Forecasting Horticultural Products Price Using ARIMA Model and Neural Network Based on a Large-Scale Data Set Collected by Web Crawler

Yuchen Weng, Xiujuan Wang, Jing Hua, Haoyu Wang, Mengzhen Kang, *Member, IEEE*, and Fei-Yue Wang, *Fellow, IEEE*

***Abstract*— The sales of agricultural products are an important component of the product supply chain. The price of agricultural products, a social signal of product supply and demand, is affected by many factors, such as climate, price, policy, and so on. Due to the asymmetry between production and marketing information, the price of many agricultural products fluctuates greatly. Horticultural products are especially sensitive to price since they are not suitable for long-term storage. Therefore, forecasting the price of horticultural products is very helpful in designing a cropping plan. In this paper, AutoRegressive Integrated Moving Average (ARIMA) model, back propagation (BP) network method, and recurrent neural network (RNN) method were tested to forecast the price of agricultural products (cucumber, tomato, and eggplant) in short term (several days) and long term (several weeks or months). A large-scale price data of agricultural products were collected from the website based on web crawler technology. Since ARIMA requires continuous and periodic data, it is suitable for small-scale periodic data. It gave good performance for average monthly data but not for daily data. Instead, the neural network methods (including BP network and RNN) can predict well daily, weekly, and monthly trend of price fluctuation. It is more suitable for large-scale data. It is expected that the deep learning method represented by a neural network will become the mainstream method of agricultural product price forecasting.**

***Index Terms*— Agricultural cyber-physical-social system (CPSS) framework, AutoRegressive Integrated Moving Average (ARIMA) model, cucumber, neural network, price forecasting, web crawler.**

Manuscript received January 29, 2019; accepted April 17, 2019. Date of publication May 20, 2019; date of current version June 10, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 31700315, Grant 31400623, and Grant 61533019. *(Corresponding author: Yuchen Weng.)*

Y. Weng and X. Wang are with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and also with the Beijing Engineering Research Center, Intelligent Systems and Technology, Beijing 100190, China (e-mail: wengyuchen2017@ia.ac.cn; xiujuan.wang@ia.ac.cn).

J. Hua and M. Kang are with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and also with the Innovation Center for Parallel Agriculture, Qingdao Academy of Intelligent Industries, Qingdao 266109, China (e-mail: mengzhen.kang@ia.ac.cn; jing.hua@ia.ac.cn).

H. Wang is with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China.

F.-Y. Wang is with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, also with the School of Computer and Control Engineering, University of Chinese Academy of Sciences, Beijing 100049, China, and also with the Research Center for Military Computational Experiments and Parallel Systems Technology, National University of Defense Technology, Changsha 410073, China (e-mail: feiyue.wang@ia.ac.cn).

Digital Object Identifier 10.1109/TCSS.2019.2914499

I. Introduction

THE prices of agricultural products in the market often fluctuate due to the season, region, economic level, and social reasons. It is a direct indicator of product supply and demand. Even in the same city, the price may be different among markets. The price variation is partly caused by the information asymmetry between production and marketing: farmers cannot forecast the balance between the market supply and demand ahead. Often they arrange the planting and harvesting dates based on their own experience without guidance. As a result, there is a risk of selling an agricultural product with low product. Especially horticultural products are not suitable for long-term storage and must be sold in a timely manner. Greenhouse environment is partially controllable, thus the cropping design, including planting area, time, and duration can adjusted to fit better product price.

In China, sales of agricultural products are concentrated in large wholesale markets, which make the prices transparent. In China, the price of agricultural products has a wide range of publicly available data on the website. The official websites of the local agricultural wholesale market and the agricultural administration department will often announce the prices of the main agricultural products. These websites are usually manned and updated with timely data. A web crawler is a generic term for computer programs that automatically extract the information from a website. With web crawl technology [1], price data can be sensed, providing a social signal of product supply and demand. Each wholesale market contains daily highest and lowest prices of various products, bringing a source of big data, bringing new chances of applying data-driven algorithms.

In China, in the past years, a lot of solar greenhouses are built, partly because of the land-transferring policy. Planting arrangement is more centralized and scaled, which demands technologies for risk reduction. A cyber-physical-social system (CPSS) for agriculture has been proposed [2] to design a cropping plan based on price information. The CPSS framework has been applied in many fields [3], [4]. In this paper, historical price data have been used in testing the nonlinear planning. However, planning for future need price forecast in long term or short term.

Long-term (some weeks or months) and short-term (several days) prediction have different purposes. Long-term prediction can support the decision of planting time since the growing season of the crop usually lasts more than 1 month. It is preferable that the harvest time coincides with the high price so that growers can get more benefit. Short-term prediction is more useful to choose a harvest date when the plants are not far from harvest, as the crop phenology can be adjusted in the greenhouse. Short-term forecasts of daily average prices can provide guidance for consumers and markets but not for production and harvesting.

Different methods have been applied to predict the price of agricultural products: regression method, times series method, and machine learning method. The regression methods, such as vector autoregressive model and vector autoregressive moving average model, usually predict the price of agricultural products by taking into account the influencing factors such as places and policies. This method can deal with only one kind of crops, thus the generalization performance is relatively poor [5], [6]. Time series methods, such as AutoRegressive Integrated Moving Average (ARIMA) model [7] and Generalized AutoRegressive Conditional Heteroskedasticity model [8], take in account of historical price information of agricultural products while ignoring other factors. Therefore, these methods are difficult to predict the fluctuations caused by some sudden factors such as whether or disasters [9]. Machine learning methods include neural network and Support Vector Machine algorithm. These methods have a wide range of applications. However, when predicting different agricultural products, big data is required and overfitting may happen [10].

As mentioned earlier, current methods usually aim at a particular agricultural product on the market because of the limitation of data. Most of the methods are not tested by large-scale data and can only make a short-term prediction. In this paper, a big data set can be built with web crawler. The monthly, weekly, and daily average prices of cucumbers were forecasted using the ARIMA model, back propagation (BP) network method, and recurrent neural network (RNN) [11], [12]. By combining monthly forecast with a weekly and daily forecast of the agricultural product price, we can provide guidance for agricultural product production and harvesting plans.

II. Materials and Methods

A. Data Source

Web crawler technology playing a role of social sensor is used to achieve price data. Generally, a crawler program can get the content of a web page according to the web link address (URL), analyze the structure of the page to extract the information needed, and can also analyze other links of the page with recursive retrieval. The most important parts of the crawler program are accessing the web server (including user authorization and file downloading) and parsing the acquired HTML files. In this paper, the Python language has been used to implement these two functions with httplib2 and beautiful soup libraries.

We chose Beijing Xinfadi Market as the target to collect and collate the price data of all kinds of agricultural products. It is a key market in Beijing which gathers product from source area (such as Shandong province and Hebei province) and then distribute products to supermarket and community retailers. The period for data collecting is from August 19, 2015 to July 30, 2018. The crawler program obtains pages every day and extracts daily prices. The monthly average price data shown in this paper is the average transaction price of agricultural products in Beijing from January 2010 to May 2018, from another data source.

B. Data Processing

The price data extracted by the crawler technology has some quality issue since the data format is not uniform and sometimes there are missing data. Therefore, data cleaning including adjustment, deletion, calibration, and unification were conducted after data collection. Finally, the data were imported into the database.

The price of agricultural products has a certain periodicity and remains stable except for major holidays or social activities. In a short period, the prices have a trend of gradually rising or declining, similar to a quadratic function. Thus, for missing data, a linear interpolation method is applied. The equation is as follows:

$$p(t) = \frac{t - t_1}{t_0 - t_1} p\left(t_o\right) + \frac{t - t_0}{t_1 - t_0} p(t_1) \tag{1}$$

where t is the date with missing price data, is the missing price, t_0 is the date immediately before t, t_1 is the date immediately after t, and $p(t_o)$ and $p(t_1)$ are the price of these two dates, respectively.

The normalized price data are used in this study to facilitate data processing and ensure faster convergence of the program. The normalization of data is to scale all the data to the interval (0, 1).

C. Data Set

Data set is designed for both methods. A prediction is made both for short term (days) and long term (weeks and months), daily, weekly, and monthly price data sets are chosen, for both training and test. Monthly and daily average prices are obtained from the Internet. Weekly average prices are calculated on the basis of daily average prices. Cucumber, which is a typical horticulture product, was chosen for price analysis.

1) *Monthly Average Price:* The training set is the average transaction price from January 2010 to December 2016, and the test set is the data from January 2017 to May 2018.
2) *Weekly Average Price:* The training set is the price data from August 19, 2015 to May 14, 2018, and the test set is the data from May 15, 2018 to July 23, 2018.
3) *Daily Average Price:* The training set is the price data from August 19, 2015 to June 30, 2018, and the test set is the data from July 1, 2018 to July 23, 2018.

The data set information is summarized in Table I, applicable for three methods.

TABLE I
TRAINING DATA AND TESTING DATA USED BY THE MONTHLY, WEEKLY, AND DAILY PREDICTION

Temporal scale	Training data	Testing data
Monthly	2010/1-2016/12	2017/1-2018/5
Weekly	2015/8/19-2018/5/14	2018/5/15-2018/7/23
Daily	2015/8/19-2018/6/30	2018/7/1-2018/7/23

III. MODEL DESCRIPTION

A. ARIMA Model

ARIMA model is a famous time series prediction method proposed by Box and Jenkins in the early 1970s [13]. The ARIMA model transforms the nonstationary time series into stationary time series and then regresses the dependent variable due to the lag value of its independent variable and the present value and lag value of its random error term [14]. The parameters p, d, and q represent the autoregressive term, the times of differences when the time series becomes stationary, and the number of moving average terms, respectively.

ARIMA algorithm is a linear regression method, which calculates data through its own historical data. Thus, the ARIMA algorithm is applicable to the data with high and stable correlation and performs well for simple and short-term prediction.

The mathematical expression of the ARIMA model is as follows:

$$\beta(t) = \mu + \rho(Y)\alpha(Y)^{-1}\epsilon(t) \tag{2}$$

where $\beta(t)$ is a sequence value, which is the average price corresponding to each time period, t is the time, $\epsilon(t)$ is the disturbance value, and Y is a backward moving operator, which can be computed in the following equation:

$$Y(W(t)) = W(t-1) \tag{3}$$

$\rho(Y)$ is a moving average operator

$$\rho(Y) = 1 - \rho_1(Y) - \rho_2(Y) - \cdots - \rho_q(Y) \tag{4}$$

$\alpha(Y)$ is an autoregressive operator

$$\alpha(Y) = \alpha_1(Y)\alpha_2(Y) - \cdots - \alpha_p(Y). \tag{5}$$

In the process of model parameter determination, we first perform the Augmented Dickey–Fuller (ADF) test. If the data does not pass the test, the original data will be differentiated until it passes. In actual price forecasting of agricultural products, data often have a strong periodicity, so d is usually set to 1. The values of p and q are integers and are determined by the Akaike information criterion (AIC) test. First, we set the range of p and q between 1 and 10, using the training set through AIC test. We get the values of parameters p and q that minimize AIC value using the exhaustive method. Because there are many kinds of agricultural products, if every agricultural product is tested to obtain p and q, it is very time-consuming. Generally, the values of p and q range from 8 to 10 [15]. The training result may be less accurate if the values of p and q are smaller than 8. It is computationally time-consuming without getting better results if you choose a larger number of p and q.

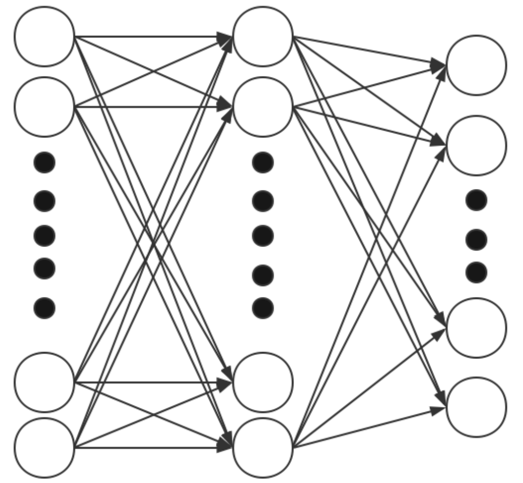


Fig. 1. Structure of the neural network. The historical prices are used as input and the output will be the future prices.

B. BP Network Method

A neural network is widely used in various fields of scientific research studies in recent years [16], [17]. Compared to the ARIMA model, the BP network algorithm does not require the periodicity of the data sequence. The price data are influenced by various factors, such as weather, price, social activities, and so on. Moreover, some factors take effect with delay, such as sudden extreme weather, which will increase the vegetable price for several days. Such effects are difficult to be forecasted accurately. We forecast the prices using BP neural network algorithm. The price of the previous period is used as input, and the future price is set as output [18], [19].

We use the Levenberg–Marquardt BP function as the training function and gradient descending beam learning function as the learning function. We take mean square error function as the performance function and S-type tangent function as the transfer function [20], [21]. The training data and testing data are showed in Table I. The structure of the neural network is showed in Fig. 1.

1) *Monthly Average Price Prediction:* The number of the input and output nodes is set to 30 and 1, respectively. The number of the hidden layer is 4, and the number of the hidden layer nodes is 10.
2) *Weekly Average Price Prediction:* The number of the input and output nodes is set to 20 and 1, respectively. The number of the hidden layer is 4, and the number of the hidden layer nodes is 10. The price data of the previous 20 weeks were used as the input to get the prices of the next 1 week.
3) *Daily Average Price Prediction:* The number of the input and output nodes is set to 30 and 7, respectively. The number of the hidden layer is 1, and the number of the hidden layer nodes is 20. The price data of the previous

TABLE II
PARAMETERS OF THE NETWORK

Numbers	Monthly	Weekly	Daily
Input nodes	12	20	30
Output nodes	4	4	7
Hidden layer	1	1	1
Hidden layer nodes	10	10	20

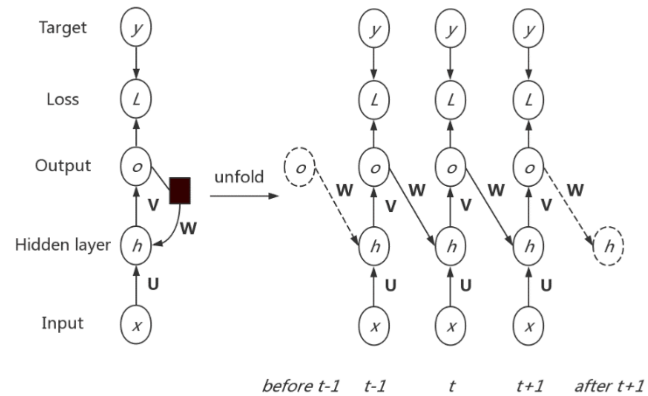


Fig. 2. Structure of the RNN.

30 days were used as the input to get the prices of the next 7 days.

The parameters of the BP network are shown in Table II. We also set the number of output nodes larger in order to see whether BP network can perform well for longer period forecasting.

C. Recurrent Neural Network

RNN is a kind of neural network used to process sequential data [22], which has been widely used in natural language processing [23], speech recognition [24], and other fields. The basic idea of RNN is that the network will memorize the previous information and apply it to the calculation of the current output, i.e., the nodes between the hidden layers are connected, and the input of the hidden layer includes not only the output of the input layer but also the output of the hidden layer at the last moment. Based on the idea of graph expansion and parameter sharing, we can design various kinds of RNN. The structure diagram of the RNN used in this paper is shown in Fig. 2.

The only loop for this type of RNN is the feedback link from the output to the hidden layer. The hidden layer and the output layer use hyperbolic tangent activation function and pure linear activation functions, respectively. RNN propagates forward from the initial state. For each time step, we can update the equation as follows:

$$a(t) = b + Wo(t-1) + Ux(t) \tag{6}$$

$$h(t) = \tanh(a(t)) \tag{7}$$

$$o(t) = c + Vh(t), \tag{8}$$

where b and c are the offset vectors of the hidden layer and the output layer, respectively. U, V, and W, respectively, represent the weighting matrices of the input layer to the hidden layer, the hidden layer to the output layer, and the formerly hidden layer to the latter one.

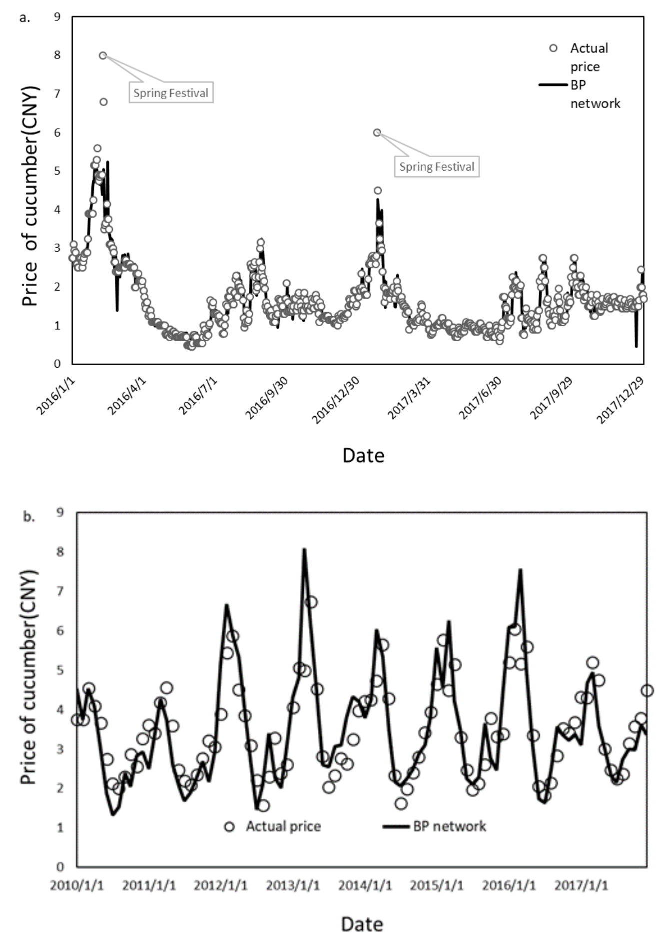


Fig. 3. (a) Actual and forecast daily price of cucumber in Beijing Xinfadi Market, from January 1, 2016 to January 1, 2018. (b) Actual and forecast monthly price of cucumber in Beijing, from January 2010 to January 2018. Lines: forecasting prices in a BP network method. Circles: actual prices.

We use the RNN to forecast price to see the differences among the three methods. The price of the previous period is used as input, and the future price is set as output. The number of the input nodes and hidden layer nodes is same as that of the BP network, and the number of the output nodes is set to 1.

IV. RESULTS

A 2-year daily average price data of cucumber in Beijing Xinfadi Market [Fig. 3(a)] and a 7-year monthly average price data of cucumber in Beijing [Fig. 3(b)] are given. As shown in Fig. 3(a), the price of cucumber has a strong periodicity throughout the year. Generally, the price of cucumber peaks in January or February of each year, due to the Spring Festival and low-temperature weather, and then gradually reaches the minimum around July. The data also show that daily average price always fluctuates [Fig. 3(b)].

A. Monthly Average Price Forecasting

As shown in Fig. 4, the monthly average price is forecasted with the ARIMA model and BP network method. The average

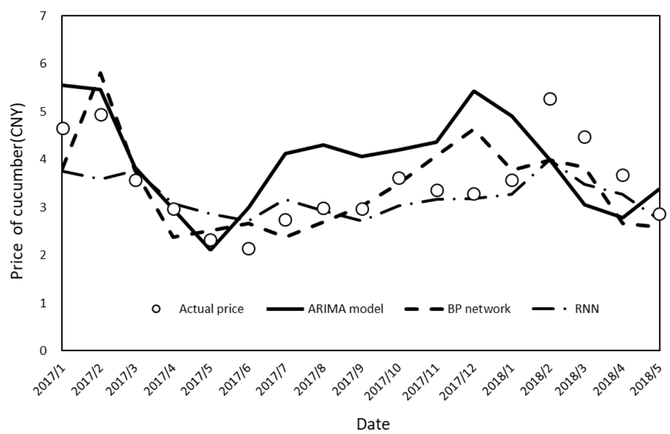


Fig. 4. Actual monthly average price and forecasting price of cucumber in Beijing from January 2017 to May 2018. Solid lines: forecasting prices in ARIMA model. Dotted lines: forecasting prices in a BP network method. Dashed-dotted lines: forecasting prices in RNN method. Circles: actual prices.

absolute error is 0.97 Chinese Yuan (CNY) and the average relative error is 28.53% with ARIMA model, whereas the average absolute error is 0.56 CNY and the average relative error is 15.6% with the BP network method. In addition, the average absolute error is 0.49 CNY and the average relative error is 13% with the RNN method. Compared with the results of the ARIMA model, the BP network method and RNN method have higher accuracy and they can predict the trend of price fluctuation more accurately. The ARIMA model performs well in the next 5 months, with the average absolute error being 0.37 CNY from January 2017 to May 2017, which increases rapidly after May 2017. The result indicates that the ARIMA model can forecast the trend well in a short-time period, but it cannot keep the accuracy for a long-time period.

B. Weekly Average Price Forecasting

The average absolute error is 0.47 CNY and the average relative error is 33.2% with the ARIMA model, whereas the average absolute error is 0.22 CNY and the average relative error is 15.4% with the BP network method. In addition, the average absolute error is 0.21 CNY and the average relative error is 14.4% with the RNN method.

As shown in Fig. 5, the ARIMA model can well predict the price of following $3-4$ weeks. The weekly average price data have a certain periodicity throughout the year, which is consistent with the monthly average periodicity. However, for some specific months, such as the Spring Festival, due to the variation of its date for different year, which may be in the first 10 days of February or in the second 10 days of February, the monthly average price in Spring Festival remains stable and reaches the highest price during the year, but the weekly average price will fluctuate greatly. The peak of the weekly average price may appear in any week in February, with a big variation. Also, the price of cucumber gets to the valley in June or July, which may appear in any week in June or July. Therefore, the time series model cannot well predict the weekly average price.

Similar to the monthly average price prediction, the forecasting results of BP network and RNN are more accurate. The two methods also perform better than ARIMA model in forecasting the fluctuation of the weekly average price.

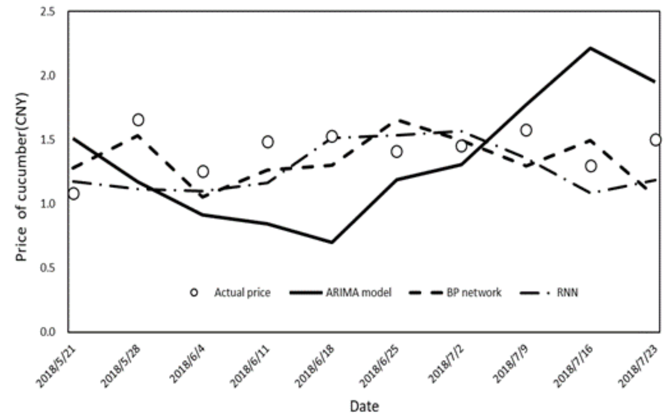


Fig. 5. Actual forecast weekly price of cucumber in Beijing Xinfadi Market from May 21, 2018 to July 23, 2018. Solid lines: forecasting prices in ARIMA model. Dotted lines: forecasting prices in a BP network method. Dashed-dotted lines: forecasting prices in RNN method. Circles: actual prices.

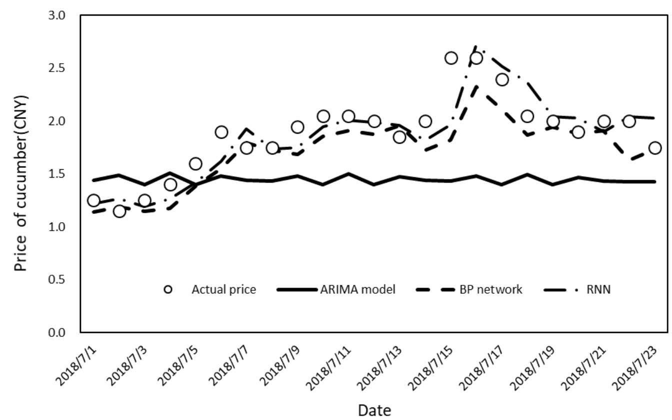


Fig. 6. Actual and forecast daily price of cucumber in Beijing Xinfadi Market from July 1, 2018 to July 23, 2018. Solid lines: forecasting prices in ARIMA model. Dotted lines: forecasting prices in a BP network method. Dashed-dotted lines: forecasting prices in RNN method. Circles: actual prices.

C. Daily Average Price Forecasting

The average absolute error is 0.55 CNY and the average relative error is 38.53% with the ARIMA model, whereas the total average absolute error is 0.3 CNY and the average relative error is 17.47% with BP network method. It is worth mentioning that the average absolute error is only 0.15 CNY and the average relative error is 8.82% with RNN method.

The daily average price predicted using the ARIMA model has large errors (Fig. 6), which indicate that the ARIMA model cannot predict the trend of the daily prices. This is because the daily average price fluctuates greatly and the data does not have obvious periodicity. Within a month, the daily average price often appears alternately a slight increasing or declining, the obvious difference in price variation can only be found according to the data of several months. When it comes to the whole year, the price actually will not vary too much. Thus, the ARIMA model will keep the slightly changing trend. This leads to the consistent trend for the forecasting values.

TABLE III
ERROR OF DAILY AVERAGE PRICE PREDICTION IN BEIJING XINFADI MARKET

Date	Average absolute error (CNY)	Average relative error
First day	0.17	10.70%
Second day	0.25	14.53%
Third day	0.26	15.20%
Forth day	0.3	17.05%
Fifth day	0.33	19.09%
Sixth day	0.37	22.01%
Seventh day	0.4	23.72%
Average	0.3	17.47%

The forecasting results of BP network are more accurate than that of the ARIMA model. The forecasting error increases with the increase of days (Table III), which is more accurate for the first 3 days, and the relative error of the first day is just 10.7%. In daily average price forecasting, neural network method succeeds to keep the accuracy even the data fluctuate a lot. The larger scale of data contributes to the accuracy of forecast results.

In addition, the BP network performed worse when we set the number of output nodes larger. We set the number of the output nodes to 12 – 15 for monthly and weekly average price forecasting, the relative errors increase to 25% – 30%. For daily average price forecasting, the error increases to 40%.

RNN method performs better than the other two methods which show the superiority of the deep learning algorithms. In the forecasting experiments of monthly and weekly average prices, it has not achieved an evidently better result than the BP network method. However, due to the increase of training data scale, the performance of the RNN method is further improved.

V. DISCUSSION

In this study, we predicted the monthly, weekly, and daily average prices of agricultural product, taking cucumber as an example. The results indicate that the accuracy of price forecasting of neural network methods including BP network and RNN are higher than that of the ARIMA model. ARIMA model can be used to predict the periodical, but it cannot predict fluctuations affected by sudden factors. Furthermore, the longer the prediction time period, the higher the accuracy of prediction is for the ARIMA model and BP network. The accuracy of monthly average price forecast is higher than that of daily average price forecast. The price trend of tomato and eggplant is very similar to that of cucumber. The forecasting results also support our conclusion.

ARIMA model has some limitations in training, which requires continuous periodic data of certain plant cultivar from the same place. The neural network methods can improve its generalization ability by training price data of different source. With the wide application of Web service, the price information of agricultural products will become more standardized and the data scale will become larger. Therefore, the method of deep learning represented by the neural network will become the mainstream method of agricultural product price forecasting.

According to the simulation results, the prediction of price can provide some guidance for production and harvesting plan. The growth period of cucumber lasts several months from planting to harvesting, thus the forecast of monthly average price can provide guidance for the planting time of cucumber. During the fruit harvesting period of cucumber, through price forecast with finer temporal scale (by day), the harvesting time of cucumber can be more accurately proposed to have a good benefit.

The source of price data is crucial for the price prediction of agricultural products. Through the investigation of agricultural market websites in Beijing, Shanghai, and Guangzhou, currently, many websites do not have updated price data of agricultural products for a long time. In addition, some problems exist, such as poor data quality, serious data loss and faked data. Regarding such problem, popular deep learning algorithms may not fit for price forecasting of agricultural products, which needs to be solved further.

In the future, the neural network methods, especially the RNN method, used in this paper should be improved by taking into account social factors (major holidays and policies), weather factors, market factors, and other new features [25], [26]. In addition, it will be interesting to collect the price data of other agricultural products and multiple regions to train the network to improve the prediction accuracy and generalization ability. Combined with specific cases, using the results of price forecasting, nonlinear programming can be designed to guide the production and harvesting of agricultural products [2], [27]. Parallel learning can be applied to train the computation system according to feedback from a real situation [28], [29].

VI. CONCLUSION

In this paper, we collected price data of cucumber in Beijing by web crawler for price forecasting. Monthly, weekly, and daily average price forecasting were performed, respectively, through the ARIMA model, BP network method, and RNN method. The results show that the RNN network method has higher accuracy. ARIMA model does have high accuracy in forecasting when the data have strong periodicity. Thanks to the expansion of data scale brought by the social sensor (web crawler), the deep learning methods bring the promising result of price forecast which is very helpful in crop planning.

REFERENCES

[1] A. Heydon and M. Najork, “Mercator: A scalable, extensible Web crawler,” *World Wide Web*, vol. 2, no. 4, pp. 219–229, 1999.

[2] M. Kang, X.-R. Fan, J. Hua, H. Wang, X. Wang, and F.-Y. Wang, “Managing traditional solar greenhouse with CPSS: A just-for-fit philosophy,” *IEEE Trans. Cybern.*, vol. 48, no. 12, pp. 3371–3380, Dec. 2018.

[3] S. Li, S. Zhao, Y. Yuan, Q. Sun, and K. Zhang, “Dynamic security risk evaluation via hybrid Bayesian risk graph in cyber-physical social systems,” *IEEE Trans. Comput. Social Syst.*, vol. 5, no. 4, pp. 1133–1141, Dec. 2018.

[4] F.-Y. Wang, “Computational social systems in a new period: A fast transition into the third axial age,” *IEEE Trans. Comput. Social Syst.*, vol. 4, no. 3, pp. 52–53, Sep. 2017.

[5] T. Park, “Forecast evaluation for multivariate time-series models: The U.S. cattle market,” *Western J. Agricult. Econ.*, vol. 15, pp. 133–143, Jul. 1990.

[6] J. Zong and Q. Zhu, "Apply grey prediction in the agriculture production price," in *Proc. 4th Int. Conf. Multimedia Inf. Netw. Secur.*, 2013, pp. 396–399.

[7] C. W. J. Granger and R. Joyeux, "An introduction to long-memory time series models and fractional differencing," *J. Time Ser. Anal.*, vol. 1, no. 1, pp. 15–29, 1980.

[8] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *J. Econ.*, vol. 31, no. 3, pp. 307–327, 1986.

[9] C. Carter and G. C. Rausser, "Commodity futures market efficiency in the soybean complex," *Rev. Econ. Statist.*, vol. 65, no. 3, pp. 469–478, 1983.

[10] G. K. Jha and K. Sinha, "Agricultural price forecasting using neural network model: An innovative information delivery system," *Agricult. Econ. Res. Rev.*, vol. 26, no. 26, pp. 229–239, 2013.

[11] P. Areekul, T. Senjyu, H. Toyama, and A. Yona, "A hybrid ARIMA and neural network model for short-term price forecasting in deregulated market," *IEEE Trans. Power Syst.*, vol. 25, no. 1, pp. 524–530, Feb. 2010.

[12] G. Qiang, C.-S. Luo, and Q.-F. Wei, "Prediction and research on vegetable price based on genetic algorithm and neural network model," *Asian Agricult. Res.* , vol. 3, pp. 148–150, Jan. 2011.

[13] G. E. P. Box, *Time Series Analysis: Forecasting and Control* (Holden-day Series in Time Series Analysis), G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, Eds. San Francisco, CA, USA: Holden-day, 1994, pp. 199–201

[14] C. S. Luo, L. Y. Zhou, and Q. F. Wei, "Application of SARIMA model in cucumber price forecast," *Appl. Mech. Mater.*, vols. 373–375, pp. 1686–1690, Aug. 2013.

[15] H. Wu, H. Wu, M. Zhu, W. Chen, and W. Chen, "A new method of large-scale short-term forecasting of agricultural commodity prices: Illustrated by the case of agricultural markets in Beijing," *J. Big Data*, vol. 4, p. 1, Jan. 2017.

[16] G. Zhang, B. E. Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks:: The state of the art," *Int. J. Forecasting*, vol. 14, pp. 35–62, Mar. 1998.

[17] S. Madisetty and M. S. Desarkar, "A neural network-based ensemble approach for spam detection in Twitter," *IEEE Trans. Comput. Social Syst.*, vol. 5, no. 4, pp. 973–984, Dec. 2018.

[18] G. K. Jha, P. Thulasiraman, and R. K. Thulasiram, "PSO based neural network for time series forecasting," in *Proc. Int. Joint Conf. Neural Netw.*, 2009, pp. 1422–1427.

[19] I. Kaastra and M. Boyd, "Designing a neural network for forecasting financial and economic time series," *Neurocomputing*, vol. 10, no. 3, pp. 215–236, 1996.

[20] S. Shanthi, P. Balamurugan, and D. Kumar, "Performance comparison of featured neural network with gradient descent and levenberg-marquart algorithm trained neural networks for prediction of blood glucose values with continuous glucose monitoring sensor data," in *Proc. Int. Conf. Emerg. Trends Sci., Eng. Technol.*, 2012, pp. 385–391.

[21] R. Venkatesan and B. Balamurugan, "A real-time hardware fault detector using an artificial neural network for distance protection," *IEEE Trans. Power Del.*, vol. 16, no. 1, pp. 75–82, Jan. 2001.

[22] R. J. Williams and D. Zipser, "A learning algorithm for continually running fully recurrent neural networks," *Neural Comput.*, vol. 1, no. 2, pp. 270–280, 1989.

[23] K. Cho *et al.* (2014). "Learning phrase representations using RNN encoder-decoder for statistical machine translation." [Online]. Available: https://arxiv.org/abs/1406.1078

[24] Y. Miao, M. Gowayyed, and F. Metze. (2015). "EESEN: End-to-end speech recognition using deep RNN models and WFST-based decoding." [Online]. Available: https://arxiv.org/abs/1507.08240

[25] X. Deng, X. Tian, S. Chen, and C. J. Harris, "Nonlinear process fault diagnosis based on serial principal component analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 3, pp. 560–572, Mar. 2018.

[26] C. Worasucheep, "A stock price forecasting application using neural networks with multi-optimizer," in *Proc. IEEE Int. Workshop Comput. Intell. Appl.*, Nov. 2016, pp. 63–68.

[27] A. Gargano and A. Timmermann, "Forecasting commodity price indexes using macroeconomic and financial predictors," *Int. J. Forecasting*, vol. 30, no. 3, pp. 825–843, 2014.

[28] S. Wang, X. Wang, P. Ye, Y. Yuan, S. Liu, and F.-Y. Wang, "Parallel crime scene analysis based on ACP approach," *IEEE Trans. Comput. Social Syst.*, vol. 5, no. 1, pp. 244–255, Mar. 2018.

[29] J. J. Zhang *et al.*, "Cyber-physical-social systems: The state of the art and perspectives," *IEEE Trans. Comput. Social Syst.*, vol. 5, no. 3, pp. 829–840, Sep. 2018, doi: 10.1109/TCSS.2018.2861224.

Yuchen Weng received the B.E. degree in electrical engineering from Agricultural University, Beijing, China, in 2017. He is currently pursuing the master's degree with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing.

His current research interests include data mining and parallel agriculture.

Xiujuan Wang received the Ph.D. degrees in applied mathematics from Ecole Central, Écully, France, and in soil resource use from China Agricultural University, Beijing, China, in 2011.

She is currently an Assistant Professor with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing. Her current research interests include plant growth modeling and parallel agriculture.

Jing Hua received the Ph.D. degree in applied computer science from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2012.

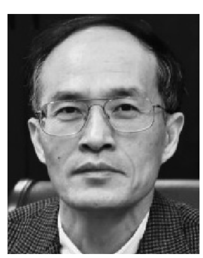
He is currently an Assistant Professor of computer science with the Institute of Automation, Chinese Academy of Sciences. His current research interest includes computer science and agriculture, including virtual plant modeling, smart agriculture, programming languages, distributed computer systems, and computer graphics.

Haoyu Wang received the Ph.D. degree in pattern recognition and intelligent system from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2014.

He is currently an Assistant Professor of computer science with the Institute of Automation, Chinese Academy of Sciences. His current research interest includes computer science and agriculture, including virtual plant modeling, smart agriculture, programming languages, and information system.

Mengzhen Kang (M'16) received the Ph.D. degree in pattern recognition and intelligent systems from the Institution of Automation, Chinese Academy of Sciences, Beijing, China, in 2003.

From 2005 to 2006, she was a Post-Doctoral Fellow with Projet DigiPlante, INRIA, Rocquencourt, France, and high-performance computing, Wageningen University, Wageningen, The Netherlands. She is currently an Associate Professor with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences. Her current research interests include parallel agriculture and computational biology.



Fei-Yue Wang (S'87–M'89–SM'94–F'03) received the Ph.D. degree in computer and systems engineering from Rensselaer Polytechnic Institute, Troy, NY, USA, in 1990.

He is currently the Director of the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China. His current research interests include methods and applications for parallel systems, social computing, and knowledge automation.

Dr. Wang is an Elected Fellow of INCOSE, IFAC, ASME, and AAAS. He was a recipient of the 2nd Class National Prize in Natural Sciences of China in 2007, the Outstanding Scientist by ACM for his research in intelligent control and social computing, the IEEE ITS Outstanding Application and Research Awards in 2009 and 2011, and the IEEE SMC Norbert Wiener Award in 2014. Since 1997, he has been serving as the General or the Program Chair for more than 20 IEEE, INFORMS, ACM, and ASME conferences. Since 2008, he has been the Vice President and the Secretary General of the Chinese Association of Automation. He is currently an Editor-in-Chief of *China's Journal of Command and Control*.