

· 中药商业 ·

基于 LSTM 神经网络的三七价格指数预测[△]李飞飞¹, 宋庆燕¹, 陈长秀¹, 刘佳泽¹, 高雪岩^{1,2*}

1. 北京珍宝岛中药产业有限公司, 北京 100800; 2. 中国中药协会中药材种植养殖专业委员会, 北京 100700

[摘要] 目的: 为了帮助药农、中药生产企业合理决策, 引导三七产业健康有序发展, 本研究通过引入 LSTM 神经网络模型对三七的价格指数进行预测。**方法:** 本研究首先分析了影响三七价格波动的各种因素, 并通过对这些因素进行相关性分析进行因素的筛选, 降低输入向量的维度; 其次, 本研究建立了基于 LSTM 神经网络的三七价格指数预测模型; 最后, 利用 2018 年三七价格数据测试模型的有效性, 并应用有效模型对 2019 年三七价格指数进行预测。**结果:** LSTM 模型能很好的拟合三七价格整体变化的趋势, 并且具有较好的预测精确度, 预测结果能为三七的生产经营提供参考。**结论:** 研究表明, LSTM 神经网络能有效应用于三七价格指数预测研究, 将其推广到其它中药材价格预测研究中, 有利于提高资源配置有效性和保证市场健康运行。

[关键词] 三七; 价格指数; 预测模型; LSTM 神经网络

[中图分类号] F326.12; F714.1 **[文献标识码]** A **[文章编号]** 1673-4890(2019)04-0536-06

doi:10.13313/j.issn.1673-4890.20190114001

Forecast of *Panax notoginseng* Price Index Based on LSTM Neural NetworkLI Fei-fei¹, SONG Qing-yan¹, CHENG Chang-xiu¹, LIU Jia-ze¹, GAO Xue-yan^{1,2*}

1. Beijing Zhenbaodao Chinese Medicine Industry Co., Ltd., Beijing 100800, China;

2. China Traditional Chinese Medicine Association Chinese Herbal Medicine Planting and Cultivation Committee, Beijing 100700, China

[Abstract] **Objective:** In order to help farmers and pharmaceutical enterprises make rational production decisions, and guide the healthy and orderly development of *Panax notoginseng* production, this paper introduced LSTM neural network model to forecast the price index of *Panax notoginseng*. **Methods:** Firstly, this paper analyzed the various factors that affect the price fluctuation of *Panax notoginseng*, then reduced the dimension of the input vector through the correlation analysis. Secondly, this paper established a forecasting model of *Panax notoginseng* price index based on LSTM neural network. Finally, the validity of the model is tested, and the effective model is applied to forecast *Panax notoginseng* price index in 2019. **Results:** The analysis results show that the LSTM model has high prediction accuracy, the prediction results can provide some suggestions for the production and operation of *Panax notoginseng*. **Conclusion:** The research shows that LSTM neural network can be effectively applied to the prediction of *Panax notoginseng* price index, and can be extended to prediction of other Chinese medicinal materials price, which can improve the efficiency of resource allocation and ensure the healthy operation of the market.

[Keywords] *Panax notoginseng*; price index; prediction model; LSTM neural network

三七是常用大宗中药材之一, 为五加科人参属植物三七 *Panax notoginseng* (Burk.) F. H. Chen 的干燥根和根茎^[1]。现代研究表明三七具有抗衰老、抗疲劳、抗缺氧、降血压、降血脂、降血糖和提高机体免疫力等药理作用^[2]。三七临床应用广泛, 近年来需求量不断增加, 目前三七年需求量约 25 000 t。

近年来, 市场炒作、种植农户不了解市场信息等原因部分中药材投机商提供了机会, 对中药材市场价格波动起到了推波助澜的作用。市场的波动造成药农和中药材经营者亏损严重, 种植和经营积极性下降, 进一步影响下游中药饮片和中成药企业的稳定生产, 影响行业的健康可持续发展。对三七

[△] **[基金项目]** 国家工业和信息化部2017年工业转型升级(中国制造2025)中药材技术保障公共服务能力建设项目(0714-EMTC-02-000195); 安徽省支持现代医疗和医药产业发展申报项目—中药材大数据云共享服务平台

* **[通信作者]** 高雪岩, 博士, 研究方向: 中药产业大数据及其应用; Tel: 18910881558, E-mail: fzghzx201901@163.com

价格进行预测,可以帮助药农和中药生产企业合理决策,引导三七产业健康有序发展,成为一个重要的研究课题。

目前,中药材相关的预测模型建设已有部分研究。马健^[3]基于时间序列对亳州的白芍产量进行了预测分析。冯烽等^[4]基于灰色系统对中药材罗汉果价格进行了预测。王诺等^[5]等基于自回归积分滑动平均模型对中药材三七价格预测。马广慧等^[6]基于遗传 BP 神经网络的三七价格预测。马健等^[7]基于遗传模拟退火三次指数平滑模型对中药材价格进行了预测。这些信息预测模型均只考虑了待预测量的历史数据对其未来数据的影响,没有考虑到影响待预测量变化的其它因素,而对于中药材价格而言,其价格变化必然与供需、成本、宏观经济环境等各种因素密切相关,缺少这些因素的预测存在很大局限性。常征宇等^[8]研究了各类因素对中药材价格的影响,并获得了各类因素对中药材价格变化的贡献度,但是未给出各因素与价格之间的定量关系。杨勇等^[9]从市场角度分析中药材价格的影响因素,建立向量自回归模型(VAR 模型),并运用脉冲响应和方差分解定量方法分析这些影响因素作用的大小,利用已知数据对中药材价格指数进行价格预测,但是模型对 3 个季度以后价格的预测精度急剧下降。

循环神经网络 RNN(Recurrent Neural Network)是一类用于处理序列数据的神经网络。与传统的神经网络相比,RNN 的特点在于将时序的概念引入到模型的设计中,RNN 网络会对前面的信息进行记忆并应用于当前输出的计算中,即隐藏层之间的节点不再无连接而是有连接的,并且隐藏层的输入不仅包括输入层的输出还包括上一时刻隐藏层的输出。这一特点使得 RNN 能够很好的应用于时序输入序列的预测。LSTM(Long Short Term)神经网络是 RNN 的一种特殊形式,利用 LSTM 模型学习长期依赖信息,在很多应用中取得了巨大成功。本研究拟首先分析影响三七价格波动的外在因素,并分析各影响因素的相关性,进行影响因素的筛选后引入 LSTM 模型,以多种影响因素的时间序列作为模型输入,对三七价格进行预测。

1 LSTM 神经网络

LSTM 是 RNN 神经网络的一种特殊类型。RNN 神经网络由输入层、隐藏层、输出层组成,其中隐藏层接收本个序列以及上一个序列通过 RNN 隐藏层计算后输出的结果,实现时间序列的历史信息通过

隐藏层计算后与新一期的序列值共同作用于隐藏层,形成新的输出。如图 1、图 2 所示,RNN 结构按照时间线展开可以更清楚的看到 RNN 的结构。

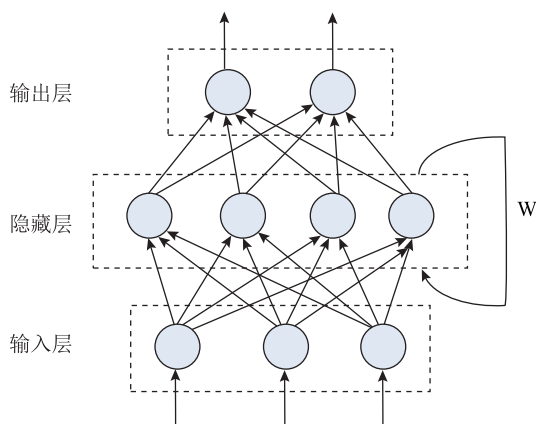


图1 RNN神经网络结构图

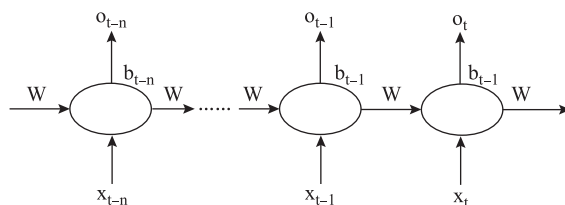


图2 RNN神经网络沿时间维度展开

假设 RNN 的输入层大小为 I , 隐层大小为 H , 输出层大小为 K 。输入层、隐藏层、输出层的向量序列分别表示为: $x_1, x_2, x_3 \cdots x_t$; $b_1, b_2, b_3 \cdots b_t$; $o_1, o_2, o_3 \cdots o_t$ 。图 2 中 x_t 为输入序列 x 的第 t 项输入, a_t 为第 t 项的隐藏层的输入, b_t 为第 t 项隐藏层的输出。 a_t 由第 t 项输入层的 x_t 以及第 $t-1$ 项的隐层输出项 b_{t-1} 共同决定:

$$a_t^h = \sum_i \omega^{ih} x_t^i + \omega^{h'h} b_{t-1}^{h'} \quad (1)$$

$$b_t^h = f(a_t^h + m_h) \quad (2)$$

其中 ω^{ih} 、 $\omega^{h'h}$ 分别为输入层和隐藏层间的权重矩阵、前项隐藏层输出传输到后项隐藏层输入间的权重矩阵, m_h 为隐层偏置, $f(x)$ 为隐层激活函数, 常见的激活函数有 Sigmoid, 双曲正切函数 tanh 等。第 t 项隐藏层输出 b_t 传导至输出层, 输出层的输出为:

$$o_t = g\left(\sum_k \omega^{hk} b_t^k + n_k\right) \quad (3)$$

其中 ω^{hk} 为隐层到输出层权重矩阵, n_k 为输出层偏置, $g(x)$ 为输出层激活函数。

对于长期依赖问题,理论上 RNN 能够处理无限长的时间序列数据,但是实际应用中,常常存在丧失学习到连接较远的信息的能力的情况。LSTM 作为

RNN 的一种特殊类型, 改进了 RNN 的结构, 通过构建特殊的控制单元解决了长期依赖的问题。LSTM 存储和计算模块内部结构的核心思想是将历史信息作为细胞状态在一条专门的信息传送路径中传播, 信息受较小干扰, 容易保持不变。图 3 为 LSTM 的一个存储处理单元的基本结构, 其中的忘记门读取 o_{t-1} 和 x_t 的信息并输出信息以决定选择忘记 b_{t-1} 的某些信息, 输入门确定什么样的新信息被存放在细胞状态中; 输出门确定何种信息被输出。

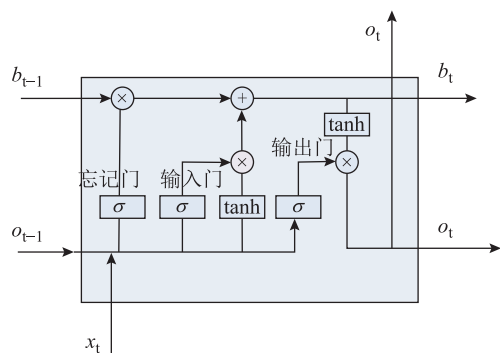


图 3 LSTM 神经网络存储处理单元

2 数据准备及预处理

2.1 三七价格波动影响因素分析及数据准备

三七价格波动的影响因素包括宏观因素、微观因素以及各类突发类事件因素^[10-14]。宏观方面, 国家的发展水平、三七所处的医药行业发展水平、国家的宏观货币政策等是影响价格波动的中药因素; 微观方面, 三七的供需和成本是影响价格波动的关键因素; 除此之外, 三七价格的波动还受到一些突发事件的影响, 如三七为原料的新产品开发、相关疫病流行、国家政策变化等。

本研究基于北京珍宝岛中药产业有限公司提供的三七 2006—2018 年在全国范围内的产量、需求量、库存量、成本、全国种植面积、亩产量等数据, 并从官方网站获取年度居民消费支出、月度全国中成药产量、月度货币供应量数据建立基于 LSTM 的价格预测模型, 用以预测三七价格走势。

根据珍宝岛中药材大数据中心提供的数据, 2017 年全国三七产量在 15 000 t 左右, 三七的年需求量在 25 000 t 左右, 目前年用量以 10% ~ 20% 递增, 预计 2018 年三七的用量在 27 000 t 左右。三七种植成本目前为 32 000 元/亩左右。三七 2006—2018 年间产量、需求量、种植面积、亩产、种植成本数据见表 1。

表 1 2006—2018 年三七供需及成本数据

年份	年产量/t	年需求量/t	总种植面积/hm ²	亩产(干重)/kg	亩种植成本/元
2006	8945	7000	120 000	165.34	14 000
2007	9166	7000	120 000	163.68	14 500
2008	8842	6750	121 000	173.33	15 000
2009	4454	6750	69 000	158.93	16 000
2010	4934	6750	84 000	140.86	19 000
2011	4707	6750	97 000	151.61	24 000
2012	6500	6750	155 000	140.00	29 000
2013	10 000	7500	292 000	144.93	34 000
2014	25 000	11 000	300 000	200.00	40 000
2015	33 861	20 000	453 200	129.49	50 000
2016	25 000	24 000	200 000	166.67	46 000
2017	15 000	25 000	280 000	170.00	30 000
2018	11 000	27 000	300 000	170.00	32 000

本研究使用的宏观数据如年度居民消费支出、月度全国中成药产量、月度货币供应量均来自国家官方统计数据。

三七产新期固定, 在产新期前后存在价格的明显波动, 因此本研究最后引入月份作为三七价格波动的一个重要影响因素。

市场上三七的规格较多, 其中出现最多的规格有: 20 头、40 头、60 头、80 头、120 头、无数头等, 不同规格的三七价格存在较大差异。本研究采用珍宝岛中药材数据库中三七的价格指数数据, 此指数综合了三七各个规格的价格。2006—2018 年部分三七价格指数见表 2, 三七价格指数变化趋势见图 4。

表 2 2006—2018 年部分三七价格数据

日期	价格指数	日期	价格指数
2006-1-31	54.00		
2006-2-28	52.00	2017-9-30	284.20
2006-3-31	51.10	2017-10-31	277.05
2006-4-30	48.80	2017-11-30	270.35
2006-5-31	46.55	2017-12-31	262.09
2006-6-30	47.00	2018-1-31	262.20
2006-7-31	47.00	2018-2-28	272.96
2006-8-31	45.90	2018-3-31	275.13
2006-9-30	45.00	2018-4-30	275.65
2006-10-31	47.74	2018-5-31	268.46
2006-11-30	48.40	2018-6-30	257.85
2006-12-31	44.81	2018-7-30	246.64
2007-1-31	45.74	2018-8-31	238.83
2007-2-28	48.00	2018-9-30	229.11
.....		2018-10-31	222.01

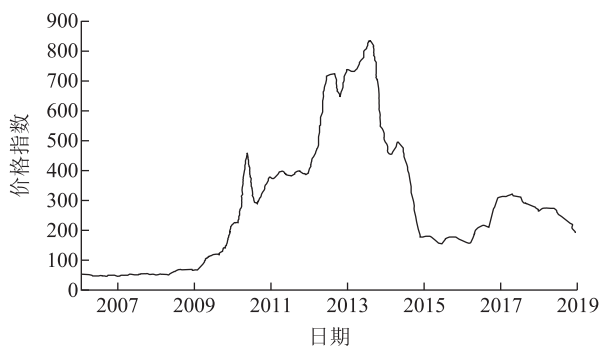


图4 三七价格指数变化趋势图

2.2 影响因素特征选择与归一化处理

本研究通过对影响三七价格变化的宏观、微观因素进行分析,初步确定使用产量、需求量、种植面积、亩产、成本、中成药产量、货币和准货币(M2)、居民消费支出、月份共9个维度的影响因素。

高维数据建模增加了模型的复杂度,降低了算法的泛化能力,增加了算法的搜索难度。建模前进行特征选择,消除冗余数据,有利于提升模型性能。本研究通过分析影响因素间相关性进行影响因素选择。

图5为三七的产量、需求量、种植面积、亩产、成本、中成药产量、货币和准货币(M2)、居民消费支出8个因素的相关性分析。从图中可以看出,中成药产量、货币和准货币(M2)、居民消费支出这3个宏观因素间存在明显的线性相关性。因此从3个因素中任意剔除其它2个因素,本研究剔除中成药产量和居民消费支出2个因素,最终保留三七的产量、需求量、种植面积、亩产、成本、货币和准货币(M2)、月份作为模型的输入变量。

完成特征选择后,对筛选后的特征进行归一化处理。本研究采用 min-max 标准化方法,即通过下列公式对数据进行标准化处理。

$$x^* = \frac{x - \min}{\max - \min} \quad (4)$$

3 预测模型构建

基于上述算法说明与数据准备,本研究构建了基于 LSTM 的三七价格指数预测模型。其预测流程主要包括数据预处理、影响因素筛选、数据归一化处理、基于 LSTM 的三七价格指数预测模型建设、训练、模型测试和应用几个部分,见图6。

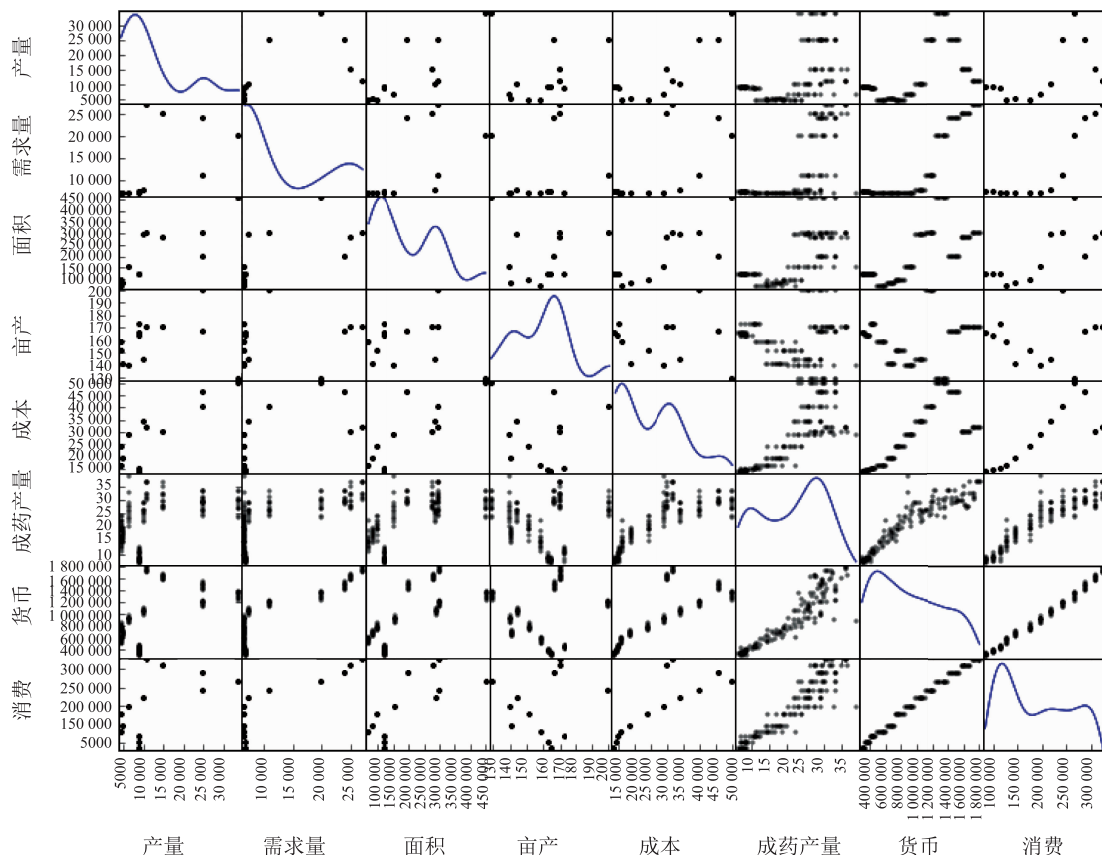


图5 影响因素间相关性分析

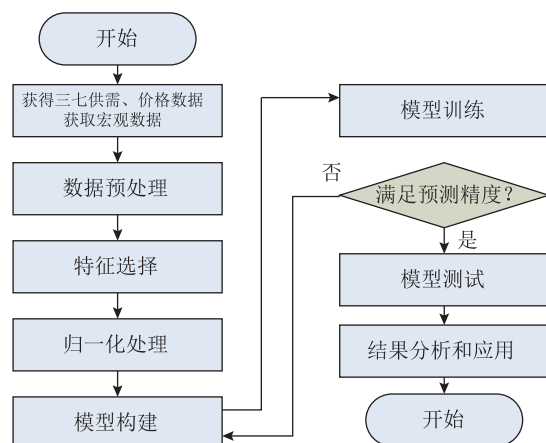


图6 影响因素间相关性分析

数据的预处理包括数据获取后对缺失值、异常值的基本处理,模型的构建和训练过程中不断检测模型的预测精度,如果精度达到要求则进入下一步模型测试阶段,精度未达到要求则通过增加训练次数、调整模型参数等直到模型达到精度要求。

4 结果与分析

本研究使用2006—2017年数据作为模型的训练数据,使用2018年数据作为模型的测试数据。基于LSTM的三七价格指数模型建立后,针对测试数据进行模型预测效果测试。

图7中实线为真实价格指数数据,虚线为拟合数据和预测数据。2018年的价格预测结果如图7所示。



图7 三七价格预测

从表3中可以看出,模型预测2018年三七价格指数数据的波动范围在206~306之间,全年价格呈现先涨后跌态势,总体比2017年略有下跌。对比2018年三七价格指数实际波动情况可以看到,预测结果能够很好的拟合三七价格变动的整体趋势。2018年上半年前期2—3月份略有上升,之后缓慢下跌。预测精度上,预测值与实际值最大误差为13.01%,最小误差为2.36%。

表3 模型预测2018年三七价格数据

日期	三七价格原指数值	三七价格指数预测值	预测误差率/%
2018-1-31	226.198 7	300.084 5	12.73
2018-2-28	272.958 5	305.977 8	12.10
2018-3-31	275.133 0	305.380 1	10.99
2018-4-30	275.648 6	301.184 4	9.26
2018-5-31	268.457 5	294.038 6	9.53
2018-6-30	257.849 5	287.346 0	11.44
2018-7-31	246.639 8	275.863 3	11.85
2018-8-31	238.834 9	264.249 2	10.64
2018-9-30	229.144 0	238.844 9	4.25
2018-10-31	222.007 9	227.245 8	2.36
2018-11-30	199.405 1	209.007 9	4.82
2018-12-31	183.077 0	206.897 0	13.01

此外,图7还展示了应用此模型另外对2019年上半年的三七价格进行预测的结果,2019年上半年的预测价格见表4。从表4中可以看出,2019年上半年三七价格范围处于216~236之间,整体上在年初略有上涨,3—4月份之后呈现缓慢下跌趋势,三七价格指数在2019年波动较为平缓。

表4 模型预测2019年上半年三七价格数据

日期	三七价格原预测值
2019-1-31	216.20
2019-2-28	232.96
2019-3-31	235.13
2019-4-30	235.65
2019-5-31	228.46
2019-6-30	217.85

5 结论和讨论

三七作为我国传统名贵中药材,是很多经典名方以及大品种中成药的重要原料,其功效也受到消费者越来越多的认同。近年来,三七的价格在市场需求、资本炒作等多方面因素的影响下,2013—2014年出现了大涨大跌,价格的非正常波动造成药农盲目进行种植扩张,继而造成供需不平衡,影响三七相关产业的健康发展。通过科学方法对三七的价格进行预测,进而指导三七药材的生产者和经营者作出正确的生产经营决策,能够有效抑制投机,保证市场健康稳定运行。

通过上述的实验数据表明,LSTM神经网络应用于三七价格指数预测取得了较好的效果。分析主要原因在

于三七等中药材价格变化为典型的时序序列,历史价格对当期价格的影响较大,而LSTM神经网络对历史信息处理的典型优势特别有利于三七价格指数的预测。

本文建立的价格指数预测模型也可以引入到其他中药材价格预测,通过信息化、智能化高效利用市场宏观数据和微观数据,有利于达到提高资源配置有效性和保证市场健康运行的目标。

参考文献

- [1] 国家药典委员会. 中华人民共和国药典:一部[M]. 北京:中国中医药科技出版社,2015:11.
- [2] 杨娟,袁一怔,尉广飞,等. 三七植物化学成分及药理作用研究进展[J]. 世界科学技术-中医药现代化,2017,19(10):1641-1647.
- [3] 马健. 基于时间序列的亳州白芍产量预测分析[J]. 呼伦贝尔学院学报,2016,24(3):78-81.
- [4] 冯峰,韦范,缪剑华. 基于灰色系统GM(1,1)模型的中药材罗汉果价格预测[J]. 广西科学,2012,19(1):15-20.
- [5] 王诺,程蒙,臧春鑫,等. 基于自回归积分滑动平均模型对中药材三七价格预测的讨论[J]. 中国中药杂志,2016,41(8):1559-1566.

- [6] 马广慧,马豆豆,邵秀丽. 基于遗传BP神经网络的三七价格预测[J]. 天津师范大学学报(自然科学版),2017,37(6):76-80.
- [7] 马健,盛魁,朱庆友,等. 基于遗传模拟退火三次指数平滑模型的中药材价格预测研究[J]. 佳木斯大学学报(自然科学版),2017(5):152-155,169.
- [8] 常征宇,王树进. 我国中药材价格波动影响因素的实证研究—基于因子分析和向量自回归模型[J]. 科技与经济,2015,28(5):101-105.
- [9] 杨勇,陶群山. 基于向量自回归模型的中药材价格影响因素分析及预测[J]. 中国现代中药,2019,21(1):111-115.
- [10] 李化. 中药材价格传导研究[J]. 卫生经济研究,2015(10):58-61.
- [11] 闫桂权,何玉成,聂飞. 我国三七价格波动特征与变动规律研究—以云南文山三七为例[J]. 四川农业大学学报,2018,36(5):143-151.
- [12] 王潇,王昌利. 试论中药材价格变动的成因及影响[J]. 商场现代化,2017,14(7):53-54.
- [13] 张晋之,杨元娟,许燕. 中药材价格波动的原因及优化策略[J]. 价格月刊,2016(2):35-38.
- [14] 张昱,王建忠,赵瑞东. 河北省中药材价格波动研究[J]. 农村经济与科技,2018,29(17):155-156.

(收稿日期:2019-01-14 编辑:周鹭)

(上接第535页)

5 总结

中药材市场因为供需失衡而进入低谷期,对于市场参与者而言,也正是转型升级的最佳时期。同样在分化下的行业时期,也是洗牌重组的密集期。参与者与其坐以待毙式的等待被吞掉,不如主动出击,寻找业务或模式上的互补和联合,以结构性升级来获得更广阔的市场。随着道地、非道地区域的差别发展及中药材生产供给侧改革的进行,中药材生产、流通端资源优化日益加快,加上新技术的继续助推,中药材市场将会愈发有质量的发生变革。

参考文献

- [1] 中药材天地网. 中药材价格综合200指数2018年走势[EB/OL]. (2018-01-01)[2019-02-11]. <https://zhishu.zyctd.com/>.

- [2] 中药材天地网. 中药材价格综合200指数价格[EB/OL]. (2018-01-01)[2019-02-11]. <https://www.zyctd.com/jiage/>.
- [3] 中药材天地网. 中药材家种100指数2018年走势[EB/OL]. (2018-01-01)[2019-02-11]. <https://zhishu.zyctd.com/jiazhong>.
- [4] 中药材天地网. 中药材野生99指数2018年走势[EB/OL]. (2018-01-01)[2019-02-11]. <https://zhishu.zyctd.com/yesheng>.
- [5] 中国人大网. 中华人民共和国环境保护税法[EB/OL]. (2018-11-05)[2019-02-18]. http://www.npc.gov.cn/npc/xinwen/2018-11/05/content_2065629.htm.
- [6] 农业农村部,国家药品监督管理局,国家中医药管理局. 关于印发《全国道地药材生产基地建设规划(2018—2025年)》的通知[EB/OL]. (2018-12-19)[2019-02-20]. http://www.moa.gov.cn/govpublic/ZZYGLS/201812/t20181219_6165190.htm.

(收稿日期:2019-03-11 编辑:周鹭)