

Price Estimation of Pseudo-Ginseng Based on ARIMA Model and BP Neural Network

Jianzhong Liu^{1*}, Huichao Li²

¹Tianjin Municipality State-Owned Assets Supervision and Administration Commission, Tianjin

²School of Computer and Control Engineering, Nankai University, Tianjin

Email: *1422624555@qq.com

Received: Jul. 12th, 2017; accepted: Jul. 24th, 2017; published: Jul. 27th, 2017

Abstract

In this paper, ARIMA and BP neural networks are used to analyze and forecast the price of Pseudo-ginseng. Firstly, the traditional time series analysis model—autoregressive moving average model is used to predict the price of Pseudo-ginseng, and it is found that it can only extract the linear characteristics of the price curve, and the effect of outlier fitting is poor. BP neural network is applied to solving the nonlinear separable problem. Combining with the fluctuation and non-linearity of the price data of Pseudo-ginseng, the BP neural network is used to analyze and forecast the price of Pseudo-ginseng, and the phase space reconstruction method is used to determine the network input dimension, to determine the network structure better. This paper analyzes the price data from January 2010 to December 2015 in Wenshan area, Yunnan Province, and forecasts the price of January to March in 2016 through the trained model. The results show that the optimized BP neural network by the method of phase space reconstruction is better than the traditional ARIMA model.

Keywords

Pseudo-Ginseng Price Forecast, ARIMA Model, BP Neural Network

基于ARIMA模型与BP神经网络的三七价格预测

刘建中^{1*}, 李慧超²

¹天津市国资委, 天津

²南开大学计算机与控制工程学院, 天津

Email: *1422624555@qq.com

收稿日期: 2017年7月12日; 录用日期: 2017年7月24日; 发布日期: 2017年7月27日

*通讯作者。

摘要

本文分别选取了ARIMA、BP神经网络方法对三七价格进行分析预测。首先,采用传统时间序列分析模型——自回归移动平均模型对三七价格进行预测,发现其仅能提取价格变化曲线中的线性特征,对于离群点拟合效果不佳。BP神经网络适用于非线性可分问题求解,结合三七价格数据的波动性和非线性的特征,使用BP神经网络对三七价格进行分析预测,并利用相空间重构方法确定网络输入维数,加快网络结构确定。本文以云南文山地区2010年1月至2015年12月每天的价格数据为实例进行分析,并通过训练得到的模型对2016年1月至3月三七价格进行预测,结果表明利用相空间重构方法优化的BP神经网络预测的结果优于传统的ARIMA模型。

关键词

三七价格预测, ARIMA模型, BP神经网络

Copyright © 2017 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近年来,中药材价格呈现普遍上涨的态势,三七作为中药材中的大宗品种,其价格波动更显剧烈。2010年三七120头的年平均单价在310元左右,最高价为510元,2011年相较2010年价格有小幅的上涨但波动不大。而到了2012与2013年120头三七最高价猛长至800元左右,2015年则跌至155元,三七价格波动的剧烈程度可见一斑。

药材价格波动严重干扰了市场的稳定,也给种植三七的药农和以三七为元材料的药企带来巨大的不确定性。因此,运用相关数学模型对药材价格的周期性和波动性进行分析,对未来三七价格走势进行预测,为药农种植和药企采买提供决策支持是很有意义的。

2. 三七价格特性分析

2.1. 三七价格上涨及波动的定性分析

影响三七价格上涨和波动的原因众多,主要因素包括需求、成本以及投机行为三个方面。

2.1.1. 需求拉动三七价格上涨

三七是治疗心血管疾病药物的主要成分,其指向性明显,具有其他药物无法替代的功效[1]。随着我国心脑血管疾病的发病群体的不断加大,对以三七为原料的治疗心脑血管疾病药物的需求也呈递增态势,特别是三七总皂苷、三七血栓通、复方丹参片和复方丹参滴丸等药品在市场中日渐走俏。因此,在需求拉动下,三七价格不断地上涨。

2.1.2. 成本推动三七价格上涨

中药材种植成本包括种子种苗价格、劳动力成本、土地成本、交通成本等。自2009年10月以来,三七种苗从过去的0.1元/株,涨到2013年的1元/株;三七种子的价格也从2009年10月之前的几十元/kg,

涨到 2009~2014 年最高时, 达到 1000 元/kg。另外, 三七对地理位置、光、热、水、土、气候等条件要求较高, 我国适合种植三七的地区有限, 随着三七价格上涨, 药农跟风种植三七, 对土地需求增加, 而适宜种植三七的土地有限, 导致地租上涨, 三七种植成本进一步增加。

2.1.3. 投机因素加剧三七价格的上涨及波动

中药材具有投机品特性, 各大药企囤货助推了三七价格的上涨。三七作为治疗心脑血管疾病药物的主要原材料, 其市场需求比较稳定, 因此, 投机资本是导致三七价格波动极大的偏离价值曲线的重要因素。

2.2. 三七价格波动定量分析

2.2.1. 自然因素对价格的影响

1) 种植面积对价格影响分析

从 2010 年到 2013 年, 三七的价格经历了历史上最大的涨幅, 从 100 元一直涨到 800 元。图 1 为 2006~2013 年云南省三七种植面积变化折线图, 图 2 为 2006~2014 年云南地区三七产量变化折线图。综合从网上搜集的大量资料可知, 此次涨价的诱因主要是云南省 2009 年到 2011 年间较低的采挖面积和连年大旱, 另外药企和经销商囤积药材也加剧了此番价格增长。

2) 降水量对价格的影响

三七生长发育期要求比较湿润的环境[2], 查看有关三七的新闻和快讯可知每年有关“干旱”的内容集中在每年的第一季度, 新栽种三七同时也在三月份左右出苗, 因此我们观察一下 1 到 4 月的降雨量。图 3 是文山州附近月值降水量统计。

直观上看旱情较严重的三年的共同特点是 2 月和 3 月的降水量偏低, 4 月份降水量比较充足, 一般是旱情解除的月份; 而同时尽管 2011 年 2 月的降水量极低, 但是 3 月极高的降水量在一定程度上缓解了旱情, 所以 2011 年的三七报道并未关注旱情。综合来看图中红色着重的年份为旱情较为严重的 3 年, 同时也是三七价格飙升的三年, 因此降雨量对三七价格具有显著影响。

3) 寒潮对价格的影响

三七产区的地貌复杂, 地势起伏大, 造成气温的垂直差异显著, 冬春季及夏末秋初, 受大气环流冷空气侵袭的影响, 造成短时性及局部的霜冻和低温灾害, 特别是低温冷害, 气温骤然降低至作物生长的临界值以下造成的危害较大, 这类情况几乎隔数年就可能出现[2] [3]。图 4 在日值价格曲线上用红圈标出了 2010~2016 年间云南地区 5 次较大的暴雪和寒潮的发生日期。

观察可知, 只有前两次寒潮之后价格曲线上升, 而后三次寒潮之后价格并没有改变甚至第 3 次寒潮后还下降。另外第一次寒潮过后价格的上升并没有超过其附近的价格波动区间, 可以认为寒潮对这次上升的贡献几乎可以忽略不计。进一步观察可知在第一次寒潮过后 11 天左右价格发生上升, 在第二次寒潮过后 17 天左右价格发生的上升, 中间间隔了很长的时间, 综上所述我们认为寒潮并不会抬升三七的价格。

在气候因素中, 目前只发现干旱这种长时间大面积气候现象会对三七的价格整体走势产生影响。另外, 降雨量和采挖面积是影响价格整体走势的两个重要因素。在整体走势的预测中, 降雨量可通过天气预报提前估计, 而今年的采挖面积也大概可以通过去年的二年生三七在地面积普查估计。

2.2.2. 三七价格波动规律分析

影响三七价格波动的外界因素众多, 并且各自之间存在较强的相关性。然而, 各因素之间的相关关系、对波动的影响程度最终都体现在价格上, 因此本文从历史价格走势来对未来价格进行预测。

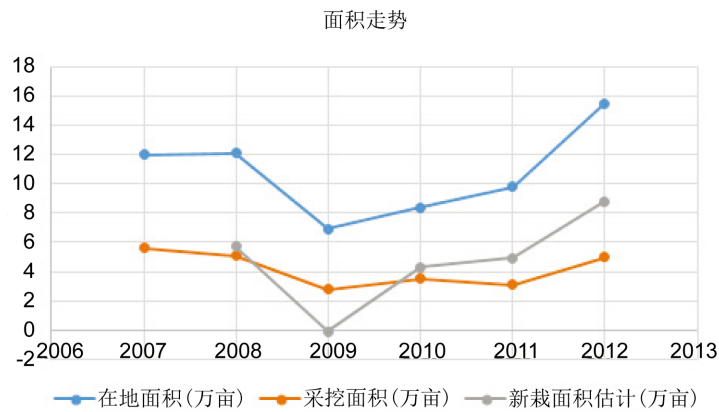


Figure 1. 2006-2013 Yunnan Province pseudo-ginseng planting area change line chart

图 1. 2006~2013 年云南省三七种植面积变化折线图

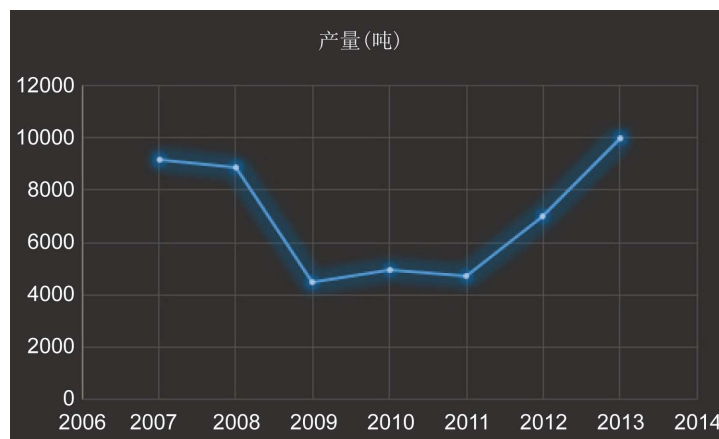


Figure 2. 2006-2014 Yunnan region pseudo-ginseng years production change line chart

图 2. 2006~2014 年云南地区三七产量变化折线图

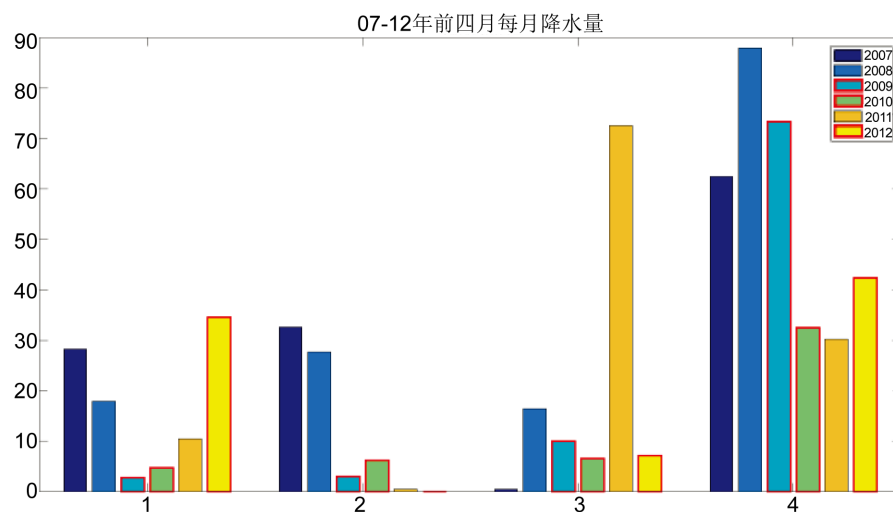


Figure 3. 2007-2012 in Yunnan 4 months before the annual precipitation histogram

图 3. 2007~2012 年云南每年前 4 个月降水量柱状图

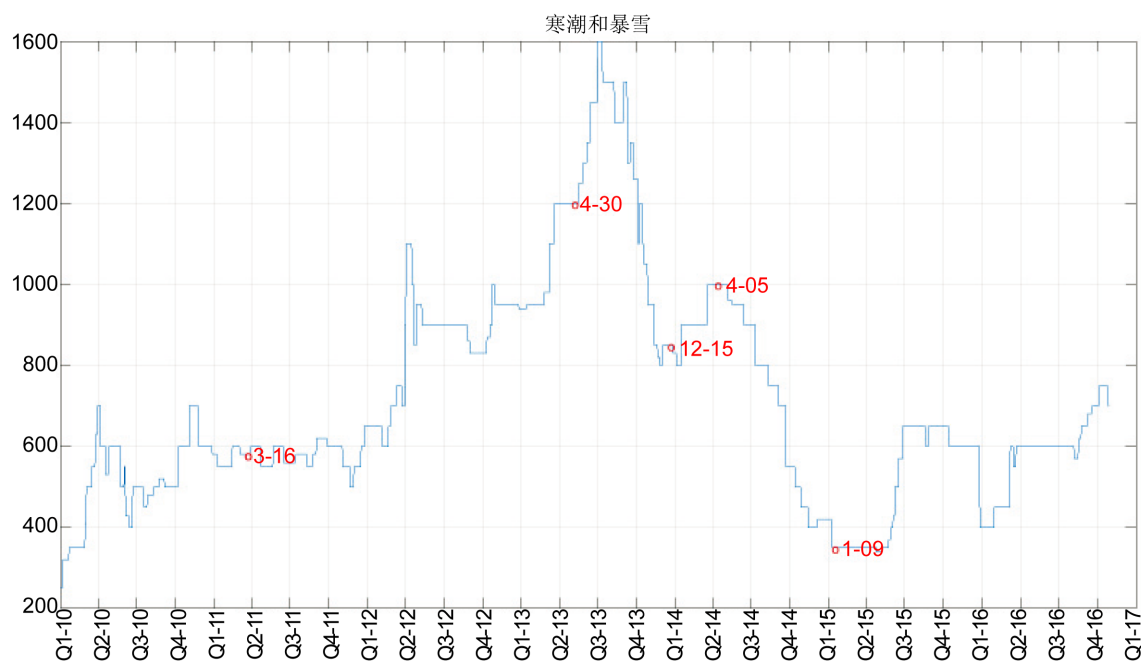


Figure 4. 2010-2016 years of cold wave in Yunnan

图 4. 2010~2016 年间云南地区寒潮记录

1) 三七价格周期性变化规律

图 5 为 2010 年至 2016 年每年 12 个月份 120 头三七的月平均价格。我们将 2010 年与 2014 年、2011 年与 2015 年、2012 年与 2016 年的价格曲线图进行对比发现：虽然整体价格高低有差距，但每组中两条曲线的走势大致相同。

为了更直观的观察，本文给出了将每组的价格曲线对比图，分别如图 6、图 7 和图 8 所示。

2010 年与 2014 年的对比图中可以看出，2010 年与 2014 年的 120 头三七价格都是在第 12 周左右开始上涨，在第 21 周左右出现下跌，又在 43 周左右小幅回升。

2011 年与 2015 年进行对比发现，这两年三七价格都相对平稳，并且都在年末时出现上扬趋势。

2012 年与 2016 年价格曲线进行对比，这两年的价格走势也是大致相同。

由以上三组价格曲线图的对比情况，我们可以初步确定，三七价格波动周期大概为四年，而这与三七的生长周期正好吻合。三七的生长周期长达 3 年，当市场上价格出现下跌时，再减少三七的种植面积已经来不及了，当年产的三七 3 年前就已经种植在地里，三七价格表现在市场上会有一些的滞后，因此确定三七的价格波动周期为四年左右。

2) 三七价格季节性变化规律

图 9 为 2010 年至 2016 年每年四个季度三七平均价格折线图。从图中我们可以清晰的看出：每年第一季度末第二季度初和第三季度末第四季度初，三七的价格会出现大幅度上涨，波动较为剧烈。

在统计学中，方差反映了数据的波动情况。为了说明以上价格波动情况的猜想的正确性，通过绘制价格方差折线图来进行验证。如图 10 所示，三七价格方差在每年的 2、3 月份和 9、10 月份较大，说明三七价格在这四个月波动明显。

三七价格的季节性波动情况也正好符合三七的采收情况。在交易市场中，三七分为冬三七和春三七。冬三七在每年春节过后进行采挖，3、4 月份上市；春三七为每年 8、9 月采挖。春节过后，随着药企购货计划的开启和新三七的上市，三七的购销热度上升，价格也随之上涨。

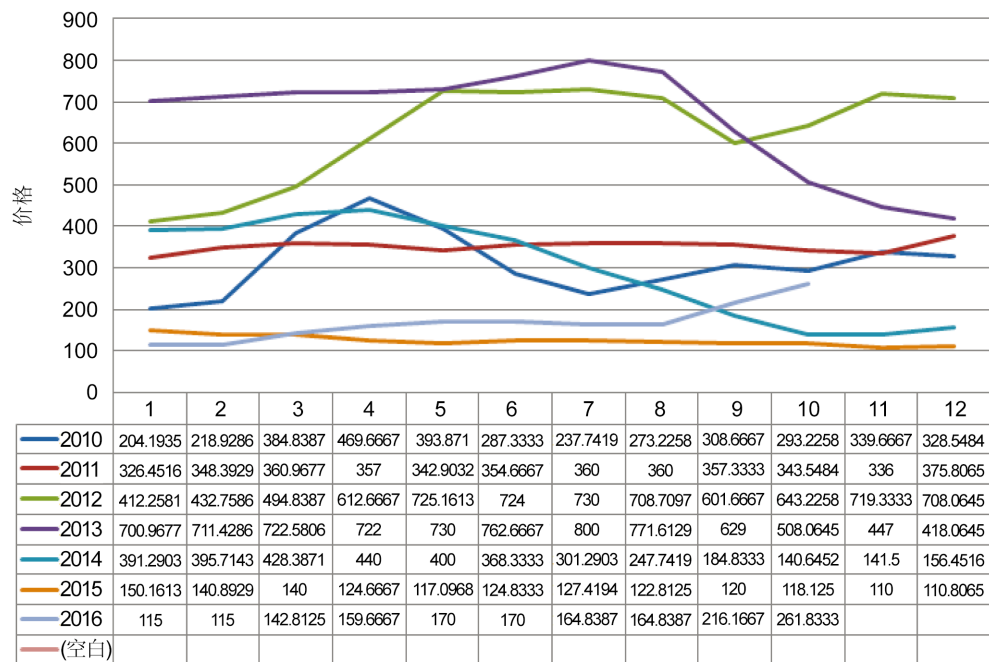


Figure 5. 2010-2016 120 price of pseudo-ginseng (monthly average) trend

图 5. 2010 年至 2016 年 120 头三七价格(月均价)走势

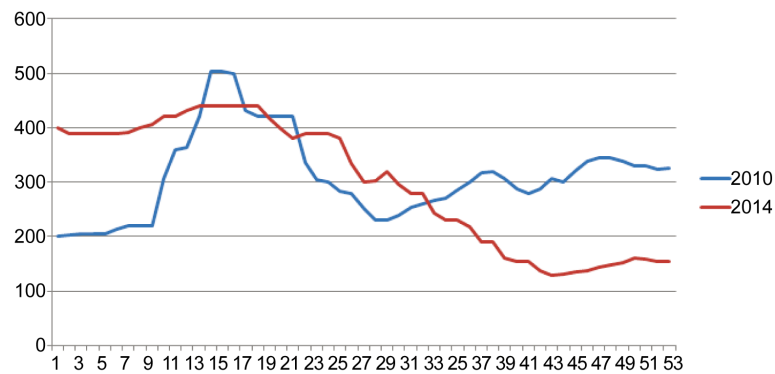


Figure 6. Price comparison between 2010 and 2014

图 6. 2010 与 2014 年价格曲线对比

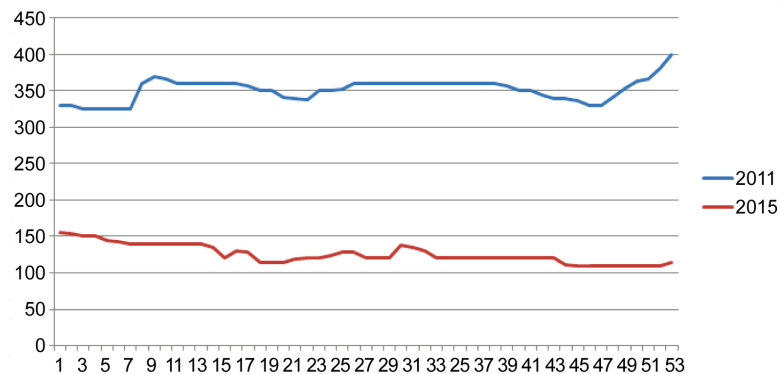


Figure 7. 2011 and 2015 price curve comparison

图 7. 2011 与 2015 年价格曲线对比

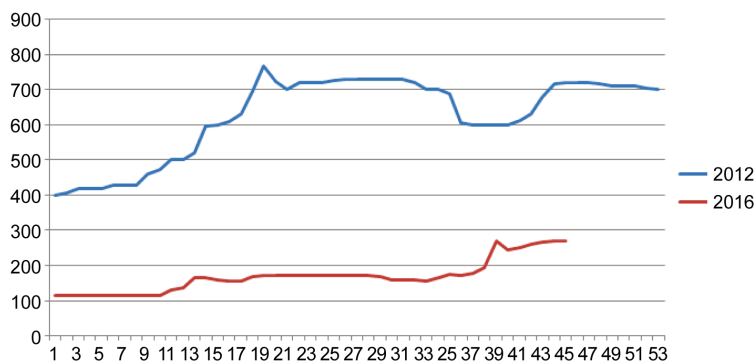


Figure 8. 2012 and 2016 price comparison curve
图 8. 2012 与 2016 年价格曲线对比

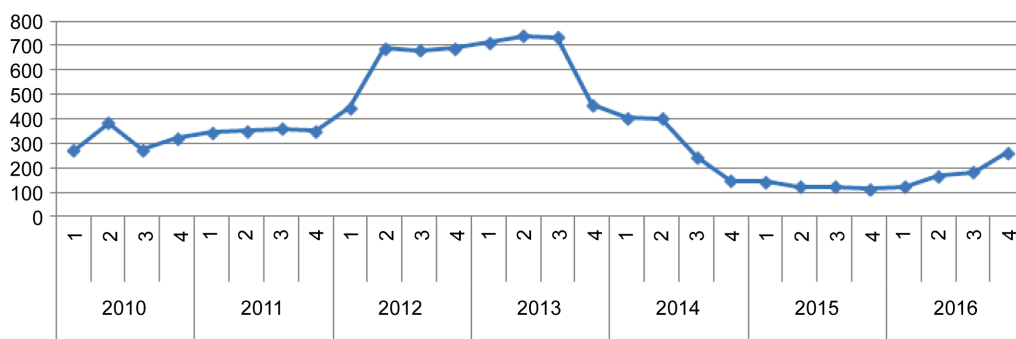


Figure 9. The average price chart of pseudo-ginseng
图 9. 120 头三七季度平均价格走势图

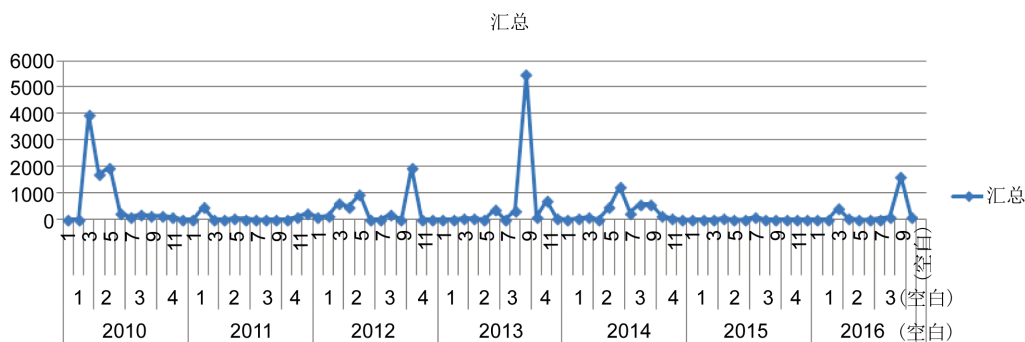


Figure 10. Line chart of price variance
图 10. 价格方差折线图

3. 基于 ARIMA 的三七价格预测

3.1. ARIMA 模型

通过三七价格数据分析, 三七价格曲线具有时序性、周期性和季节性的特点, 因此三七价格预测属于典型的时间序列分析问题。

自回归移动平均(ARMA)模型是传统的时间序列分析模型, 其广泛应用于股票、期货等时间序列分析[4]。ARIMA 模型以 AR 模型和 MA 模型为基础“混合”构成, 包含了自回归和移动平均 2 种成分[5]。

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t + \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + \dots + \varphi_q \varepsilon_{t-q}$$

模型记作 $ARMA(p,d,q)$, 其中 d 代表对原序列进行差分的阶数, p 代表自回归成分的阶数, q 代表移动平均成分的阶数; $\phi_1, \phi_2, \dots, \phi_{1-p}$ 是自回归系数, 描述了 y_{t-j} ($j=1, 2, \dots, p$) 对 y_t 的影响程度; $\varphi_1, \varphi_2, \dots, \varphi_q$ 为移动平均系数, ε_t 是随机干扰误差项[6]。

3.2. 数据预处理

1) 时间维度确定

实验分析所用数据为 2010 年 1 月至 2016 年 3 月每天的三七价格, 为了提高计算精度和效率, 需要首先确定价格频域(时间维度)。

2010 至 2016 年间, 2010 年的价格小幅波动最为频繁, 因此选取 2010 年的价格数据作为分析对象, 来确定最大的有效时间维度。

图 11 为 2010 年三七日价格曲线, 从中可以看出, 一周 7 天内的价格无明显差异; 而一个月中各周间的波动则较为明显, 因此选取前 1 个月的数据来预测下月里每周的平均价格。

2) 数据归一化

不同年代经济水平不同使得价格基数不同, 为了消除这种影响需要对数据进行归一化处理。另外, 进行归一化处理后的数据可以消除不同量纲之间的影响, 同时加快程序运行时的收敛速度。

3.3. 数据集介绍

实验所用数据集为 2010 年 1 月至 2016 年 10 月每周的平均价格, 其中 2010 年至 2015 年的数据作为训练集, 2016 年 1 月至 3 月的数据作为测试集。每一样本包括每月五周的周均价, 即: $X = (P_1, P_2, P_3, P_4, P_5)$ 。2010 年至 2015 年共 5 年, 60 条样本, 因此测试集为 60×5 的矩阵。如表 1 所示为 2010 年的样本资料阵。

3.4. 模型建立

通过使用 EViews 软件对云南文山地区 2010 年 1 月至 2015 年 12 月每周的价格进行分析和预测。

1) 序列平稳性检验

时间序列分析过程中, 为防止回归分析中存在伪回归, 不能够真实的反映因变量和解释变量之间存在的均衡关系, 因此需要对序列进行单位根检验, 以确定序列的平稳性。通过对 2010 年 1 月至 2015 年 12 月每周的三七价格进行 ADF 检验, 图 12 为 ADF 检验图。从图中可以看出 ADF 值为 -0.88426 均大于显著性水平分别为 1%, 5%, 10% 时的临界值, 因此该序列不稳定。

2) 差分平稳化

对原始序列进行差分变换可以将不平稳序列转换成平稳序列。实验中对原序列进行一阶差分, 再次对差分后的序列进行单位根检验, 得到一阶差分序列单位根检验结果如图 13 所示, 其 ADF 值为 -21.12798 , 均小于显著性水平为 1%, 5%, 10% 的临界值, 因此可确定差分后的序列为平稳序列, 且 $d=1$ 。

3) 模型定阶与数据预测

模型定阶过程主要是为了确定 p 、 q 两个参数, 通过逐步试探方法最终确定模型 $ARIMA(2,1,2)$ 为最佳模型。利用 $ARIMA(2,1,2)$ 模型对 2010 年 1 月至 2015 年 12 月的价格进行拟合预测, 从图 14 可以看出模型拟合偏差为 0, 方差为 0.000024, 所以 ARIMA 对训练数据的拟合效果较好。

使用所得模型对 2016 年 1 月至 3 月的价格进行预测, 预测值与实际值对比如图 15 所示。

从图 15 中可以看出 ARIMA 模型对 2016 年前两个月的三七价格预测比较准确, 但 3 月份数据偏差较大。综上所述, ARIMA 模型能较好的预测短期数据的趋势, 但对于非线性变化规律预测效果较差。

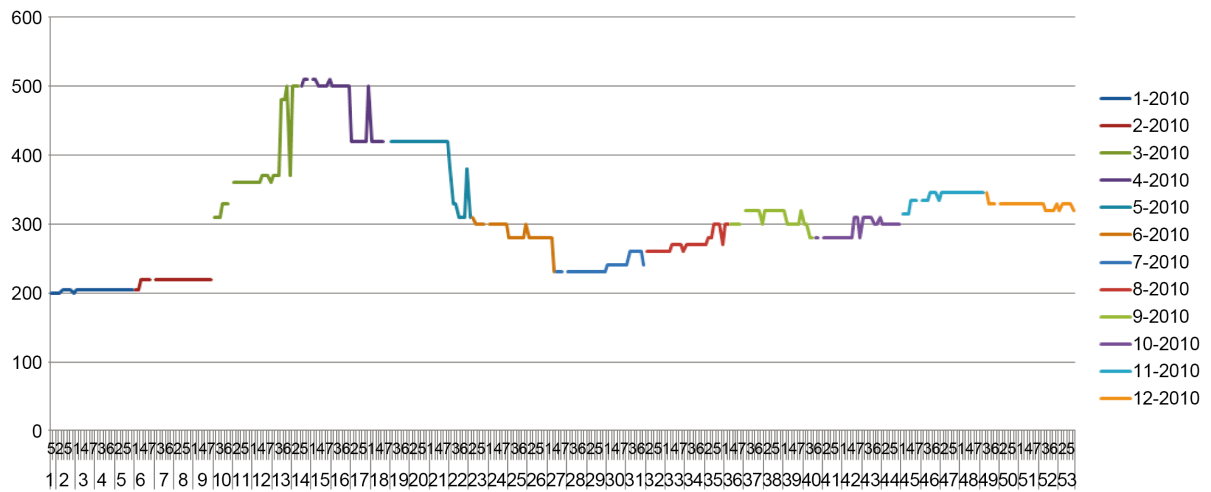


Figure 11. Price fluctuation curve of pseudo-ginseng in the year of 2010

图 11. 2010 年三周日价格波动曲线

Null Hypothesis: SER01 has a unit root
 Exogenous: Constant
 Lag Length: 1 (Automatic - based on SIC, maxlag = 15)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-0.884260	0.7925
Test critical values:		
1% level	-3.451214	
5% level	-2.870621	
10% level	-2.571679	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(SER01)
 Method: Least Squares
 Date: 01/26/17 Time:23:02
 Sample (adjusted): 1/15/2010 12/25/2015
 Included observations: 311 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
SER01(-1)	-0.007459	0.008435	-0.884260	0.3772
D(SER01(-1))	-0.177637	0.056162	-3.162947	0.0017
C	2.722532	3.842710	0.708493	0.4792
R-squared	0.035524	Mean dependent var		-0.266422
Adjusted R-squared	0.029262	S.D. dependent var		30.89936
S.E. of regression	30.44392	Akaike info criterion		9.649249
Sum squared resid	285464.4	Schwarz criterion		9.715325
Log likelihood	-1502.123	Hannan-Quinn criter.		9.693669
F-statistic	5.672258	Durbin-Watson stat		1.988495
Prob(F-statistic)	0.003809			

Figure 12. Initial sequence unit root detection results

图 12. 初始序列单位根检测结果图

Null Hypothesis: D(SER01) has a unit root
 Exogenous: Constant
 Lag Length: 0 (Automatic - based on SIC, maxlag = 15)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-21.12798	0.0000
Test critical values:		
1% level	-3.451214	
5% level	-2.870621	
10% level	-2.571679	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(SER01,2)
 Method: Least Squares
 Date: 01/26/17 Time:23:05
 Sample (adjusted): 1/15/2010 12/25/2015
 Included observations: 311 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(SER01(-1))	-1.181865	0.055938	-21.12798	0.0000
C	-0.313204	1.725767	-0.181487	0.8561
R-squared	0.590941	Mean dependent var	-0.009187	
Adjusted R-squared	0.589617	S.D. dependent var	47.50644	
S.E. of regression	30.43318	Akaike info criterion	9.675354	
Sum squared resid	286189.1	Schwarz criterion	9.699404	
Log likelihood	-1502.518	Hannan-Quinn criter.	9.684967	
F-statistic	446.3914	Durbin-Watson stat	1.989658	
Prob(F-statistic)	0.000000			

Figure 13. First-order differential unit root test

图 13. 一阶差分单位根检验

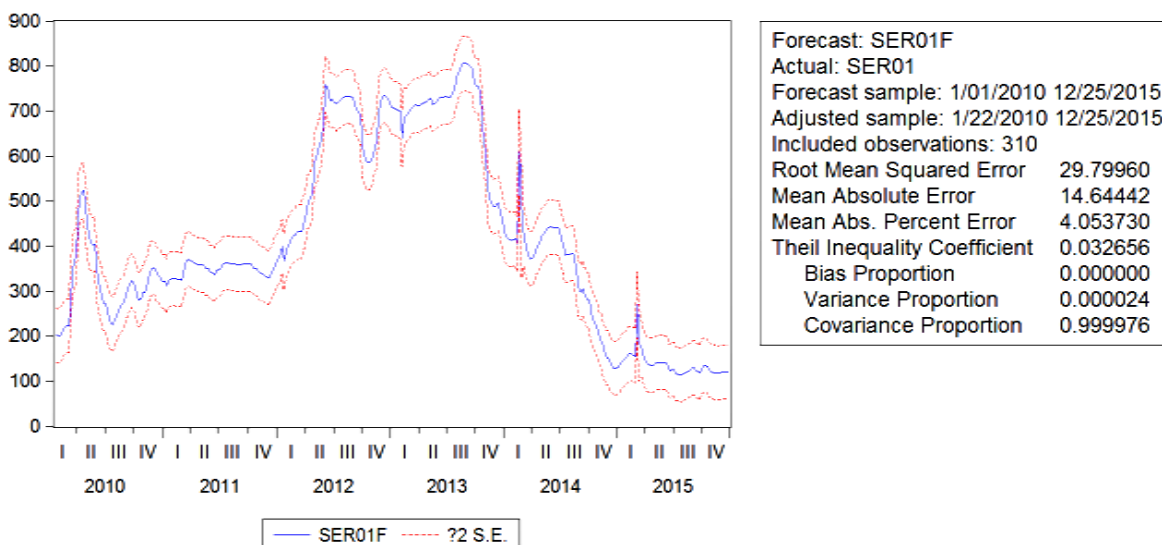


Figure 14. Model prediction results

图 14. 模型预测结果图

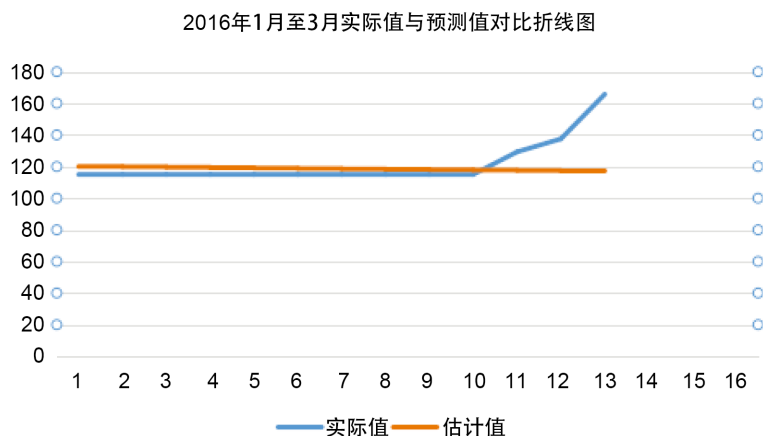


Figure 15. Comparison of the mean and predicted values of the monthly average for January to March 2016

图 15. 2016 年 1 月至 3 月周均值实际值与预测值对比图

Table 1. Weekly price of pseudo-ginseng sample data matrix in 2010

表 1. 2010 年 120 头三七周价格样本资料阵

	第一周	第二周	第三周	第四周	第五周
2010~1	200	202.8571429	205	205	205
2010~2	215	220	220	220	220
2010~3	320	360	364.2857143	420	500
2010~4	506.6666667	504.2857143	500	431.4285714	420
2010~5	420	420	420	420	335.7142857
2010~6	302	300	282.8571429	280	267.5
2010~7	230	230	230	240	254.2857143
2010~8	260	265.7142857	270	285.7142857	300
2010~9	300	317.1428571	320	305.7142857	292
2010~10	280	280	288.5714286	307.1428571	300
2010~11	325	339.2857143	345	345	345
2010~12	333.75	330	330	324.2857143	326.6666667

4. 基于 BP 神经网络的三七价格预测

ARIMA 模型虽然对短期数据变化趋势具有较好的预测效果,但其对非线性变化因素的感知能力较差 [7],对价格波动曲线中的离群点预测效果不佳。而 BP 神经网络属于多层感知器,适用于求解非线性可分问题。为解决这个问题,引入 BP 神经网络算法。

4.1. BP 网络预测模型

要解决非线性可分问题,需要使用具有多层功能神经元的神经网络模型。本实验选用“多层前馈神经网络”作为预测模型,使用误差反向传播算法进行参数调整。

如图 16 所示为“M-P 神经元模型”,神经元接收来自 n 个其他神经元传递过来的输入信号,并通过带权重的连接进行传递,通过“激活函数”处理以产生神经元的输出。多隐层前馈神经网络模型,输入层神经元接收外界输入,隐层与输入层神经元对信号进行加工,最终结果由输出层神经元输出。

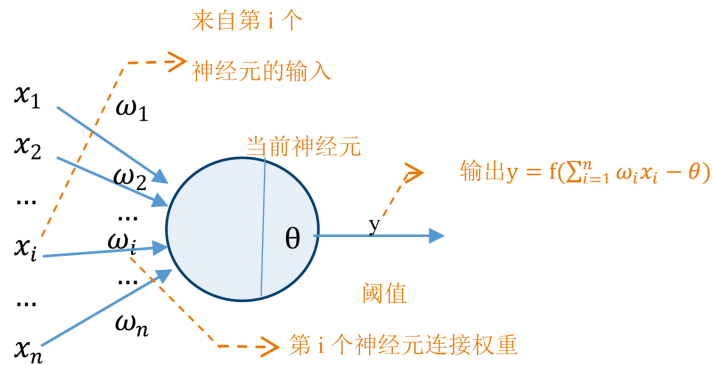


Figure 16. M-P neuron model
图 16. M-P 神经元模型

实验所用数据集为 2010 年 1 月至 2016 年 10 月每周的平均价格, 训练集、测试集、检验集按 70%、15%、15% 进行随机划分。由于周数据的数据量较少(360 组), 根据 Haykin 提出的经验性公式[8]决定使用 10 周的数据来预测下一周的数据。因此, BP 网络输入层包含 10 个节点, 输出层为一个节点。

网络层次结构确定之后, 需要设置层与层之间信息传递所用传递函数。根据实际交易经验, 将初始学习率设为 0.1, 期望达到的最小错误率为 4×10^{-5} ; 另外, 具体实验表明当迭代次数达到 10,000 次之后, 模型错误率趋于稳定, 因此迭代次数设为 10,000 次。网络结构参数设定如表 2 所示。

在其他条件不变的情况下, 分别用 *logsig* 和 *tansig* 作为激活函数对整理后的数据进行测试, 选取预测结果误差较小的 *tansig* 函数为输入层与隐含层的传递函数; 通过控制变量法在保证其他条件不变只改变隐层神经元个数的情况下发现隐含层节点为 9 个时预测误差较小。输出层传递函数一般选用 *purelin* 函数, 以保持输出的范围[9]。

实验结果如图 17 所示, 神经网络的预测结果相较于 ARIMA 有了较大提升, 价格在跌势时下降的更快, 对跌势预测地更精确, 嗅觉更灵敏。

4.2. BP 网络结构优化

神经网络的关键问题为确定网络结构, 包括输入节点个数, 隐层数, 隐层节点个数和输出层节点个数[10] [11]。实际问题中, 根据经验直观判断网络输入输出节点个数往往无法取得好的预测结果, 既费时又费力。因此, 不能仅根据经验值来确定网络的结构, 为了解决此问题, 引入相空间重构[12] [13]对 BP 神经网络结构进行优化。

1999 年, H.S. Kim 等人基于嵌入窗法思想提出了 C-C 方法, 该方法使用关联积分同时估计出时延与嵌入窗, 是估计相空间重构参数延迟时间 τ_d 和延迟时间窗 τ_w 的有效方法[16]。

估计延迟时间 τ_d 和延迟时间窗 τ_w 的 C-C 方法通过计算以下 3 个量:

$$\bar{S}(t) = 1/16 \sum S(m, r_j, t) \quad (1)$$

$$\Delta \bar{S}(t) = 1/4 \sum \Delta S(m, t) \quad (2)$$

$$S_{cor}(t) = \Delta \bar{S}(t) + |\bar{S}(t)| \quad (3)$$

其中, $r_j = \sigma_j / 2$, σ 为给定时间序列的标准差, $\Delta S(m, t) = \max \{S(m, r_j, t)\} - \min \{S(m, r_j, t)\}$ 。由 $\Delta \bar{S}(t)$ 的第一个极小值所对应的时滞 t 计算 $\tau_d = t\tau_s$, 由 $S_{cor}(t)$ 的最小值所对应的时滞 t 计算时间窗 $\tau_w = t\tau_s$, τ_s 为采样间隔。

Table 2. BP network structure related parameters table
表 2. BP 网络结构相关参数表

参数	值(单位)
迭代次数(Epochs)	10000 (次)
学习率(lr)	0.1
目标(goal)	4×10^{-5}

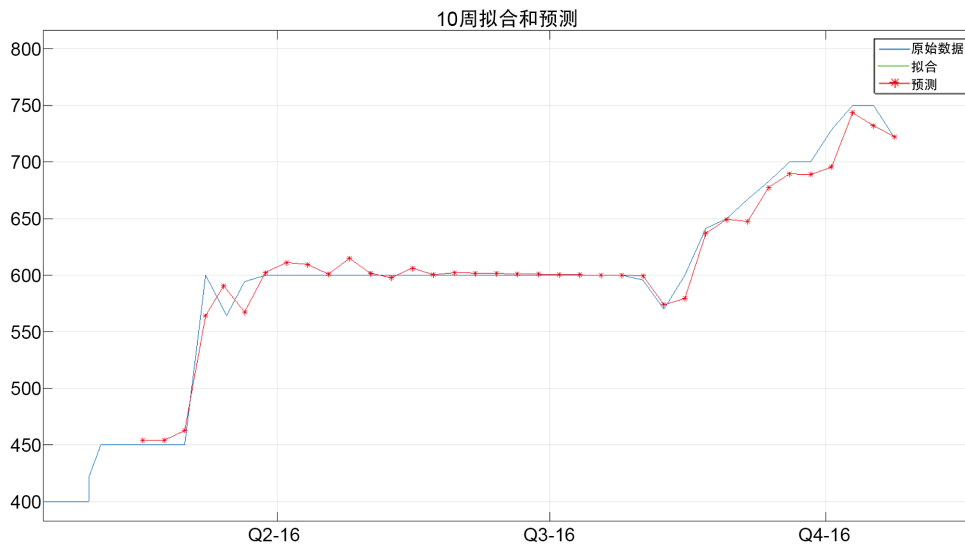


Figure 17. Comparison of predicted and true values of preliminary neural networks
图 17. 初步神经网络预测值与真实值对比图

C-C 方法中关键的量 $S(m, r, t)$ 定义为

$$S(m, r, t) = 1/t \sum_{s=1}^t [C_s(m, r, t) - C_s^m(1, r, t)] \quad (4)$$

其中, $m = 2, 3, \dots$, 关联积分为

$$C(m, r, t) = \lim_{N \rightarrow \infty} C(m, N, r, t) \quad (5)$$

价格曲线具有一定的混沌性[14], 借助相空间重构确定 BP 网络的输入维数。首先使用 C-C 方法计算相空间的嵌入维和时间延迟[15]。结果图 18 所示。

从图 18 观察得到时延和嵌入窗口的变化趋势, 时延在 7 s 时取得最小值。根据 Haykin *et al.* 提出的经验性公式确定 BP 网络的输入维数为 7, 即使用 7 周的数据来预测下一周的数据。另外, 为了让网络学习到周期性额外增加一维 ID, 用来标志目标周在一年内的位置, 如一年 52 周中的第 3 周, 其 ID 为 3。网络只设置一层隐含层, 隐含层有 8 个节点, 包括上一时刻的输出结果和本时刻的输入。最终所得优化后的 BP 神经网络结构如图 19 所示。

2) 网络训练与预测

优化后的 BP 神经网络的三七价格预测模型所用数据集仍为 2010 年 1 月 1 日~2015 年 12 月 31 日的三七周价格, 训练集、测试集、检验集按 70%、15%、15% 进行随机划分。优化后 BP 神经网络对 2016 年的预测效果如图 20 所示。从图 20 中可以相较于输入节点为 10 个的神经网络, 优化后的 BP 神经网络在涨势预测地更精准一些, 在跌势时预测价格下降缓慢, 预测值曲线更贴近于真实值曲线。

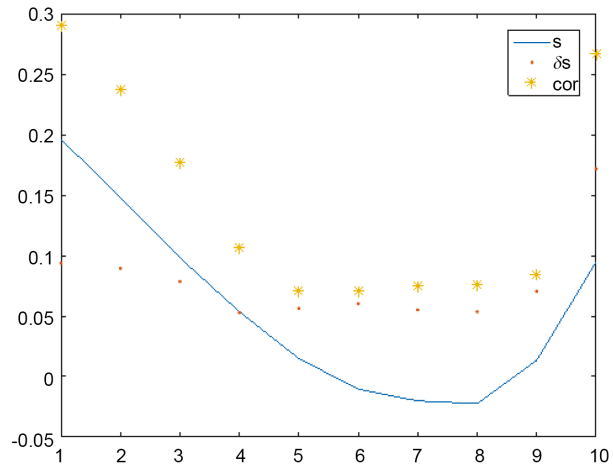


Figure 18. Phase space embedded dimension and time delay analysis results

图 18. 相空间嵌入维与时间延迟分析结果图

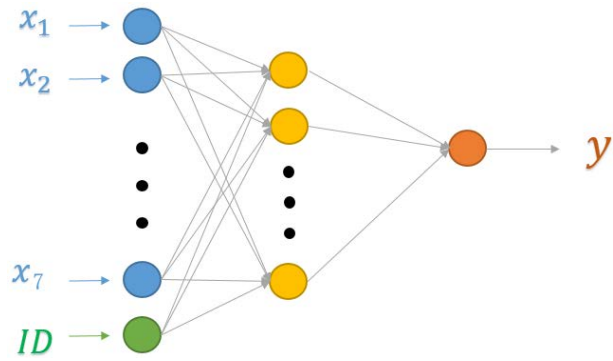


Figure 19. Optimized BP neural network structure

图 19. 优化后 BP 神经网络结构图

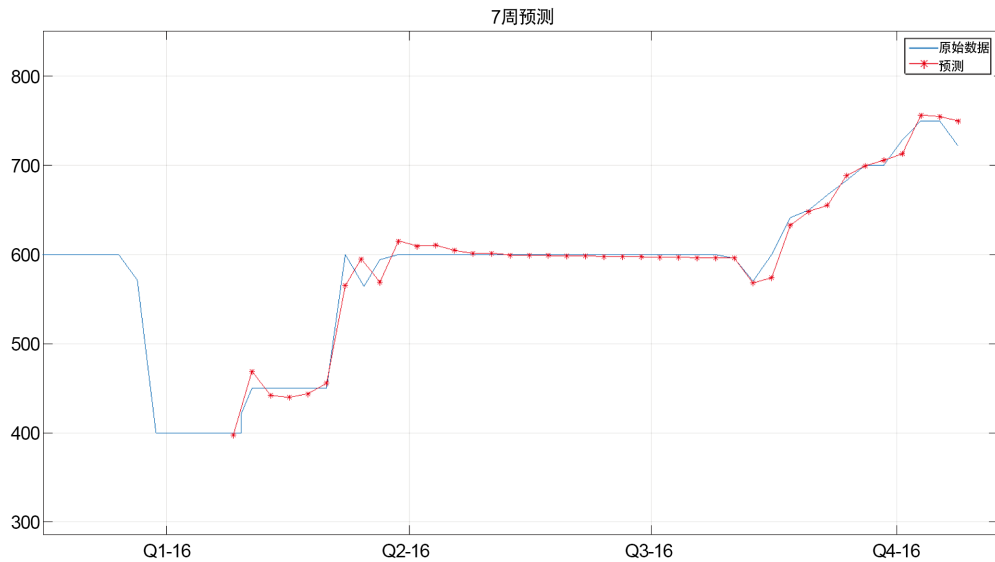


Figure 20. BP neural network 2016 price evaluation

图 20. BP 神经网络 2016 年三七价格预测

5. 结论

此次实验中共使用了 ARIMA、BP 神经网络对 2010 年 1 月至 2015 年 12 月的周价格进行学习, 利用 2016 年 1 月至 3 月的周价格对模型进行预测, 实验结构表明 BP 神经网络对非线性价格波动曲线具有更好的预测效果。另外, 为了更客观的确定神经网络结构, 提出了相空间重构的方法来确定神经网络输入节点的个数, 进一步提高了神经网络的预测精度。

未来的研究方向可以考虑使用集成算法将 ARIMA 算法和 BP 神经网络进行结合, 使用组合模型同时提取价格曲线的线性和非线性规律, 达到更好的预测效果。

基金项目

天津市重大科技专项(14ZXDZGX00867, 15ZXDSGX00090); 天津市互联网先进制造专项(15ZXHLGX00360, 15ZXHLGX00380)。

参考文献 (References)

- [1] 罗群, 游春梅, 官会林. 环境因素对三七生长影响的分析[J]. 中国西部科技, 2010, 9(9): 7-8.
- [2] 陈中坚. 三七的无公害栽培[J]. 云南农业科技, 1999(6): 44-45.
- [3] 刘峰, 王儒敬, 李传席. ARIMA 模型在农产品价格预测中的应用[J]. 计算机工程与应用, 2009, 45(25): 238-239.
- [4] 张立杰, 朱新杰. 我国棉花价格长期走势与短期预测——基于差分自回归移动平均模型(ARIMA)的分析[J]. 价格理论与实践, 2012(6): 53-54.
- [5] Contreras, J., Espinola, R., Nogales, F.J., et al. (2002) ARIMA Models to Predict Next-Day Electricity Prices. *IEEE Power Engineering Review*, **18**, 1014-1020. <https://doi.org/10.1109/MPER.2002.4312577>
- [6] Xue, D.-M. and Hua, Z.-Q. (2016) ARIMA Based Time Series Forecasting Model. *Recent Advances in Electrical and Electronic Engineering*, **9**, 93-98. <https://doi.org/10.2174/2352096509999160819164242>
- [7] Adebisi, A.A., Adewumi, A.O. and Ayo, C.K. (2014) Comparison of ARIMA and Artificial Neural Networks Models for Stock Price Prediction. *Journal of Applied Mathematics*, **2014**, Article ID: 614342.
- [8] SIMON HAYKIN. 神经网络原理[M]. 第 2 版, 北京: 机械工业出版社, 2006.
- [9] 高述涛. CS 算法优化 BP 神经网络的短时交通流量预测[J]. 计算机工程与应用, 2013, 49(9): 106-109.
- [10] 李萍, 曾令可, 税安泽, 等. 基于 MATLAB 的 BP 神经网络预测系统的设计[J]. 计算机应用与软件, 2008, 25(4): 149-150.
- [11] Yang, S.E. and Huang, L. (2005) Financial Crisis Warning Model based on BP Neural Network. *Systems Engineering-theory & Practice*, **25**, 12-19.
- [12] Ma, W., Wang, Y. and Dong, N. (2010) Study on Stock Price Prediction Based on BP Neural Network. *IEEE International Conference on Emergency Management and Management Sciences*, Beijing, 8-10 August 2010, 57-60.
- [13] 陈敏, 徐德智, 罗庆云. 时间序列相空间重构及其应用研究[J]. 计算机与信息技术, 2005(11): 11-12.
- [14] 郁俊莉, 王其文, 韩文秀. 经济时间序列相空间重构与混沌特性判定研究[J]. 武汉大学学报(理学版), 2004, 50(1): 33-37.
- [15] 修春波, 刘向东, 张宇河. 相空间重构延迟时间与嵌入维数的选择[J]. 北京理工大学学报, 2003, 23(2): 219-224.
- [16] 徐自励, 王一扬, 周激流. 估计非线性时间序列嵌入延迟时间和延迟时间窗的 C-C 平均方法[J]. 四川大学学报工程科学版, 2007, 39(1): 151-155.

期刊投稿者将享受如下服务：

1. 投稿前咨询服务 (QQ、微信、邮箱皆可)
2. 为您匹配最合适的期刊
3. 24 小时以内解答您的所有疑问
4. 友好的在线投稿界面
5. 专业的同行评审
6. 知网检索
7. 全网络覆盖式推广您的研究

投稿请点击：<http://www.hanspub.org/Submission.aspx>

期刊邮箱：csa@hanspub.org