

基于时间序列分析的自动售货机销量预测

洪 鹏 余世明

(浙江工业大学信息工程学院 杭州 310023)

摘 要 针对自动售货机单一商品售空而频繁加货导致耗费大量人力物力资源的问题,提出用 RBF 神经网络对各产品的销量进行预测,以便制定合理的销售方案来减少这方面费用。考虑到自动售货机的历史销售数据受非最优销售方案制约,从而导致预测结果不理想的特点,建立一个 ARMA 模型预测受制约销量来补偿实际销量以优化预测模型,加快销售方案趋于最优的速度。并将观测数据修正前后的预测结果与实际销量进行对比,验证了该模型在售货机销量预测方面的有效性。

关键词 自动售货机, RBF 神经网络, ARMA

中图法分类号 TP301.4 **文献标识码** A

Vending Machine Sales Forecasting Based on Time Series Analysis

HONG Peng YU Shi-ming

(College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China)

Abstract As vending machine takes a lot of material resources to replenish goods while a single commodity sells out, we developed a reasonable method using RBF neural network for forecasting sales of each product to reduce the costs. Taking the sales data of vending machine maybe constrained by poor sales program into account, we established ARMA model to optimize the actual sales and accelerate the adjustment of the speed of the sales program. We compared before and after correction of the observational data to predict the result with the actual sales to verify the validity of the model in vending machine sales forecasting.

Keywords Vending machine, RBF neural network, ARMA

1 引言

随着自动售货机的销售系统越来越完善,包括支付系统、监控维护系统等技术都趋于成熟,正确的销售决策的制定开始成为售货机厂商关注的热点。传统的销售决策都是凭借商家的长期经验进行决定,而商品的销售受很多因素的影响,如季节、周边人流量以及受众群体的偏好等,因此自动售货机上的商品供求往往得不到合理的平衡,收益达不到最大化。

在售货机销售方案制定的决策分析上,目前不少人对此已有了一定的研究。Tomás 等人^[1]提出由监控系统获取产品的销售信息,通过不断地对销售表现不佳的产品进行替换来达到销售方案的优化。Feng-Cheng Lin 等人^[2]将产品根据其销售业绩进行聚类,并通过决策树提取受欢迎产品的关键属性,最后用贝叶斯网络过滤销售产品。Hidetaka Sakai 等人^[3]通过组合模糊逻辑和多元回归预测模型来动态地修正售货机销量预测的精度,以节约售货机制冷所消耗的能量。

以上这些研究都是基于售货机的历史销售数据,但是由于售货机初始销售方案的良好性未知,产品的实际销量往往受到销售方案的制约。因此统计观测的实际数据只能定性地

反映产品的销售表现的优良,而不能定量地反映实际的销售情况,这使得销售方案的优化速度较为缓慢。并且售货机货架上的产品经常随市场变动,因此需要一个快速的销售方案优化方法。本文为了满足实际需求,提出一种混合 ARMA 和人工神经网络的预测模型。

2 预测方案的设计

自动售货机的主要稀缺资源是货架空间,因此销售方案是否合理的评判标准是售货机货架空间能否合理分配,使得单位空间的收益率最大。文献^[4]证明了最优货架空间分配应具有这样的性质:产品货架空间分配之比等于产品对于零售商的收益之比。

$$\frac{s_1}{s_2} = \frac{(p_1 - w_1)d_1}{(p_2 - w_2)d_2} \quad (1)$$

其中, s 表示产品获得的货架分配空间, p 为产品价格, w 为产品成本, d 为产品的市场需求。根据上述公式,最优销售方案的制定需要价格、成本和市场需求 3 个参数,产品的销售价格及成本比较容易获得,所以本文着眼于通过产品的历史销量来分析产品的市场需求。

根据自动售货机的货道设计以及销售方式等实际情况的

洪 鹏(1990—),男,硕士生,主要研究方向为模型预测控制及嵌入式系统, E-mail: hongpeng1990@126.com; 余世明(1962—),男,博士,教授,博士生导师,主要研究方向为模型预测控制、大系统建模与优化、数字控制理论及其在机电一化系统中的应用。

分析,售货机的销售方案变动不宜过于频繁。因此将销售方案定为以月为更改周期,即商品未来一年每月的销量作为预测对象,选用预测对象的过去几年每月销量作为训练样本和测试样本来训练神经网络,以达到预测的目的。

本文提到的实际销售数据受销售方案制约,即货架上的产品提前售罄的情况,从微观角度看这是一种数据样本的不完备。而缺乏准确的训练样本,神经网络预测模型往往得不到较高的预测精度。由于这种现象只有在每天最后阶段才会出现且并非必然出现,本文没有采用内插和外推法对需要的数据进行估计,而是在出现制约现象时,对售货机当天每隔半小时的销量进行观测,然后对空缺数据进行预测估计。整个预测方案结构图如图1所示。

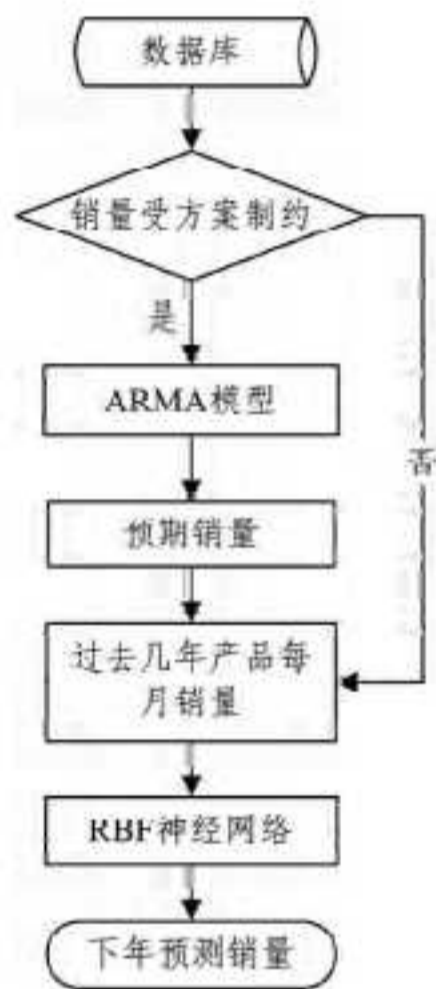


图1 预测过程结构

3 ARMA 模型建立

通过对售货机每日实际销售数据的分析发现在每日 22:00 至次日 7:00 的销售数据数值不大且具有较强的随机性,不具备分析价值。因此售货机在当日 22:00 前出现产品售罄时,判定其销量受非最优方案制约,需要进行对空缺销量进行预测。将前 4 日以及当日 7:00 至达到饱和销量前的销售数据作为样本,对达到饱和销量后至 22:00 的空缺数据进行预测估计。

由于自回归滑动平均模型(ARMA)在短期预测上表现良好并且技术成熟,本文采用该模型拟合样本时间序列曲线,对出现销量受制的情况进行预期销量的预测。样本数据的 ARMA 模型可以描述为:

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (2)$$

其中, p 和 q 分别是自回归部分和滑动平均部分的阶数; ϕ 和 θ 分别是自回归系数和滑动平均系数。

首先求出该观测值序列的样本自相关系数和偏自相关系数,根据这两者表现的拖尾性和截尾性选择阶数适当的 ARMA 模型。然后利用序列的观测值通过极大似然估计来确定该模型中的未知参数值。最后对模型和参数进行显著性检

验,并通过 AIC 准则对最优模型进行选择。建立好合适的模型后,通过预测方差最小原则对产品的预期销量进行预测。

4 RBF 神经网络预测模型

4.1 RBF 神经网络概述

一个有多输入单输出的 RBF 网络结构如图 2 所示,它是由输入层、隐含层和输出层构成的三层前向网络。其中输入层由输入节点组成,输入节点数等于输入向量 x 的维数;隐含层由直接与所有输入节点相连的非线性单元组成,输出层由与隐藏单元相连的线性单元组成。

本文采用 Gauss 分布函数:

$$\varphi(x)_j = \exp\left(-\frac{\|x - c_j\|^2}{2\sigma_j^2}\right) \quad (3)$$

作为隐含层基函数。其中, x 为 n 维输入向量, c_j, σ_j 分别表示高斯函数的中心向量和形状参数, $\|\cdot\|$ 表示欧式范数。输出层神经元采用线性激活函数:

$$f_r(x) = \lambda_0 + \sum_{i=1}^{n_r} \lambda_i \varphi(x) \quad (4)$$

其中, n_r 表示隐层单元个数, λ 表示隐层与输出层间的突触权值。

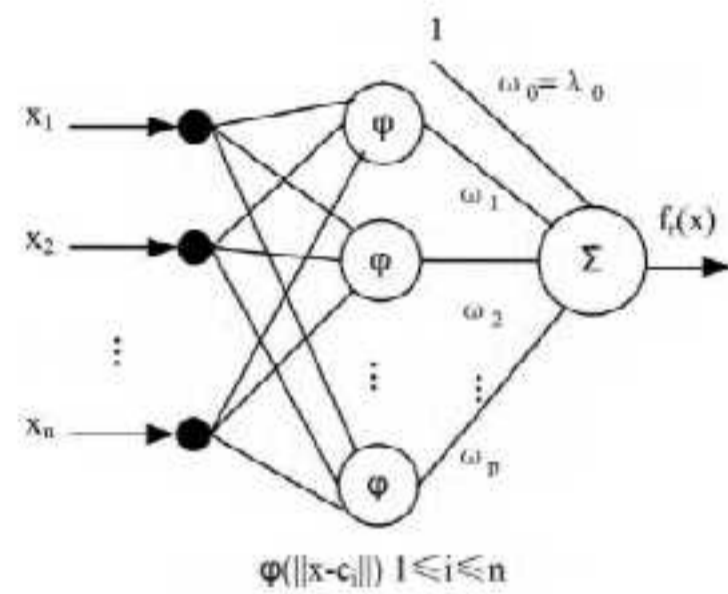


图2 RBF 网络

4.2 RBF 网络预测模型的确定

从本方案例的实际销量数据中可以看到有明显以年为固定周期的季节波动和一定的趋势波动作用于该时间序列,因此将神经网络的输入维数设为 12。为了预测时间序列中的期望输出序列 $y(t) = [y(1), \dots, y(12)]^T$, 令输入矢量为 $x(t) = [y(t-1), \dots, y(t-n)]^T$ 代表过去的 n 个样本,整个网络结构可以表示为 $y(t) = f(y(t-1), y(t-2), \dots, y(t-n))$ 。

在确定神经网络输入维数后,用正交最小二乘法[5]通过保留样本数据中对期望输出影响最大的算子来选取数据中心,然后对 RBF 网络的隐层单元数、基函数中心、扩展常数和输出层权值进行监督学习,得到自动售货机销量的 RBF 神经网络模型。

5 仿真研究

5.1 样本数据的处理

本文选用某地售货机 2008 年至 2013 年瓶装水的历史销售数据作为售货机销量预测的原始数据。表 1 为 2008 年 6 月每天每隔半小时水的实测销量,其中有多次出现提前售罄状态,即前文提到的销售制约情况。

表1 2008年6月每天的销量

日期	7:00 (瓶)	7:30 (瓶)	8:00 (瓶)	...	21:30 (瓶)	22:00 (瓶)
6/1	1	3	1	...	2	0
6/2	2	1	3	...	1	1
...
6/30	0	2	4	...	0	0

因此需要对出现制约现象的实测销售数据进行预期可达销售量的预测,选用前4天及当天可用样本(共146个)进行建模,根据该观测值序列求出的自相关系数和偏自相关系数值表现的性质,选用ARMA(1,1)模型拟合序列,使用极大似然估计得到该模型的口径为:

$$x_t = 0.51x_{t-1} + 0.21\varepsilon_{t-1} + 1.48, \text{Var}(\sigma^2) = 0.75$$

通过最小均方误差理论,对序列 $x(t)$ 的最后4个值进行预测,其预测结果及95%置信区间如表2所列。

表2 6月12日受制约水销量预测

时间	Forecast	Lo 95	Hi 95
20:30	1.7278	0.0109	3.4447
21:00	1.5668	-0.5106	3.6441
21:30	1.4932	-0.6516	3.6281
22:00	1.4597	-0.6990	3.6184

5.2 RBF神经网络预测

本文在Matlab下采用RBF神经网络进行售货机销量预测,以2008至2012年每月水销量作为训练样本,2013年每月销量作为测试样本。由于直接用原始数据对神经网络进行训练,数值大的变量对输出的影响较大,因此对样本数据进行归一化处理,将样本数据变换到同一数量级。同时为避免出现过拟合情况,将训练允许误差设为0.001,对模型进行训练。

本文选用原始数据和通过ARMA模型修正后数据两种样本分别进行神经网络训练,得到两种预测结果。然后将预测结果进行简单处理,对每月总销量按当月天数进行平均,针对售货机货道容量限制,向上取整,制定两种新的销售方案。得到原始数据(方案A)和数据修正后(方案B)两种方案下预测结果的相对误差,如图3所示。

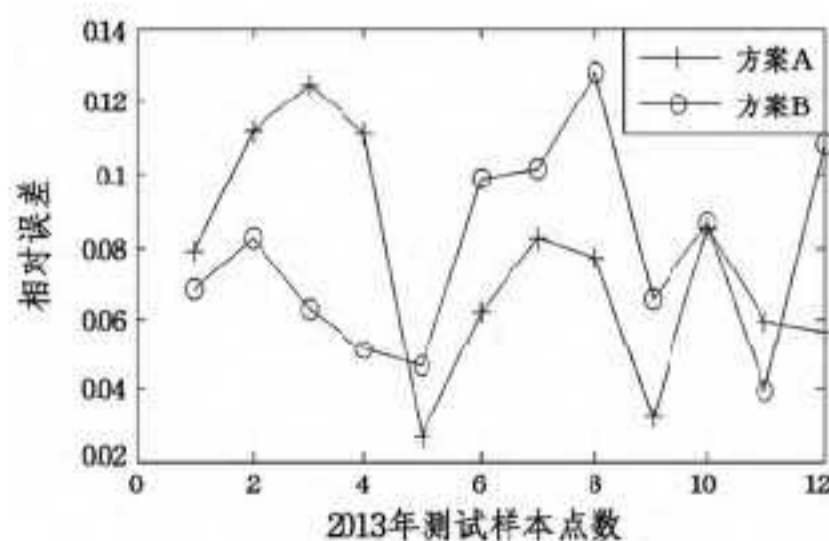


图3 两种方案下预测结果的相对误差曲线

5.3 仿真结果比较

售货机两种方案水销量的预测曲线以及实际销量曲线如图4所示。

从单一产品销售量预测结果来看,方案B对比方案A预测结果的相对误差并没有明显减小,但是方案B设置的产品销售空间比方案A要大。因此对其他产品采用相同方法进行市场需求的预测后,在销售淡季时,对于剩余空间的竞争,

方案B体现更大的作用。剩余空间的竞争通过以下方法:预测各产品的市场需求,减去每天都售完的基础销量,再根据式(1),对利润增量和空间增量求比值,得出各产品的单位空间效益。除去各产品基础销量所占空间后,遵循单位空间效益最高的产品进行销售方案的设计,从而促使空间利益最大化。因此本文认为方案B能给出各产品更加确切的市场需求。

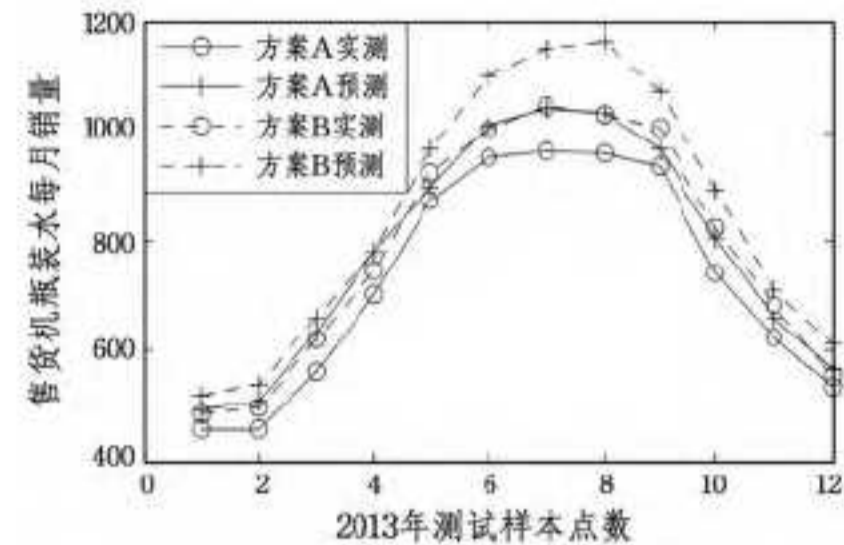


图4 2013年水销量预测及实测曲线

结束语 神经网络已经被广泛地运用于时间序列预测上,但是目前的研究大多集中在提高识别时间序列模型和非线性特性的能力上。他们结合传统时间序列方法和人工神经网络的优势,将混合预测模型用于预测水质[6]、风速[7]、电价[8]等现实生活中极其复杂的时间序列的预测。

本文根据应用对象自身特点,借鉴分层混合预测模型思路,将统计数据分为宏观和微观两部分。在微观部分通过传统的时间序列方案进行空缺数据的预测估计,在宏观部分通过人工神经网络进行销量预测,为合理更改售货机销售方案提供更好的数据支持。

参考文献

- [1] Murray T, Jansson E R. Methods and system for managing vending operations based on wireless data[P]. US, 2010031261A1, 2010-12-9
- [2] Lin Feng-cheng, Yu Hsin-wen, Hsu Chih-hao, et al. Recommendation system for localized products in vending machines[J]. Expert Systems with Applications, 2011, 38(8): 9129-9138
- [3] Sakai H, Nakajima H, Higashihara M, et al. Development of a fuzzy sales forecasting system for vending machines[J]. Computers & Industrial Engineering, 1999, 36: 427-449
- [4] 鲁其辉, 朱道立. 供应链中基于货架空间分配的质量改进策略研究[J]. 管理科学学报, 2010, 13(1): 31-39
- [5] Chen S, Cowan C F N, Grant P M. Orthogonal Least Squares Learning Algorithm for Radial Basis Function Networks[J]. IEEE transactions on neural networks, 1991, 2(2): 302-309
- [6] Faruk D Ö. A hybrid neural network and ARIMA model for water quality time series prediction[J]. Engineering Application of Artificial Intelligence, 2010, 23(4): 586-594
- [7] Erasmo C, Wifrido R. Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA-ANN model[J]. Renewable Energy, 2010, 35(12): 2732-2738
- [8] Dipti S, Zhang G F, Abbas K, et al. Hybrid Neural-Evolutionary Model for Electricity Price Forecasting[C] // Proceeding of International joint conference on neural networks, 2011. San Jose, CA, 2011: 3164-3169