

基于组合模型的农产品物价预测算法^①



苏照军^{1,2}, 郭锐锋², 高 岑², 王美吉², 李冬梅²

¹(中国科学院大学 计算机控制与工程学院, 北京 100049)

²(中国科学院 沈阳计算技术研究所, 沈阳 110168)

通讯作者: 苏照军, E-mail: 718280975@qq.com

摘 要: 当今时代, 科学技术高速发展, 涌现出一批新技术, 数据挖掘、机器学习等新科学领域被深入研究, 众多智能算法逐渐出现, 同时被应用到了不同的领域中. 本文构建了一种基于 BP (Back Propagation) 神经网络和 SVR (Support Vector Regression) 支持向量回归机的组合模型. 依托于农产品价格数据进行实例验证分析, 结果表明相对于单一的预测模型, BP-SVR-BP 组合模型在预测精度上有了很大的提升, 拟合效果更加逼近真实数据曲线, 能够客观真实的反应农产品物价变化规律.

关键词: 组合模型; BP 神经网络; 物价预测; SVR 预测; 农产品

引用格式: 苏照军, 郭锐锋, 高岑, 王美吉, 李冬梅. 基于组合模型的农产品物价预测算法. 计算机系统应用, 2019, 28(5): 185-189. <http://www.c-s-a.org.cn/1003-3254/6910.html>

Agricultural Product Price Forecasting Algorithm Based on Combination Model

SU Zhao-Jun^{1,2}, GUO Rui-Feng², GAO Cen², WANG Mei-Ji², LI Dong-Mei²

¹(School of Computer and Control Engineering, University of Chinese Academy of Sciences, Beijing 100049, China)

²(Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Shenyang 110168, China)

Abstract: Nowadays, with the rapid development of science and technology, a number of new technologies have emerged. New scientific fields such as data mining and machine learning have been deeply studied. Many intelligent algorithms have emerged and applied to different fields. This paper constructs a combined model based on BP (Back Propagation) neural network and SVR (Support Vector Regression). Based on the agricultural product price data, the example verification analysis shows that compared with the single prediction model, the BP-SVR-BP prediction model has greatly improved the prediction accuracy. The fitting effect is closer to the real data curve, which can objectively and truly reflect the law of agricultural product price changes.

Key words: combined model; BP neural network; price forecasting; SVR forecasting; agricultural products

1 引言

我国是一个农业大国, 农业在国民经济中占据着重要地位, 农业是否平稳健康发展, 关系着社会的稳定. 所以, 对农产品物价进行预测具有迫切的现实意义.

之前人们大多采用单一模型进行预测, 近年来人们更多的是采用组合模型进行预测. 陈咏梅^[1]等使用基于 ARIMA 和 BP 神经网络^[2]的组合模型对福建滨海旅

游市场规模进行预测, 表明该组合模型的预测准确率高, 更加接近真实情况. 朱叶^[3]等使用灰色模型和 BP 神经网络构建的组合模型对兵团农产品物流市场需求进行预测, 表明组合模型在预测精度上有了很大的提升. 刘金源^[4]等使用基于 BP 神经网络、GM 灰色预测理论和 PCA 主成分分析的组合模型对城市燃气日负荷进行预测, 表明该组合模型^[5,6]是一种更为有效

① 收稿时间: 2018-12-04; 修改时间: 2018-12-26; 采用时间: 2018-12-28; csa 在线出版时间: 2019-05-01

的预测方法. 组合模型展现出了更大的优势. 本文使用粒子群和动态学习率优化 BP 神经网络模型, 结合支持向量机模型对农产品物价进行预测, 将两组预测结果使用优化的 BP 神经网络进行组合得到预测结果, 提高模型的泛化能力和预测准确度.

2 BP-SVR-BP 组合模型

2.1 组合模型组合方式

组合模型是利用两种或者两种以上的模型, 分别对实际问题进行模型训练, 将结果使用权重进行组合, 得到最终结果的过程. 组合模型目标就是利用一组最优权重系数, 对单独模型结果进行加权求值, 使得结果

更加接近真实值.

目前, 单一预测模型准确率不高, 考虑因素单一, 没有结合模型的适用场景, 所以本文研究了组合模型. 由于 BP 神经网络能够更好地模拟线性和非线性数据, 适合于长期预测模型的训练; SVR 是基于时间序列的回归预测模型, 适合于短期预测. 将适合中长期预测的 BP 模型和适合于短期预测的 SVR 模型组合, 使模型具有更高的准确率, 更强的泛化能力. 所以选用了优化后的 BP 神经网络模型和 SVR 模型进行组合. 提出了基于 BP-SVR-BP 的组合模型算法, 即 BP 神经网络-支持向量机回归-BP 神经网络的组合模型, 并且使用农产品价格数据对模型进行验证评估.

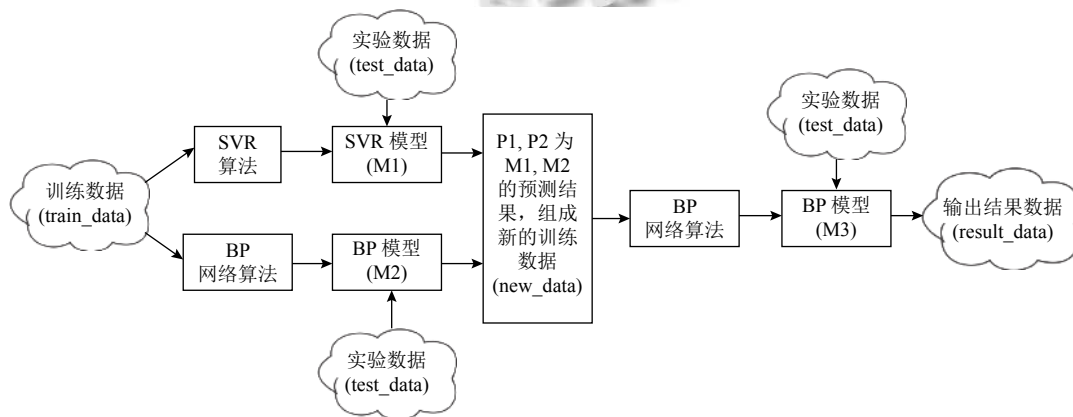


图 1 BP-SVR-BP 组合模型

图 1 为 BP-SVR-BP 组合模型的结构图. 第一步将经过数据预处理的训练集数据 (train_data) 输入到 SVR 算法中, 经过训练, 得到 SVR 模型 (M1); 同时将训练集数据输入到第一轮的 BP 神经网络算法中, 经过训练, 得到 BP 神经网络模型 (M2). 第二步将实验数据或者称之为测试数据 (test_data) 输入到 SVR 和 BP 模型中, 得到两组实验结果数据 (P1, P2), 将其组成两列数据, 并且将其对应的真实数据 (Price) 加入, 作为第三列, 形如 (P1, P2, Price) 组成第二轮 BP 神经网络算法的训练集数据 (new_data). 第三步将新的训练集数据 (new_data) 输入 BP 神经网络算法, 这一步主要是利用 BP 拟合组合模型的权重系数, 寻找最优的权重系数组合. 经过训练, 得到第二轮的 BP 模型 (M3). 第四步将实验数据 (test_data1) 输入模型 (M3), 得到预测结果, 评价组合模型的预测准确度.

在组合模型中, 第一轮我使用了优化的 BP 神经网络

模型和基于多项式核函数的 SVR 模型, 第二轮使用了优化的 BP 神经网络模型训练组合模型权重系数.

2.2 优化 BP 神经网络

本文采用三层结构, 即输入层、隐含层、输出层. 假设输入层 n 个节点, 隐含层 k 个节点, 输出层 1 个节点. $X(X_1, X_2, \dots, X_n)$ 为输入层输入数据; $M(M_1, M_2, \dots, M_n)$ 为经输入层加权求和的结果, 即隐含层的输入数据; $N(N_1, N_2, \dots, N_k)$ 为隐含层由激活函数求得的结果, 即隐含层的输出数据; Z 为经隐含层加权求和的结果, 即输出层的输入数据; Y 为预测值, W 为各层链接的权值, $B(B_1, B_2, \dots, B_l)$ 表示偏置值.

根据输入 X , 得隐含层的输入为:

$$M_j = \sum_{i=1}^n W_{ij}^{(l)} X_i + B_i^{(l)} \quad (1)$$

其中, l 为层数, b 为偏置值. 则隐含层的输出为:

$$N_j = \psi_1(M_j) = \psi_1 \left(\sum_{i=1}^n W_{ij}^{(l)} X_i + B_j^{(l)} \right) \quad (2)$$

其中, ψ_1 为隐含层 Sigmoid 激活函数, $j=1, 2, \dots, n$.

输出层的输入为:

$$Z = \sum_{j=1}^k W_{j1}^{(l)} N_j + B_j^{(l)} \quad (3)$$

输出层的输出为:

$$Y = \psi_2(Z) = \psi_2 \left(\sum_{j=1}^k W_{j1}^{(l)} N_j + B_j^{(l)} \right) \quad (4)$$

由于 BP 神经网络的权值和阈值对价格的预测准确度有一定影响, 可以通过粒子群算法对网络中的权值和阈值进行全局寻优, 合理优化调整. 假设给定 n 个粒子:

$p_i = (p_{i1}, p_{i2}, \dots, p_{iD})$ 粒子 i 的位置 $x_i = (x_{i1}, x_{i2}, \dots, x_{iD})$; 粒子 i 的速度 $v_i = (v_{i1}, v_{i2}, \dots, v_{iD})$; 粒子 i 经过的历史最优位置; 粒子群整体经历的最优位置: $g = (g_1, g_2, \dots, g_D)$; 则 PSO 速度更新公式为:

$$v_{iD}^{k+1} = \omega v_{iD}^k + c_1 r_1 (p_{iD}^k - x_{iD}^k) + c_2 r_2 (p_{gD}^k - x_{iD}^k) \quad (5)$$

位置更新为:

$$x_{iD}^{k+1} = x_{iD}^k + v_{iD}^{k+1} \quad (6)$$

k 为: 第 k 次迭代; c_1, c_2 为学习率变量, r_1, r_2 为 $[0, 1]$ 范围内的随机数, ω 为粒子群权重变量.

假设输入层、隐含层、输出层节点数为 $L1_Num, L2_Num, L3_Num$. 种群参数初始化为: 种群数目 N ; 维度为 $D=L1_Num \times L2_Num + L2_Num \times L3_Num$; 使用真实值与预测值的均方误差作为粒子群适应度函数:

$$e = \frac{1}{n} \sum_{j=1}^n (O_j - Y_j)^2 \quad (7)$$

其中, O_j 为第 j 个记录的真实值, Y_j 为第 j 个记录的预测值. 得到的最优位置数据, 分配到网络中训练.

BP 神经网络的学习率对网络的收敛速度有影响. η 过大, 收敛速度过快; η 过小, 导致收敛速度过慢. 因此需采用一种动态改变的学习速率 η .

$$\eta(k) = \begin{cases} 1.05\eta(k-1), & E(k+1) < E(k) \\ 0.75\eta(k-1), & E(k+1) > E(k) \\ \eta(k), & E(k+1) = E(k) \end{cases} \quad (8)$$

2.3 支持向量回归机 (SVR) 模型

支持向量回归机 (SVR) 模型是在 SVM 基础上, 加

以改进, 可以用于基于回归的预测模型. 模型主要是通过核函数将数据映射到一个高维空间, 然后在高维空间寻找一个最优的超平面, 使所有点到该平面的距离误差最小. 这个函数即是 SVR 函数模型. SVR 函数为:

$$f(X) = W^T \phi(X) + b \quad (9)$$

其中, W 为权值向量; $\phi(X)$ 为映射函数; b 为常数. 在此基础增加松弛度变量 ξ_i 和 $\hat{\xi}_i$, 则式子可转化为:

$$\min_{w, \xi_i, \hat{\xi}_i} \left(\frac{1}{2} \|w\|^2 + C \sum_{i=1}^m (\xi_i + \hat{\xi}_i) \right) \quad (10)$$

同时必须满足:

$$\begin{cases} f(x_i) - y_i + b \leq \varepsilon + \xi_i \\ y_i - f(x_i) - b \leq \varepsilon + \hat{\xi}_i \\ \xi_i \geq 0, \hat{\xi}_i \geq 0, i = 1, 2, \dots, m \end{cases} \quad (11)$$

3 训练组合模型

3.1 数据集

数据选用 2009-2016 年上海农产品日价格数据作为实验数据.

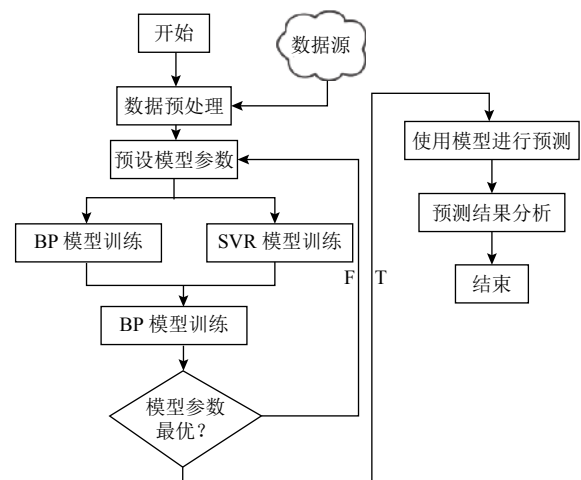


图2 构建模型流程图

3.2 数据预处理

选用数据的平均交易价格作为预测目标, 数据发布时间、最低交易价格、最高交易价格、最高温度、最低温度作为模型输入数据, 训练模型.

(1) 将数据发布时间转换为 YYMMDD 的时间格式, 例如 2016/6/11 转换为 20160611.

(2) 采用前后 k 个数据的平均值填充缺失值.

(3) 采用 MIN-MAX 方法对特征数据归一化.

3.3 组合模型训练

(1) 预处理之后的数据, 选用集合数据 (train_data) 的 70% 作为训练数据 (Train1), train_data 的 30% 的数据作为测试数据 (Test1). 将 Train1 数据分别输入到 SVR 算法和 BP 算法中训练出 SVR 模型 (M1) 和 BP 模型 (M2). 将 Test1 输入到 M1 和 M2 模型中, 得到两组预测结果 (P1, P2), 和真实数据 (Price) 组成三列数据 (train_data2), 形如 (P1, P2, Price). 将 train_data2 的 80% 作为训练集 (Train2), train_data2 的 20% 作为测试集 (Test2), Train2 训练集训练 BP 模型, 得到最优权重系数组合, 得到模型 M3. 将 Test2 输入 M3 模型得到预测结果值.

(2) 对 BP 神经网络训练, 选择最优隐含层节点数目. 使用均方根误差 (RMSE) 衡量真实值和预测值之间的准确率, 如图 3.

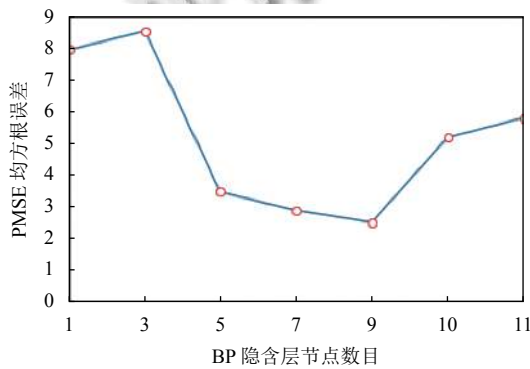


图 3 隐藏层节点数-RMSE 图

根据图 3, BP 隐含层节点数目为 9.

(3) SVR 模型, 选用多项式核函数 (Poly).

4 实验结果与分析

4.1 与单一模型对比分析

本文基于 BP-SVR-BP 组合模型, 第一个 BP 网络模型使用了 5 个节点的输入层, 9 个节点的隐含层, 1 个节点的输出层, 网络权值和阈值使用粒子群算法进行全局优化, 加以自适应动态调整学习率; SVR 选用多项式核函数进行训练; 第二个 BP 神经网络使用 2 个节点的输入层, 5 个节点的隐含层, 1 个节点的输出层. 预测结果如图 4.

图 4 描绘了时间与价格预测值的折线图, 从图中可以得出 BP 神经网络, SVR 模型, BP-SVR-BP 模型,

总体趋势相近, 但 BP, SVR 局部存在较大差距, BP-SVR-BP 组合模型局部更符合价格数据趋势. 将折线图放大, 如图 5 所示.

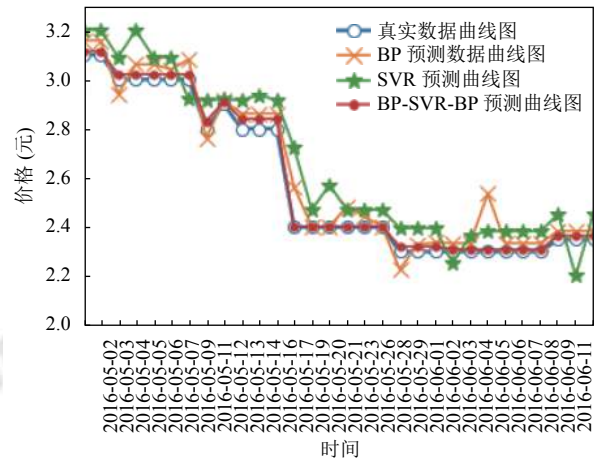


图 4 时间-价格图

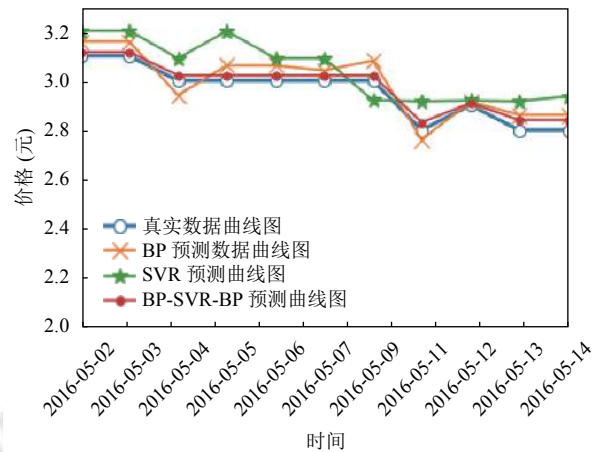


图 5 时间-价格图

从图 5 中, 可以得出, SVR 模型预测的值离真实值曲线最远, BP 神经网络次之, BP-SVR-BP 模型最近, 说明 BP-SVR-BP 模型的预测准确度最高, 所以 BP-SVR-BP 预测准确率上优于单一的 BP, SVR 模型.

为了更加准确地评价三个预测模型, 采用量化体系进行评价. 采用 MAE, MSE, RMSE 进行量化评价.

测得的结果见表 1.

表 1 单一模型对比

模型	MAE(%)	MSE(%)	RMSE(%)
BP 模型	3.3815	0.1504	3.87799
SVR 模型	9.0615	0.8417	9.17422
BP-SVR-BP 模型	1.5744	0.0380	1.94813

从表1中,可以从量化的角度分析得出, BP-SVR-BP模型, MAE, MSE, RMSE 的值都比 BP 和 SVR 模型小, 基于以上综合分析, 得出 BP-SVR-BP 组合预测模型对农产品价格曲线拟合效果更优, 预测准确率更高。

4.2 与组合模型对比分析

前人的研究中, 也尝试了很多的组合模型预测算法。比如使用 BP 神经网络模型和 ARIMA 模型进行组合, 首先通过 ARIMA 模型预测结果值, 接着使用预测值作为参数, 输入 BP 神经网络预测, 得出预测结果, 即为 BP-ARIMA 模型。也有采用 BP 神经网络模型和灰色模型 GM(1, 1) 进行组合, 将结果采用标准差确定组合模型组合权重, 将单一模型结果组合, 得出预测结果, 即为 BP-GM(1, 1) 组合模型。现将提到的预测方法用在农产品价格预测上, 对组合模型进行对比分析, 如图6。

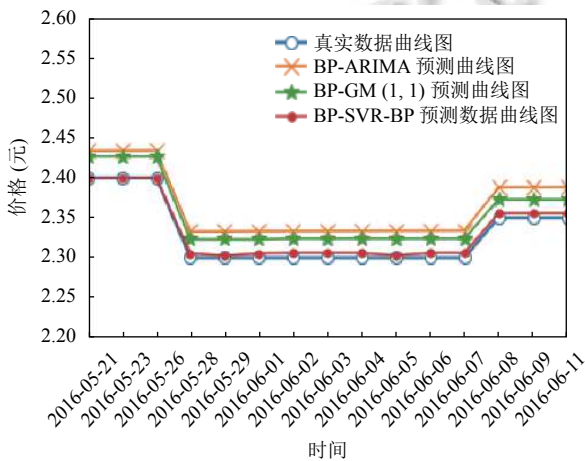


图6 组合模型时间-价格图

图6分别描绘了 BP-SVR-BP、BP-ARIMA、BP-GM(1, 1) 组合模型在农产品价格中的预测结果, 从图中可以清晰地看出, 三种组合模型对农产品价格曲线模拟较好, 都能够比较准确地模拟出农产品价格走势, 和真实农产品价格走势相同, 但是从图6中也可以得出, 三种模型虽然走势相同, 但是准确率上却有差别, BP-SVR-BP 模型描绘的曲线相对于 BP-ARIMA、BP-GM(1, 1) 模型描绘的曲线, 更加逼近真实数据曲线, 即为准确率更高。采用量化体系分析结果如表2。

从表2中, 可以从量化的角度分析, BP-SVR-BP 组合模型的 MAE, MSE, RMSE 值都比 BP-ARIMA、BP-GM(1, 1) 组合模型的小, 所以可以得出 BP-SVR-BP 组合模型的准确率更好, 泛化能力更强, 能够更好地实现

农产品物价的预测。

表2 组合模型对比分析

组合模型	MAE(%)	MSE(%)	RMSE(%)
BP-ARIMA 模型	3.6445	0.1407	3.7518
BP-GM(1, 1 模型)	2.8694	0.0881	2.9671
BP-SVR-BP 模型	1.5744	0.0380	1.94813

5 结语

本文提出了基于 BP-SVR-BP 的组合模型, 使用图分析和数据量化分析, 分析了 BP 模型、SVR 模型、BP-SVR-BP 组合模型, 得到组合模型准确率更高, 拟合效果更优。紧接着对常见的组合模型进行了对比分析, 对 BP-SVR-BP 组合模型、BP-ARIMA 组合模型、BP-GM(1, 1) 组合模型三种组合模型进行了图分析和数据量化分析, 得出 BP-SVR-BP 组合模型相对于 BP-ARIMA 和 BP-GM(1, 1) 组合模型准确率更好, 泛化能力更强, 能够较好地实现农产品物价的预测。本方法还有继续需要改进的地方, 比如说, 将对农产品价格产生影响的更多影响因素纳入模型训练中, 对 SVR 核函数进行优化, 对组合模型组合预测结果的加权值进行优化等等。未来将继续对组合模型进行深入研究, 深化改进模型, 得到更优组合。

参考文献

- 1 陈咏梅. 基于 BP-ARIMA 组合模型的福建滨海旅游市场规模预测. 厦门理工学院学报, 2018, 26(4): 22-26.
- 2 马健, 盛魁, 董辉. 基于改进型误差函数的复值 BP 神经网络对股票价格的预测研究. 长春师范大学学报, 2018, 37(6): 85-90.
- 3 朱叶, 喻晓玲. 基于组合模型的兵团农产品物流市场需求预测. 塔里木大学学报, 2016, 28(2): 52-60. [doi: 10.3969/j.issn.1009-0568.2016.02.010]
- 4 刘金源, 王寿喜, 李婵. 城市燃气日负荷 PCA-GM-BPNN 组合预测模型. 天然气与石油, 2018, 36(5): 13-19. [doi: 10.3969/j.issn.1006-5539.2018.05.003]
- 5 吴叶, 刘婷婷, 方少勇. 基于 MIV-GA-BP 神经网络的我国棉价预测研究. 棉纺织技术, 2018, 46(7): 77-80. [doi: 10.3969/j.issn.1001-7415.2018.07.020]
- 6 江粉桃. 非线性方法的时间序列组合模型在农产品价格预测中的应用研究[硕士学位论文]. 兰州: 兰州交通大学, 2018.