Copper price forecasting

Abstract

This research has been completed to establish a 2-layer neural network with the hyper-parameter settings such as activation function, initializer, without or with the regularization term (L2), optimizer, epochs, hidden nodes, with or without learning rate decay schedule. Finally, the top five models which performed well in the in-sample data were ensembles to predict out-of-sample data. In this study, we tried a total of 927 model configurations and applied root mean square error (RMSE) to verify the effectiveness and accuracy. The empirical results show that the final forecasting model can obtain a favorable prediction effect in the time series data with higher volatility compared to those of the traditional economic method and traditional SVM techniques. The result also demonstrates the effectiveness of the proposed predictive techniques in revealing the underlying nonlinear patterns of copper prices.

Preprocessing

At this stage, we do time lag establishment, feature scaling, and train/test split. Firstly, according to previous literature on forecasting time series, time lag has a significant impact on forecasting results (Bouktif et al., 2018; Rubaszek et al., 2020). In this research, we consider the previous one to four weeks' historical copper prices as input factors respectively. Furthermore, the feature with the larger scale will dominate the other when the features are on different scales. To cope with this problem, we used min-max method to scale the features within a specific range to truly present the impact of each variable change on copper prices.

Finally, the predictive model is also very important to maintain a certain degree of effectiveness for out-of-sample. The convenient ratios of splitting samples such as 70:30%, 80:20%, and 90:10% (Alameer et al., 2019). From the perspective of data volume, this research divided the training and test data into the ratio of 80:20, in which the first 324 observations (from October 31, 2011 to February 13, 2019) were used for training the model, whereas the final 36 observations (from February 20, 2019 to January 1, 2021) were employed to test the performance and accuracy of the proposed model.

Experiment

In this section, we combined 927 model configurations of a 2-layer neural network based on activation function, initializer, without or with the regularization term (L2), optimizer, epochs, hidden nodes, with or without learning rate decay schedule. The configuration of Each hyperparameter is depicted in Table 1. Furthermore, the experimental result of the top five best models which effectively adapt to the in-sample data is shown in Table 2.

Table 1. A combination of hyperparameters for 2 layer neural network

Hyperparameter	Search space
Activation function	{sigmoid, tanh, relu}
Initializer	{small_random, Xavier}
Regularization term (L2)	$\lambda = \{0, 0.001, 0.0001\}$
Optimizer	{SGD, Adam, Momentum}
Epochs	{100, 200, 300}
Hidden nodes	{5, 8, 11}
Learning rate decay schedule	{none, Cosine decay}

Table 2. The experimental result of the top five best models

	Epochs	Activation function	Hidden nodes	Optimizer	Initializer	Regularization term (λ)	In-sample RMSE
M1	300	tanh	11	Adam	Xavier	0	0.0392
M2	300	relu	11	Adam	Xavier	0	0.0397
M3	300	relu	8	Adam	Xavier	0.001	0.0398
M4	300	relu	8	Adam	Xavier	0.0001	0.0400
M5	300	relu	8	Adam	Xavier	0	0.0405

From the Table 2, we can find that M1 to M5 have some similar characteristics, such as epoch and hidden nodes are the larger options in the search space, and most of models adopt relu as activation function. In addition, Adam, and Xavier are used for optimizer and weight initialization.

Table 3. In-sample and out-of-sample results from M1 to M5

	In-sample RMSE	Out-of-sample RMSE
M1	0.0392	0.0896
M2	0.0397	0.1465
M3	0.0398	0.0793
M4	0.0400	0.0679
M5	0.0405	0.0798

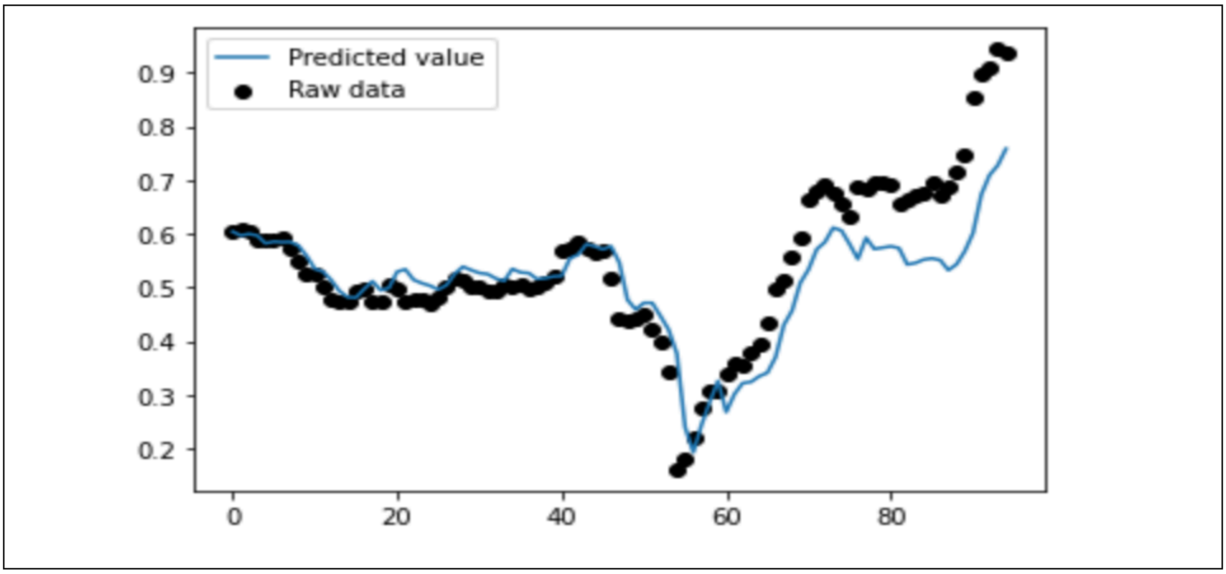

Fig 1. The prediction result of the ensemble model

Although the model performs well in the in-sample data, it may not have the same effect as reflected in the out-of-sample. In Table 3, we found that the M2 model obtained very good results in the in-sample data, but it performed abnormally in the out-of-sample. In order to reduce the effect of such uncertain predictions, the ensemble models are used to reduce the generalization error of the prediction. The study used the top five best models which performed well in the in-sample data to forecast the out-of-sample data. Then, the prediction results for each base model are averaged. The experimental results show that an error of RMSE =0.0856 is achieved, and it also demonstrates the prediction of the ensemble model for the unseen data is better than other 95% of model configurations, i.e., the prediction error is smaller than other 924 models. It corresponds to Kotu & Deshpande (2015) claimed that the base models are diverse and independent, the prediction error of the model decreases when the ensemble approach is used. Although such a model is composed of multiple base models, it acts and performs as a single model.

Reference

Alameer, Z., Elaziz, M. A., Ewees, A. A., Ye, H., & Jianhua, Z. (2019). Forecasting Copper Prices Using Hybrid Adaptive Neuro-Fuzzy Inference System and Genetic Algorithms. *Natural Resources Research*, *28*(4), 1385–1401.

Bouktif, S., Fiaz, A., Ouni, A., & Serhani, M. A. (2018). Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. *Energies*, *11*(7), 1636.

Kotu, V., & Deshpande, B. (2015). Data Mining Process. In *Predictive Analytics and Data Mining* (pp. 17–36). Elsevier.

Rubaszek, M., Karolak, Z., & Kwas, M. (2020). Mean-reversion, non-linearities and the dynamics of industrial metal prices. A forecasting perspective. *Resources Policy*, *65*, 101538.