Agricultural Product Price Prediction Based on Improved Temporal Convolutional Network

Written by 林海, 吴连伟, 仲仁豪, 徐心怡, 胡鹏
School of Informatics Xiamen University

Abstract

As the price of agricultural product indirectly reflects the relationship between supply and demand of the agricultural market, so predicted price could be a reference of farmers and the country to modify their strategy to smooth the impact from the market. The Agricultural product price prediction has been widely studied in recent years, but it generally stays in the stage of traditional linear regression analysis or deep learning model such as LSTM neural network. Therefore, combined with the latest research on time series prediction in deep learning, we apply the temporal convolutional network TCN model to agricultural product price prediction. We also propose an improved TCN neural network model, which introduces LSTM neural network on the basis of TCN model. Thanks to the gate mechanism of LSTM cell, the model's control to long-term memory is strengthened. Compared with TCN, LSTM neural network and GRU neural network, all single models, we evaluate the TCN stacked LSTM model and results shows that it gets better performance in some cases.

Introduction

Agricultural products market is an unseparatable part of China's market economic system. The stability of agriculture is the cornerstone of China's economic development. A common problem in agricultural production is the waste of resources and the loss of benefits caused by too much or too little planting. It is mainly due to the failure to follow the Law of market economy. This would be greatly improved if farmers were given prices over a period of time as a reference for farming and the a signal of supply-demand balance. So it is a meaningful study to forcast the price of agricutual products.

At present, with the development of big data and artificial intelligence, more agricultural product data is stored as a format of electronic files which will help further data mining and analysis. Agricultural product price data is a special sub-category of time series, which has many influencing factors. Natural disasters, supply, weather, national policies, international exchange rates, epidemics and other factors are all essential factors that affect agricultural product prices. For example, at the beginning of this year, the world's economy was affected by COVID-19, which also impacted on agricultural products. However, this unpredictable factor is currently difficult to predict or quantify. How to accurately predict the price trend of agricultural products in a certain period of time in the future to provide farmers references is one of the most challenging agricultural problems at present.

In the past, the empirical estimation method, survey analysis method and post-statistical analysis method were usually used to predict the price of agricultural products. However, the first two methods were too subjective, while the traditional statistical analysis method applied to this problem was often based on the assumption that "the time series of the price of agricultural products is linear", while obviously the market is volatile. Now in the era of big data,the massive data makes it possible to use deep learning methods. Many scholars have focused their research on predicting the prices of agricultural products. They tried different methods such as gray theory, RBF neural network, etc. (陈佳珊 and 张丹 2019; 刘锦源 2019; 罗洪奔 2014; 景秋玉 2018) . They rarely conduct research on the analysis of agricultural product price trends in combination with agricultural product factors. From data acquisition, influencing factor analysis to combined factor forecasting process, this problem is extremely challenging. There are many options when predicting the trend of time series, which can be roughly divided into the following categories, statistical methods, machine learning methods and deep learning methods. For example, Kim et al. proposed an improved SVM model, which used polynomial kernel function and Gaussian radial basis function as the kernel function, and used 12 indicators as inputs to study the Korea's Comprehensive Stock Price Index (KOSPI)(Kim 2003).

Related Work

The forecast of agricultural product price trend is similar to the classical time series problem, which has transited from the first qualitative analysis to the later quantitative analysis. The analysis basis of problems has gradually changed from the original subjective ideas based on experience and feeling to the application of big data and scientific statistical methods. Based on the past agricultural product price prediction models, it can be divided into the traditional statistical analytical model and the machine learning prediction algorithm from the persepective of algorithm, whereas from the perspective of model components, it can be divided into into single prediction model, combined prediction model and integrated prediction model(MacQueen 1967).

Traditional statistical analysis models composed of linear theory and nonlinear theory, and linear theory models are generally based on the assumption that the past sequence is linear. In 2013, Guihong Wang and other researchers used 8 methods respectively, including average prediction method and timing average growth quantity prediction method, to conduct price prediction and experimental comparative study on 13 types of agricultural products(Zhang 2003). However, the price sequence of agricultural products cannot always form a simple linear relationship or a linear superposition. Since the price sequence of agricultural products has a large number of potential factors and great uncertainties, the linear prediction method would assert certain limitations and thus is not applicable.

As researchers continue to take the influence factors on agricultural products into account, the application of nonlinear theoretical models in this problem has been developed. In 2012, Ganqiong Li proposed to divide the impact factors of agricultural products into strong fluctuation factors and volatile fluctuation factors. Li took vegetables as an example to establish non-parametric kernel density estimation method, multi-layer feedforward neural network and other models. The research results showed that the prediction accuracy was above 90% (李干琼 2012).

Later, the intelligent prediction algorithm represented by machine learning and deep learning algorithm gradually becomes the mainstream algorithm due to their higher prediction accuracy and generalization ability(刘锦源 2019; Arthur and Vassilvitskii 2006; 陈佳珊 and 张丹 2019; Olah 2015; 姚缙然 2019). Jinyuan Liu introduced THE EEMD method to improve the LSTM model for predicting the futures prices of agricultural products. This method is more accurate than the traditional machine learning method (LSTM and SVR) (刘锦源 2019). Ning Jia , Chunjun Zheng further combined the convolutional neural network, Long-Short Term Memory model and the attention mechanism. Compring with the traditional single model, the combined model not only improved the accuracy, but also predict the overall trend of vegetable products in the coming week accurately (贾宁 and 郑纯军 2019).

Reseachers found through experiments that both the traditional statistical analysis model have complex factors that would affect the model, and thus it is not sufficient to only rely on a single prediction model. The combined model and the integrated model have thereby been proposed. Different models are chosen and combined according to certain weight. Jiashan Chen and other researchers predicted the agricultural product price index for the next five quarters. They extracted the linear information from SARIMA and nonlinear information with LSSVM. The result showed that the combined model had higher robustness and prediction accuracy than the single model (陈佳珊 and 张丹 2019). To achieve better forecasting for the output of grain, Jinran Yao attempted to obtain a combination forecasting model that has a wider range of application and better stability (姚缙然 2019).

The combination model slices the research object horizontally followed by its superposition, while the corresponding integrated prediction model slices the research object vertically, followed by divide and conquer and reintegration. Lin Chen studied the ARIMA-based prediction model, neural network prediction model and VAR model to predict the settlement price of cotton and other futures on Zhengzhou Commodity Exchange, which achieved satisfactory results (陈林 2011).

At present, more researchers are trying to apply deep learning to solve the problem. As a common tool in artificial intelligence, deep learning technology is also applied on time series problems, which solves many practical issues including the price prediction of stocks and horticultural plants. In literature(Gudelek, Boluk, and Ozbayoglu 2017), convolutional neural network (CNN) is used to predict the stock price fluctuations in a two-dimensional space. Zhang adopted a hybrid framework of ARIMA model and neural network model to solve the time series problem(Zhang 2003). In literature(Chang, Zhang, and Chen 2018), Long Short-Term Memory (LSTM) is used to predict the price of electricity. Therefore, in order to obtain better results than the previous prediction algorithm, we will use four different deep learning models to study the prediction of agricultural product price trend.

Deep Learning Models

In this section, we'll introduce four different deep learning models we implement in our experiences. They are LSTM neural network, GRU neural network, TCN(temporal convolution neural network) and TCN stacked LSTM neural network respectively.

LSTM neural network

In traditional recurrent neural networks, two issues brought great affect on model performance: (1)Models have difficulty learning long-term memory, for they use a relatively single mechanism to deliver hidden state which makes information from long time ago easily forgotten;(2) The gradient blows up or decays exponentially over time due to a scaling effect on the gradient of the loss function of the neural network. The two issues above make traditional recurrent neural networks a fair performance on long-term memory tasks and hard to train.

Hochreiter S et al.(Hochreiter and Schmidhuber 1997) redesigned the cell of recurrent neural network and introduced LSTM neural network, which have been widely used in time series problems in recent years. Thanks to its 'gate mechanism' in LSTM cell, it perfected the shortcomings of classical original RNN models. The cell structure is depicted in Figure1.

A LSTM cell can be a described as a collection of three gates: input gate ,forget gate and output gate. The input gate receives the input and serves as a filter to controls update information $u^{(t)}$.

$$i^{(t)} = \sigma\left(W_{ih}h^{(t-1)} + W_{ix}x^{(t)} + b_i\right) \tag{1}$$

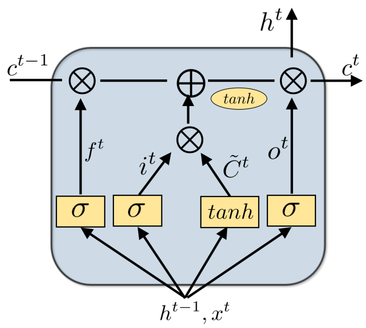


Figure 1: The architecture of long-short term memory cell
.

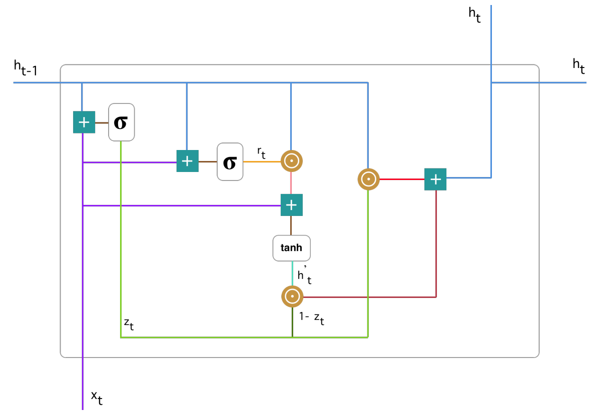


Figure 2: The architecture of gated recurrent unit
.

$$u^{(t)} = \tanh\left(W_{ch}h^{(t-1)} + W_{cx}x^{(t)} + b_c\right) \quad (2)$$

Where $x^{(t)}$ is the input at time step t , h^{t-1} denotes the hidden state of last time step and σ represents sigmoid activation function.

The forget gate decides the extent cell at time step t wil forget the information in c^{t-1}, in which the memory of previous time steps stored.

$$g^{(t)} = \sigma\left(W_{fh}h^{(t-1)} + W_{fx}x^{(t)} + b_g\right) \quad (3)$$

Then the cell computes the c^t based on the above two gates.

$$c^{(t)} = g^{(t)} \odot c^{(t-1)} + i^{(t)} \odot u^{(t)} \quad (4)$$

The output gate finally expose a part of cell state $c^{(t)}$ as the hidden state of time step t.

$$o^{(t)} = \sigma\left(W_{oh}h^{(t-1)} + W_{ox}x^{(t)} + b_o\right) \quad (5)$$

$$h^{(t)} = o^{(t)} \odot \tanh\left(c^{(t)}\right) \quad (6)$$

With the help of 'gate mechanism' and cell state $c^{(t)}$, LSTM neural networks are able to learn to preserve long-term memory.

GRU neural network

Introduced by Cho, et al. in 2014 (Cho et al. 2014), GRU (Gated Recurrent Unit) aims to solve the vanishing gradient problem which comes with a standard recurrent neural network. GRU can also be considered as a variation on the LSTM because both are designed similarly and, in some cases, produce equally excellent results.

To solve the vanishing gradient problem of a standard RNN, GRU uses, so-called, update gate and reset gate. Basically, these are two vectors which decide what information should be passed to the output. The special thing about them is that they can be trained to keep information from long ago, without washing it through time or remove information which is irrelevant to the prediction. The architecture of GRU is shown in Figure 2.

The update gate z_t is calculated using the fomula:

$$z_t = \sigma\left(W^{(z)}x_t + U^{(z)}h_{t-1}\right) \quad (7)$$

When x_t is plugged into the network unit, it is multiplied by its own weight $W(z)$. The same goes for $h_(t-1)$ which holds the information for the previous $t-1$ units and is multiplied by its own weight $U(z)$. Both results are added together and a sigmoid activation function is applied to squash the result between 0 and 1.

The reset gate is used from the model to decide how much of the past information to forget. To calculate it, we use:

$$r_t = \sigma\left(W^{(r)}x_t + U^{(r)}h_{t-1}\right) \quad (8)$$

Then we take the usage of reset gate and calculate a new memory content h_t, in which reset gate is to restore the relevant information of the past:

$$h'_t = \tanh\left(Wx_t + r_t \odot Uh_{t-1}\right) \quad (9)$$

Finally, based on the update gate and the introduced new memory content, we'll filter the content from the past the new memory content, which means gettting useful from the past and current with the help of update gate and combined as a new content vector. The fomula is:

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot h'_t \quad (10)$$

By introducing two gates, the architecture of GRU neural network is simpler. In some cases can perfrom as well as LSTM neural network which has three gates and save training time faced with large samples.

Temporal Convolution Networks

The TCNs model was firstly designed in 2018(Bai, Kolter, and Koltun 2018) as a simple sequence predicting architechture. The distinguishing characteristics of TCNs are: 1) The convolutions in the architecture are causal; 2) the architecture can take a sequence of any length and map it to an output sequence of the same length, just as with an RNN.

Beyond this, the TCNs build very long effective history sizes by using a combination of very deep networks and dilated convolutions without gating mechanisms like above two recurrent neural network models.

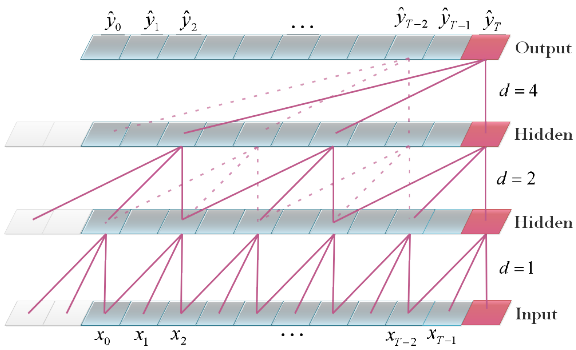


Figure 3: A dilated causal convolution with dilation factors d = 1, 2, 4 and filter size k = 3. The receptive field is able to cover all values from the input sequence.

.

Causal Convolutions The TCN is based on two principles: the fact that the network produces an output of the same length as the input; the fact that there can be no leakage from the future into the past. The accomplish the first point, the TCN uses a 1D fully-convolutional network (FCN) architecture(Long, Shelhamer, and Darrell 2015). To achieve the second point, the TCN uses causal convolutions, convolutions where an output at time t is convolved only with elements from time t and earlier in the previous layer.

The above design can be describe as: TCN = 1D FCN + causal convolutions.

Dilated Convolutions But how TCNs preserve long history? The solution in TCN is the dilated convolutions. Formally, for a 1-D sequence input $\mathbf{x} \in \mathbf{R}^n$ and a filter $f : \{0, \ldots, k-1\} \to \mathbf{R}$, the dilated convolution operation F on element s of the sequence is defined as:

$$F(s) = (\mathbf{x} *_d f)(s) = \sum_{i=0}^{k-1} f(i) \cdot \mathbf{x}_{s-d \cdot i} \tag{11}$$

Where d is the dilation factor, k is the filter size, and $s - d \cdot i$ accounts for the direction of the past. We dipicted an example of dilated convolution architecture in Figure 3.

Either we choose larger filter size k and increasing the dilation factor d will expand the receptive field of the TCN. When the layers go higher, the increase of dilation factor d ensures the filter of the layer hits the input from long history; the larger filter size k will ensure the filter catch the information from in relatively long history from previous layer, which will also expand the receptive field.

Residual Connections As the TCN's receptive field depends on the network depth n as well as filter size k and dilation factor d, stabilization of deeper and larger TCNs becomes important. Due to the effectiveness the residual block (He et al. 2016) helps layers to learn modifications to the identity mapping rather than the entire tranformation, which has repeatedly been shown to benefit very deep net- works, the TCN employ a generic residual module in place of a convolutional layer.

The mechanism of residual block can be shown as:

$$o = a(\mathbf{x} + \mathbf{F}(\mathbf{x})) \tag{12}$$

Where $\mathbf{x}$ is the input of the block, the a is the active function, and the $\mathbf{F}$ denotes a series of transformations.

TCN stacked LSTM

Given both LSTM neural network and TCN perform well in some cases and are brought into various applications, we combine the two models into a new one named TCN stacked LSTM.

On the basis of TCN, we add LSTM layer on it. The stacked model contains CNN and RNN, we designed the model to extract and combine the advantages from them. The detail of these two models have been descripted above.

Experiments

Data Collection

In our work, we are not using open source dataset but collect history data by outselves. The factors affecting the agricultural product price we considered are temperature ,impact on production, and exchange rate, impact on sales. The data range from 2014-01-01 to 2020-10-20. The following websites are where we crawl data:

- **temperature** http://lishi.tianqi.com/
- **exchange rate** https://srh.bankofchina.com/search/whpj/search_cn.jsp
- **fruit price** http://nc.mofcom.gov.cn/channel/jghq2017

Totally we collect history price data of five fruit categories in some Beijing agricultural wholesale markets, the fruits contain Cantaloupe, Strawberry, Pineapple, Fuji Apple, Mango. The temperature consists of highest temperature, lowest temperature and average temperature for each day.

Preprocessing

Due to the exchange rate is in hours, we average the data and observe the change in a day.

In addition, for the benefit of stability in training, we use min-max transformation to normalize the factors. The min-max transformation can be shown as:

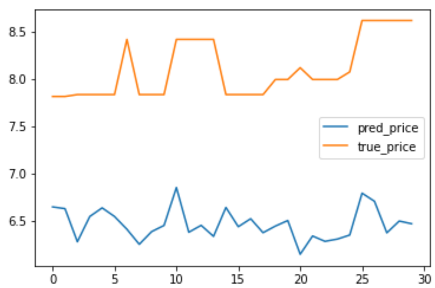
$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{13}$$

Finally, we leave out data of last 30 days for evaluating the performance of models.

Training

All the networks are trained using adam optimizer and Mean Square Error(MSE) loss function and Mean Absolute Error(MAE) is used to evaluate the models. We train models for 500 epochs with a batch size of 30. In the training process, we use 10% of the data as the validation set. When 40 epochs validation set loss is not improved, training will be stopped. After training, we save loss values and models.

(a) strawberry



(b) cantaloupe

Figure 4: LSTM's prediction results for strawberry and cantaloupe prices

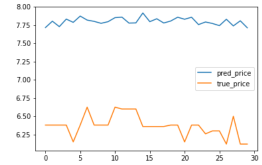
类别	GRU	LSTM	TCN	Improved TCN
草莓	83.151	19.329	12.924	**6.448**
哈密瓜	3.136	2.802	4.639	**2.114**
富士苹果	2.067	0.347	**0.082**	0.114
芒果	1.54	0.992	5.751	**0.942**
菠萝	0.359	0.163	0.135	**0.133**

Table 1: losses for four models in different fruits

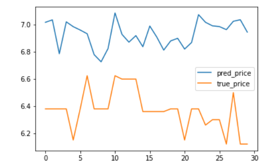
Evaluation and Comparison

We trained GRU, LSTM, TCN and TCN stacked LSTM models for each fruit and then using the test data set to evaluate different models. We calculate the mean square error of different models on different fruit test sets, which is convenient for comparison between models.The losses of training are shown in Table1. On the strawberry and mango data sets, LSTM performs better; on other data sets, TCN performs better. Although LSTM performs best on the strawberry and cantaloupe data sets, the gap between the predicted price and the real price is still large. Figure 4 shows LSTM's prediction results for strawberry and cantaloupe prices. The horizontal axis is the number of days in the future, and the vertical axis is the price.

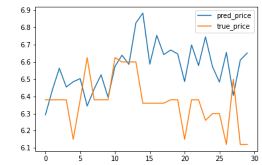
For TCN, compared with other models, its prediction of Fuji apple price is very close to the real price. Figure 5 shows four models' prediction results for Fuji apple. For GRU and TCN stacked LSTM, the difference between the predicted price of different fruits and the real price is large, about 2 to 3 yuan. In general, the performance of TCN is slightly better, but there is still a certain gap with the real price.



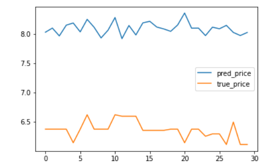
(a) GRU



(b) LSTM



(c) TCN



(d) TCN stacked LSTM

Figure 5: four models' prediction results for Fuji apple

Discussion

The research on agricultural product price trend prediction combined with deep learning is complicated and diverse, but it is difficult to find a recognized generalization ability and high accuracy algorithm model. This paper studies the network structure of recurrent neural network LSTM and time neural network TCN, introduces the advantages of LSTM on the basis of TCN, and proposes an improved TCN model.Compared with TCN, LSTM neural network and GRU neural network, we evaluate the TCN stacked LSTM model and results shows that it gets better performance in some cases.

Although we have implemented an algorithm model with better performance in agricultural product price prediction, we also encountered some shortcomings and regrets in the research process, which may be provided as a future work direction.

Improvement of data set Although we can easily obtain the price data of various agricultural products with the help of the Internet, it is not enough to obtain the price data of agricultural products. There are many factors that affect the prices of agricultural products, and the accumulation, acquisition, and mining of influencing factors are crucial. How to quantify natural factors such as disasters is also worth exploring in the future.

Improvement of algorithm designment Although the TCN, LSTM, and improved TCN models used in this paper perform better in agricultural product price prediction. But as research continues to deepen, better algorithm models, combined models, and integrated models will continue to emerge. This requires us to keep pace with the times and combine the latest research results. At the same time, the research on the adaptive optimal algorithm of model initial parameters is also a direction worth studying in the future.

References

Arthur, D.; and Vassilvitskii, S. 2006. k-means++: The advantages of careful seeding. Technical report, Stanford.

Bai, S.; Kolter, J. Z.; and Koltun, V. 2018. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271* .

Chang, Z.; Zhang, Y.; and Chen, W. 2018. Effective adam-optimized LSTM neural network for electricity price forecasting. In *2018 IEEE 9th international conference on software engineering and service science (ICSESS)*, 245–248. IEEE.

Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; and Bengio, Y. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078* .

Gudelek, M. U.; Boluk, S. A.; and Ozbayoglu, A. M. 2017. A deep learning based stock trading model with 2-D CNN trend detection. In *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, 1–8. IEEE.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Hochreiter, S.; and Schmidhuber, J. 1997. Long short-term memory. *Neural computation* 9(8): 1735–1780.

Kim, K.-j. 2003. Financial time series forecasting using support vector machines. *Neurocomputing* 55(1-2): 307–319.

Long, J.; Shelhamer, E.; and Darrell, T. 2015. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3431–3440.

MacQueen, J. 1967. Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. *Some methods for classification and analysis of multivariate observations* 281–297.

Olah, C. 2015. Understanding lstm networks .

Zhang, G. P. 2003. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* 50: 159–175.

刘锦源. 2019. 面向农产品期货价格预测的改进 LSTM 方法. 江苏科技信息 (27): 13.

姚缙然. 2019. 基于*Python*的粮食产量组合预测模型研究及应用. Ph.D. thesis.

景秋玉. 2018. 基于 *PCA-SVM-GARCH* 模型的股价预测. Master's thesis, 首都经济贸易大学.

李干琼. 2012. *SV* 因子分析框架下的农产品市场短期预测. Ph.D. thesis, 北京: 中国农业科学院.

罗洪奔. 2014. 基于灰色-ARIMA的金融时间序列智能混合预测研究. 财经理论与实践 (02): 29–36.

贾宁; and 郑纯军. 2019. 基于LSTM-DA神经网络的农产品价格指数短期预测模型. 计算机科学 046(0z2): 62–65,71.

陈佳珊; and 张丹. 2019. 基于SARIMA组合模型的农产品生产价格指数的短期预测. 时代金融 (30).

陈林. 2011. 基于非线性集成预测技术的期货价格预测研究. Ph.D. thesis, 福州大学.