当 $|\hat{\phi}_{kk}| < 2/\sqrt{N}$ 时, 则认为在第 k 步具有截尾性, 则模型的阶数 $p = k$ 。

(2) ARMA 模型的参数估计。模型中的参数 α_i, β_i 通过最小二乘法来确定^[4]。残差平方和函数

$$\delta = (x_t - z_t)^2 = (\alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_p x_{t-p} + \beta_1 z_{t-1} + \beta_2 z_{t-2} + \dots + \beta_q z_{t-q})^2 \quad (9)$$

按照多元函数取极小值理论, 当 $\frac{\partial \delta}{\partial \alpha_i} = 0, i=1, 2, \dots,$

$p; \frac{\partial \delta}{\partial \beta_j} = 0, j=1, 2, \dots, q$ 时, 唯一确定一组参数。

(3) ARMA 模型的预报方程。根据最小方差预报标准和时间序列的正交特性 [$E(x_t, x_{t+k}) = 0, k \neq 0$], 得出基于 ARMA 模型的 k 步线性预报值

$$\hat{x}_{L_t}(k) = \sum_{j=1}^{k-1} \alpha_j \hat{x}_{L_t}(k-j) + \sum_{j=k}^p \alpha_j x_{t-j+k} + \sum_{j=k}^q \beta_j z_{t-j+k} \quad (10)$$

式中: $\hat{x}_{t+k}(k)$ 为在 t 时刻根据 t 时刻及之前的数据基于 ARMA 模型预测第 $t+k$ 时刻的值^[5]。

2. 神经网络的非线性预测

(1) 确定径向基函数的隐含层节点中心值 c_j 。隐含节点中心 c_j 的确定可以采用“K-MEANS”方程求得,其基本原理就是尽可能地使 c_j 对输入的样本数据在数据密集处的中心点分布也密集。

(2) 求归一化参数 λ_j 。 λ_j 表示每个中心相关的数据分布测度,通常采用隐含层节点中心与训练模式之间的平均距离度量

$$\lambda_j = \sqrt{\frac{1}{n} \sum_{i=1}^n (x - c_j)^T (x - c_j)} \quad (11)$$

当 c_j 和 λ_j 确定后,基于训练样本,求取使能量函数最小的权值参数。采用 Chen 等人提出的正交最小二乘法 OLS (Orthogonal Least Squares) 确定学习网络权值 $\eta_j (j=1, 2, \dots, n)$ 。

RBF 神经网络的 k 步非线性预测输出

$$\begin{aligned} \hat{e}_{t+k} &= f(e_{t-1}, e_{t-2}, \dots, e_{t-n}) = \\ &= \sum_{j=0}^n \eta_j^{(k)} \phi_j(\cdot, \cdot) = \\ &= \sum_{j=0}^n \eta_j^{(k)} \phi(\|e_{t-j} - c_j\|, \lambda_j) = \\ &= \sum_{j=0}^n \eta_j^{(k)} \exp\left(-\frac{\|e_{t-j} - c_j\|^2}{\lambda_j^2}\right) \end{aligned} \quad (12)$$

3. 预测评估标准

为了比较 ARMA 模型、RBF 模型和本文提出的 ARMA-RBF 混合模型的预测性能,采用了误差均方值 MSE、绝对误差均值 MAE 和绝对百分比误差均值 MAPE 来评估预测的精确性。

此外,全面考虑各种因素对销售量的突发影响,对销售变化转折点的预测也具有重要意义。转折点的预测可以用 Cumby 和 Modest^[7] 提出的评测方法来评估。定义模型的预测的方向变化量为 F_t , 实际的方向变化量为 A_t , 其公式

$$\text{若 } \Delta Z_t > 0, \text{ 则 } A_t = 1, \text{ 否则 } A_t = 0 \quad (13)$$

$$\text{若 } \Delta X_t > 0, \text{ 则 } F_t = 1, \text{ 否则 } F_t = 0 \quad (14)$$

式中: ΔZ_t 为时刻 $t-1$ 到 t 之间的变量; ΔX_t 为时刻 $t-1$ 到 t 之间的预测变量。

预测销售数据的转折点

$$F_t = \alpha A_t + \theta_t \quad (15)$$

式中: θ_t 为误差项; α 为这个线性方程的斜率,是一个远远大于零的正数,以表明 F_t 和 A_t 是线性关系。

通过 α 这个系数将这一变化放大。如果实际销售数据 ΔZ_t 的变化方向为增加的趋势,则 A_t 取正数,由式(20)得出销售预测变化趋势 F_t 取正数,则表明实际销售数据的变化也是增加的趋势,这样就可以很容易识别出转折点。因此, α 体现了预测的转折能力。

二、汽车销售量混合预测的仿真分析

通过天津某汽车集团的销售管理系统得到 2004 年 1 月到 2004 年 5 月连续 130 天的销售成交率统计数据,即 $L=130$ 。先用前 100 天数据训练混合预测模型,再用后 30 天数据检验模型预测效果。

按照前面所述步骤,先搭建线性部分预测模型,并最终确定线性模型阶次 $p=2, q=4$, 即 ARMA(2, 4) 及相应的线性预报方程

$$\begin{aligned} \hat{x}_{tL} &= 0.645x_{t-1} - 0.07x_{t-2} + z_t + 0.308z_{t-1} - \\ &= 0.069z_{t-2} - 0.12z_{t-3} + 0.08z_{t-4} \end{aligned} \quad (16)$$

针对线性预测误差构建非线性预测模型即 RBF 神经网络预测模型,希望通过前 S 天的销售成交率预测出第 $S+1$ 天的销售成交率。根据经验,取 RBF 网络输入层节点个数 $S=1$, 输出节点个数 $M=1$, 隐藏层节点个数 $n=10$, 则样本个数 $K=L-(M+S)+1=120$, 前 90 个样本作为学习样本, 学习经过 1 500 次迭代后的平均相对误差为 0.023 7。这样,混合模型各子模型训练后,再利用后 30 天的数据作为检验样本,将线性预测结果和非线性预测结果求和以得到最终预测数据,预测效果拟合曲线见图 2。从图 2 可以看出预测呈现一定的滞后,但总的来看该混合预测方法能够及时地跟踪成交率的变化,尤其是对销售旺盛阶段能够得到准确的预测。

与 ARMA 模型、RBF 模型的预测误差对比见表 1。混合预测方法的预测精度明显优于其他两种方法。

预测转折点的对比结果见表 2, 混合预测方法的倾斜系数 α 值对 t 比率表明,对于在时间序列统计上是不为零,且 α 的符号没有变化,这表明混合模型预测方法具有很好的转折点预测能力;然而 ARMA 预测模型的系数 α 在时间序列中并不明显区别于零。在时间序列上,当历史数据的样本从 100 天减至 50 天时, α 甚至是负的,这体现出 ARMA 模型不能很好预报转折点。可见,在样本数据量不

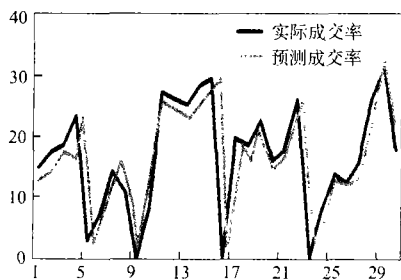


图 2 2004 年 5 月成交率预测值与实际值对照图

是很充足的情况下,采用混合模型预测方法进行预测,可以得到很好的预测效果。

表 1 不同预测方法的预测误差对比

| 误差标准 | | ARMA 预测 | RBF 预测 | 混合预测 |
|------|--------|---------|--------|-------|
| 成交率 | MSE | 0.13 | 0.09 | 0.05 |
| | MAE | 3.27 | 2.33 | 1.78 |
| | MAPE/% | 2.13 | 2.89 | 1.8 |
| 销量 | MSE | 3 426 | 2 137 | 1 809 |
| | MAE | 63 | 54 | 30 |
| | MAPE/% | 0.145 | 0.085 | 0.046 |

表 2 不同预测方法对成交率转折点的预测结果比较
(选取 100 天的样本数据)

| 预测方法 | ARMA | RBF | 混合法 |
|----------|------------|------------|------------|
| α | 0.40(1.97) | 0.43(2.26) | 0.20(1.10) |
| α | 0.46(1.72) | 0.37(1.27) | 0.66(2.76) |

注:括号中的数据为 α 与 t 的比例

总之,利用 ARMA - RBF 混合预测方法可以比较精确地预测出汽车销售情况。仿真结果验证了该

方法的有效性。

三、结 语

本文提出了一种汽车销售量混合预测方法,主要分两个步骤实现。第一步,利用 ARMA 模型来分析销量数据中的线性部分;第二步,采用 RBF 神经网络解决 ARMA 模型的预测误差的建模问题。该方法充分结合了 ARMA 模型和神经网络各自的优势,以实现汽车销量的准确预测。通过仿真试验,表明该混合预测方法预测效果优于单独使用 ARMA 模型或神经网络的预测效果,验证了方法的有效性和优越性。

参考文献:

- [1] Zhang G P. Time series forecasting using a hybrid ARIMA and neural network model [J]. *Neurocomputing*, 2003, 50: 159—175.
- [2] Apostolikas G, Tzafestas S. On-line RBFNN based identification of rapidly time-varying nonlinear systems with optimal structure-adaptation [J]. *Mathematics and Computers in Simulation*, 2003, 60: 1—13.
- [3] Brian H H. Insights into neural-network forecasting of time series corresponding to ARMA (p, q) structures [J]. *Omega*, 2001, 29 (3): 273—289.
- [4] Box G E P, Gwilym M J, Gregory C R. *Time Series Analysis: Forecasting and Control* [M]. New Jersey: Prentice Hall, 1994.
- [5] 张有为. 预测的数学方法 [M]. 北京: 国防工业出版社, 1991.
- [6] Cumby R, Modest D. Testing for market timing ability—A framework for forecast evaluation [J]. *Journal of Financial Economics*, 1987, 19: 169—189.

Hybrid Forecasting Method for Automobile Sale

LI Xiang¹, ZONG Qun², TONG Ling²

(1. School of Management, Tianjin University, Tianjin 300072, China;

2. School of Electrical and Automation Engineering, Tianjin University, Tianjin 300072, China)

Abstract: A hybrid forecasting method for automobile sale based on autoregressive moving average (ARMA) model and radial basis function neural network (RBFNN) is presented. In the first step, ARMA model is applied to analyzing the linear part of the automobile sale volume. In the second step, an RBFNN is developed to model the residuals from the ARMA model. The final forecast value combines both outputs of ARMA model and RBFNN. Through the proposed method, not only the linear structure of the sale data is exposed but also the nonlinear relationship between data is captured. Simulating experiment results with real data sets indicates that the combined model can be an effective way to improve forecasting accuracy achieved by either of the models used separately.

Keywords: hybrid forecasting; ARMA model; RBF neural network; automobile sale forecasting