

基于SARIMA-BP神经网络方法的汽车销量预测研究^①

浙江农业商贸职业学院 范庆科

摘要: 对汽车销量时间序列的预测方法进行了研究,针对汽车销量影响因素表现出的多样性,应用灰色关联分析法对其进行筛选。针对汽车销量时间序列表现出的线性特征和环境动态变化导致的非线性规律,单一的线性预测方法和非线性预测方法都无法满足时间序列的预测要求。提出了一种SARIMA-BP神经网络预测方法,利用SARIMA方法对时间序列的线性部分进行建模,利用BP神经网络方法对时间序列的非线性部分进行建模。仿真结果表明, SARIMA-BP神经网络方法比单一模型的预测准确率更高。

关键词: 时间序列 灰色关联分析法 SARIMA-BP神经网络方法

中图分类号: F714

文献标识码: A

文章编号: 2096-0298(2017)08(a)-170-03

随着经济供给侧改革力度的持续加大和宏观政策效应的不断释放,汽车销量逐年稳步增高。如何根据市场需求来准确预测汽车销量成为当前中国汽车产业经济的研究热点。

当前应用于汽车销量的预测方法归纳起来主要有两类:(1)单模型预测,如BP神经网络、SARIMA方法、灰色模型等;(2)基于融合模型的预测,如神经网络和粒子群算法的融合、主成分分析和神经网络的融合等。张健^[1]利用SARIMA模型提取了通货膨胀中的月度频率波动特征,有效降低了预测误差。针对GDP表现出的季节性波动,赵喜仓和周作杰应用SARIMA方法建立季节时间序列模型,提高了预测准确性^[2]。SARIMA方法在样本数据特征提取、序列周期波动抑制、季节性干扰消除等方面得到了广泛应用^[3]。但由于其本质为线性控制,不能较好的拟合数据中的非线性的特征。神经网络控制不依赖所建的模型精度,具有较强的自学习能力和任意非线性函数逼近能力,能较好的抑制环境干扰和非线性引起的误差。曾纳等人利用BP神经网络模型对草地生物量信息和分布格局进行模拟和空间估算,仿真结果与实测值拟合度较高^[4]。皮进修等人提出了一种基于SARIMA-GMDH组合预测方法^[5],应用该方法对我国CPI月度序列进行预测,结果表明对具备优势互补的两种单一模型进行有效组合,大大提高了预测的误差。

针对汽车销量数据中包含的线性特征和非线性特征,本文提出了一种基于SARIMA-BP神经网络的预测方法,利用BP神经网络自学习能力强的特征实现对非线性动态变化的自适应调控,通过神经网络的非线性函数逼近和收敛优势保证了预测的精度和稳定性。

1 汽车销量影响因素的选取

当前影响汽车销量的因素^[6]归纳起来大体有以下几种:(1)社会因素,如人均可支配收入、道路基础设施建设;(2)国家政策,如购置税补贴、限购政策等;(3)价格因素,如燃油价格、汽车售价等;(4)技术因素,如汽车用料质量、制造工艺等。

灰色关联分析法^[7]主要通过对系统特征序列进行数据分析,得出相关因素的曲线形状相似度,依据灰色关联度的大小判断因素的影响程度。针对汽车销量影响因素呈现出多样化且关联性较

强的特点,应用灰色关联分析法对其进行筛选。

汽车销量序列为 $X_0 = (x_0(1), x_0(2), \dots, x_0(n))$,影响因素的特征序列为 $X_i = (x_i(1), x_i(2), \dots, x_i(n))$,汽车销量与影响因素的灰色关联度计算步骤为:

初始化特征关联数据:

$$x_i(k) = \frac{x_i(k)}{x_i(1)} \quad (1)$$

计算特征序列与影响因素序列的指标差异值:

$$\Delta_i(k) = |x_0(k) - x_i(k)| \quad (2)$$

计算关联系数:

$$\gamma(x_0(k), x_i(k)) = \frac{m + \varepsilon M}{\Delta_i(k) + \varepsilon M} = \frac{\min_i \min_k |x_0(k) - x_i(k)| + \varepsilon \max_i \max_k |x_0(k) - x_i(k)|}{|x_0(k) - x_i(k)| + \varepsilon \max_i \max_k |x_0(k) - x_i(k)|} \quad (3)$$

式中, $i = 1, 2, \dots, m$, $k = 1, 2, \dots, n$, $\varepsilon \in (0, 1)$ 。

计算关联度:

$$\gamma(X_0, X_i) = \frac{1}{n} \sum_{k=1}^n \gamma(x_0(k), x_i(k)) \quad (4)$$

通过MATLAB编程求解出各因素对汽车销量的影响关联值,并筛选出关联度较高的6个影响因素,筛选结果如表1所示。

表1 销量影响因素筛选

影响因素	关联度	影响因素	关联度
人均可支配收入	0.9125	购置税补贴	0.8120
整车价格	0.8729	燃油价格	0.7869
汽车材料质量	0.8543	公路里程	0.7523

2 模型与研究方法

2.1 SARIMA模型

ARIMA模型是一种通过差分变换将非平稳时间序列转化为稳定序列的建模方法^[8]。

ARIMA(p, d, q)模型公式为:

$$\varphi_p(B) \nabla^d Y_t = \theta_q(B) \omega_t \quad (5)$$

$$\varphi_p(B) = (1 - \varphi_1 B - \varphi_2 B^2 \dots - \varphi_p B^p) \quad (6)$$

$$\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 \dots - \theta_q B^q) \quad (7)$$

式中, φ_p 为自回归量, p 为自回归阶数, B 为后移量, $\nabla^d = (1 - B)^d$ 为 d 阶差分, Y_t 为时间序列, θ_q 为移动量, q 为自回归阶数, ω_t 为正态分布 $N(0, \sigma^2)$ 白噪声。

针对ARIMA模型在处理带有季节性的时间序列时存在拟合准确率不高的缺陷,函数通过季节差分变化扩展生成

①基金项目: 浙江省教育厅一般科研项目(Y201635219)。

作者简介: 范庆科(1991-),男,浙江江山人,助教,硕士,主要从事非线性控制方面的研究。

表2 模型实验样本数据

(单位: 万辆)

年	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
2014	215.6	159.6	216.9	220.4	191.1	184.6	161.8	171.6	198.4	198.7	209.1	241.0
2015	232.0	159.3	224.1	199.5	190.4	180.3	150.3	166.5	202.5	222.2	250.9	278.6
2016	250.1	158.1	244.1	212.2	209.2	207.1	185.2	207.1	256.4	265.0	293.9	305.7

SARIMA(p, d, q)(P, D, Q)模型,其表达式为:

$$\phi_p(B)\Phi_P(B^S)\nabla^d\nabla_S^D Y_t = \theta_q(B)\Theta_Q(B^S)\omega_t \quad (8)$$

式中, Φ_P 为季节回归量, p 为季节回归阶数, $\nabla_S^D = (1 - B^S)^D$ 为 D 阶季节差分, S 为季节周期, Θ_Q 为季节移动量, Q 为季节移动阶数。

SARIMA模型的构建与预测步骤为:(1)通过季节差分变化将季节性时间序列转化为稳定序列,满足模型的平稳观测要求;(2)提取稳定序列的趋势特征,采用AIC或SBC方法来确定模型的阶数(p, d, q, P, D, Q);(3)参数的估算(有效值、标准误差和统计值)和显著性检验(白噪声值);(4)序列未来值的预估。

2.2 BP神经网络模型

BP神经网络是一种基于神经单元的多层前馈网络,包含输入层、输出层和隐含层,其核心是逆向传播学习过程(输入信号的正向传播和输出信号的反向传播)。BP神经网络原理采用梯度下降法调整权值并更新网络,使误差函数满足迭代要求,完成对样本数据的训练和学习过程。结合汽车销量选取的6个汽车销量影响因素,建立的BP神经网络模型如图1所示。

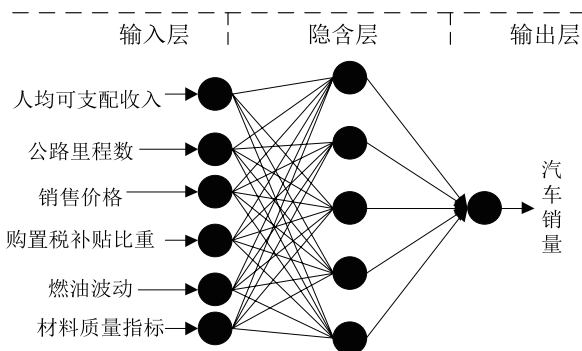


图1 BP神经网络模型

BP神经网络模型中,定义各层神经元节点为 net 节点数分别为 i, j, k , 即输入层输入为 μ_i , 隐含层神经元输入为 net_j , 输出为 o_j , 输出层神经元输入为 net_k , 输出为 y_k , 其表达式分别为:

$$net_j = \sum \alpha_{ij} \mu_i + \beta_j \quad (9)$$

$$o_j = \gamma(net_j) = \gamma(\sum \alpha_{ij} \mu_i + \beta_j) \quad (10)$$

$$net_k = \sum \alpha_{jk} o_j + \beta_k \quad (11)$$

$$y_k = \tau(net_k) = \tau(\sum \alpha_{jk} o_j + \beta_k) \quad (12)$$

式中, $\gamma(\cdot)$ 和 $\tau(\cdot)$ 分别为隐含层和输出层的激活函数 α_{ij}, β_j 和 α_{jk}, β_k 分别为输入层和输出层的权值和阈值。权值和阈值的迭代公式为:

$$\Delta \alpha_{ij(k)} = \sigma \rho_{j(k)} o_{i(j)} \quad (13)$$

$$\Delta \beta_{j(k)} = \sigma \rho_{j(k)} \quad (14)$$

式中, σ 为学习速率, $\sigma \in (0, 1)$, $\rho_{j(k)}$ 为单元层误差函数。

2.3 SARIMA-BP神经网络模型

时间序列可分解为线性序列和非线性序列两部分,将汽车销量的特征序列进行以中心、4为周期的二阶滤波后,线性部分 L_t 和非线性部分 S_t 表达式为:

$$L_t = \frac{1}{4} \left(\sum_{i=1}^3 x(t+i) + \frac{1}{2}(t+2) \right) \quad (15)$$

$$S_t = x_0(t) - L_t \quad (16)$$

式中, $t = 4(i-1) + j$, i 为年份, $n = 4i$, j 为季度。

SARIMA-BP神经网络模型预测步骤为:(1)利用SARIMA模型预测方法拟合线性部分 L_t , 并计算 S_t , 预测值记为 \tilde{L}_t ; (2)提取非线性部分 S_t 信息特征构建BP神经网络模型, 记预测值为 \tilde{S}_t , BP神经网络的参数设置如图2所示;(3)SARIMA-BP方法预测值为 \tilde{L}_t 和 \tilde{S}_t 之和, 记为 \tilde{X}_t 。

2.4 方法评价体系

方法评价体系包含采用均方根误差RMSE、平均绝对百分比误差MAPE和平均绝对误差MAE, 其表达式分别为:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (L_t - \tilde{L}_t)^2} \quad (17)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{L_t - \tilde{L}_t}{L_t} \right| \times 100\% \quad (18)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |L_t - \tilde{L}_t| \quad (19)$$

3 模型比较与结果分析

以2014年~2016年度的汽车月销量构成时间序列, 共计36组实验样本。随机从样本数据中筛选出30组数据, 分别使用SARIMA模型、BP神经网络模型对其进行模拟。表1给出了实验样本数据(单位: 万辆), 图2给出了各模型对采用数据的预测效果。

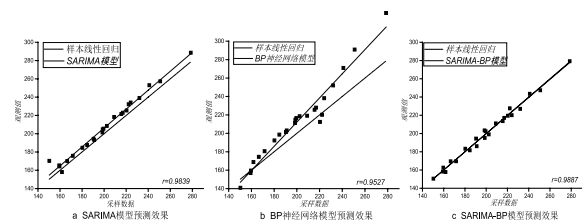


图2 不同模型样本数据预测效果图

表3 模型预测方法性能比较

	RMSE	MAPE	MAE
SARIMA 模型	0.487	3.25%	0.288
BP 神经网络模型	0.469	2.73%	0.253
SARIMA-BP 模型	0.221	0.96%	0.094

由图2可知, SARIMA模型和BP神经网络模型对样本数据的

关于大型企业财务共享服务中心运行优化的几点思考

中国电信股份有限公司山东分公司 高雷

摘要: 本文结合大型企业财务共享服务中心进入稳定运行期后的实际情况,深入分析报账事项的真实性风险、次级单位财务指标关注度不足、会计核算效率持续提升受限、系统集成规则亟待优化、服务职能缺乏拓展等方面的问题,提出建立风险预警机制、模拟账套细化核算、改良报账均衡性、考核激励提质增效、会计核算智慧化、内部有偿服务等针对性解决措施,以期能使财务共享服务中心在风险管控和效益提升方面发挥更好的作用。

关键词: 大型企业 财务共享 运行优化

中图分类号: F275

文献标识码: A

文章编号: 2096-0298(2017)08(a)-172-02

1 财务共享服务中心的运行情况

在过去的几十年里,财务共享服务中心模式已经被90%的世界500强公司接受和采用。近年来,为了适应全球化发展,提高企业竞争力,我国越来越多的企业开始借鉴国际先进管理模式建立财务共享服务中心,从职能、组织、流程、系统方面进行了积极的探索和实践,加强对下属公司的财务管控,降低财务管理成本。财务共享服务中心模式在财务运营成本、会计核算标准、财务控制能力、流程风险管控方面取得了成效,对大型企业增强内部活力、提升运行效率起到积极的推动作用。

随着财务共享服务中心运行的日趋成熟,共享服务模式的一些风险点逐渐显现,面对国有企业改革的大背景,新形势下战略转型、提升效益的要求,大型企业更加需要认识到财务共享服务中心运行中未能突破的瓶颈和问题,把握风险管控和提升效率两大重点,探讨可行的研究对策,以期更好的发挥集约财务核算的优势。

2 运行中显现的问题

2.1 财务与业务部门的分隔导致报账真实性存在风险

实行财务共享服务之后,会计核算人员集中到财务共享服务

中心,远离市场趋势和业务情况,对经济业务的实质情况缺乏直接了解,对业务操作过程中的突发事项主要依靠与前端业务部门沟通取得的间接信息,降低了财务信息的可靠性与财务决策及时性。比如,财务共享服务中心的记账过程是分散派单到不同的会计进行核算记账的,如果报账部门为了规避大额项目审批而拆分项目分别报账,此时,如果流转至财务共享服务中心后分散至不同的会计,按照独立的报账项目核算,就可能会产生经营的内控风险。

2.2 合并账套后对次级利润中心的财务指标关注度不够

某些大型企业引进财务共享服务后,实施的是集约财务核算模式,其将省级公司本部和地市分公司的账套合并,财务核算和资金支付集中处理,全省作为一个会计主体,各分支机构是利润中心组。这种报表模式下,在全省报表数据存在异常时比较容易对次级利润中心的数据有所分析和关注,但在全省报表数据正常的情况下,可能会忽略分支机构的财务指标异动。另外,对于不同利润中心之间的往来账项,从报表角度进行调整和清理的驱动力不足,容易形成长账龄未清账项,影响会计核算质量。

2.3 会计核算效率的持续提升出现瓶颈

预测值整体高于样本实际值,SARIMA模型预测精度比BP神经网络模型高,这是由于SARIMA模型对带有季节性的时间序列具有较高的辨识能力。SARIMA-BP神经网络模型具备较高的样本数据拟合能力,SARIMA模型和BP神经网络模型对样本数据的预测效果总体低于SARIMA-BP神经网络模型。SARIMA模型、BP神经网络模型和SARIMA-BP神经网络模型的线性拟合度分别为0.9539、0.9527和0.9557,SARIMA-BP神经网络方法的预测误差最小。由表3可知,构建的预测方法的RMSE误差、MAPE误差和MAE误差均大幅度低于SARIMA方法和BP神经网络方法,证明所构建的SARIMA-BP神经网络方法的预测效果较好。

4 结语

在汽车销量预测中,由于汽车销量时间序列表现出的影响因素多样性、非线性和线性规律,本文通过灰色关联分析法提取关联度较高的销量影响因素,融合SARIMA模型和BP神经网络模型,提出一种SARIMA-BP神经网络预测方法,应用该方法对汽车销量时间序列数据进行建模及预测。试验结果表明,所建模型比单一模型的预测准确率更高,验证了所提出预测方法的可行性。

参考文献

- [1] 张健.SARIMA模型在预测中国CPI中的应用[J].统计与决策,2011(05).
- [2] 赵喜仓,周作杰.基于SARIMA模型的我国季度GDP时间序列分析与预测[J].统计与决策,2010(22).
- [3] 姜春雷,张树清,张策.等.基于SARIMA-BP神经网络组合方法的MODIS叶面积指数时间序列建模与预测[J].光谱学与光谱分析,2017(01).
- [4] 曾纳,任小丽,何洪林,等.基于神经网络的三江源区草地上生物量估算[J].环境科学研究,2017(01).
- [5] 皮进修,赵清俊,彭建文.基于SARIMA-GMDH的CPI组合预测模型[J].统计与决策,2016(17).
- [6] 尹小平,王艳秀.中国汽车销量影响因素的实证分析[J].统计与决策,2011(08).
- [7] 董少阳,夏清国,李宁.基于灰色关联分析法的软件缺陷类型预测[J].计算机工程与应用,2013(02).
- [8] 杨峰,薛斌,刘剑.基于非平稳时序ARIMA模型的W频段雨衰预测[J].电子与信息学报,2015(10).