Predicting the tendency of stock's price with factor-cross method based on deep learning

Zhujun Zhang*
Nanjing University of Information Science and Technology, Nanjing, China
gracejin@shinyway.com.cn

Abstract. In this paper, a solution for tendency prediction in the A-share market is presented. In data processing, the problem of non-normal factor distribution is solved by using compound processing method, and the case of non-independent distribution among factors is focused. The DNNLinearCombinedEstimator model in TensorFlow is applied to explore the first-order relationship between these factors. In order to intuitively understand the advantages of the model, an experiment is conducted between our solution and traditional machine learning methods. The result exhibits that the performance of our work is more excellent in all prediction metrics.

1. Introduction
The quantitative transaction is a subject that aims to select excess returns from huge historical data. In 1952, Markowitz initiated the theory of modern portfolio management[1]. In 1964, W.F.Sharpe proposed the capital asset pricing model(CAPM)[2]. In 1973, F.Black, and M.Scholes proposed the BS model for the pricing of financial derivatives[3]. After that, S.A.Ross et al put forward Arbitrage Pricing theory[4] and the multi-factor model is still one of the most widely used quantitative investment methods up to now.

With the development of computer science, many research efforts make a difference in the field of quantitative finance[5]. In recent years, it is also developing rapidly in China. In three quarters of 2020, the management amount of quantitative private placement has exceeded 500 billion yuan.[5].

Fig.1 Quantitative fund size and its proportion in equity funds[6]

Compared with other more mature financial markets, the A-share market still has many problems[5-7]. Thus, more significant risks are suffered by ordinary investors. To make a change, it is necessary to improve the capacity of stock selection and trading choice. Generally speaking, stock selection depends on a better stock pool. In addition, making a better trading choice is based on machine learning.
Stock selection is to choose stocks with greater potential for earnings. Therefore, the high-quality’s “blue-chip” stocks and “white horse” stocks in the A-shares market is considered in our research. In trading, the processes of data-set acquisition, factor preprocessing, modeling and performance are focused on.

2. Data preparation
In data collection, the stock ids and historical opening and closing prices are got to make the corresponding label according to whether the closing price is rise or fall within two days. Then, according to the stock id, the historical factor data of each stock is obtained on the Joinquant platform[7].

| Data category       | Paraphrase                                      |
|---------------------|-------------------------------------------------|
| Stock data          | Comprehensive market, financial and fundamental data |
| Industry concept data | Industry sector and conceptual sector data      |
| Macroeconomic data  | Important domestic macroeconomic data            |
| Option data         | Contract information and market data of stock options and commodity options |
| Market fund data    | ETF, LOF, classification fund, monetary fund complete market, net value data |
| OTC fund data       | OTC fund unit net value, recovery net value, investment portfolio and other data |
| Alpha101 factor     | 101 alphas factors in papers published         |
| Alpha191 factor     | Short cycle trading alpha factor                |

While data preprocessing, in addition to some conventional feature processing method, the experience of the box-cox transformation[8] from Kaggle are introduced. These parts will be described in detail below.

3. Background

3.1. Factor processing
Generally speaking, there are some problems in the raw features. To solve the problem of dimension difference, min-max transformation is usually adopted to make the range of data in the same interval [9]. For sequence $x_1, x_2, \ldots, x_n$, min-max transformation is:

$$y_i = \frac{x_i - \min_{1 \leq j \leq n} \{x_j\}}{\max_{1 \leq j \leq n} \{x_j\} - \min_{1 \leq j \leq n} \{x_j\}}$$

The range of the new sequence $y_1, y_2, \ldots, y_n$ is mapped between $[0,1]$. For some non-linear measurements, Z-score conversion can also be used to remove dimensions

$$y_i = \frac{x_i - \bar{x}}{s}$$

among them, $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$, $s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}$[8].

One-hot transformation is used to solve the second problem of representation of the attribution factor. It represents a qualitative variable as a binary vector. To achieve it, mapping the category value to an integer value is needed, then each integer value is represented as a binary vector, [10]. To solve the problem of peak distribution, Box-Cox transformation is considered[9]. This method is aimed to reduce the correlation between unobservable errors and predictive variables to a certain extent and increasing the correlation between factors and predictive targets. The equations is as follows[9].

$$y = \begin{cases} 
\frac{(x+c)^\lambda - 1}{\lambda}, & \lambda \neq 0, \\
\ln(x+c), & \lambda = 0
\end{cases}$$
3.2. Model

Although models based on deep learning are the first choice now, noticed that one of the assumptions are the input features must satisfy the restricted condition of independent distribution, which means that most of the deep learning models do not consider the relationship among different factors, and it is worth studying. Thus, the Wide&Deep model is introduced in our work[10].

![Wide&Deep schematic diagram of model construction](image1)

The Wide&Deep model contains a linear model (Wide part) and a neural network model (Deep part) and it has both memory and generalization abilities[11]. In the Wide part, the transformation is repeatedly mentioned to generate composite features:

\[
\phi_k(x) = \prod_{i=1}^{d} \chi_i^{c_{ki}}, \quad c_{ki} \in \{0, 1\}
\]

where \(k\) is the \(k\)th feature, \(i\) represent the \(i\)th dimension, \(c_{ki}\) indicates whether the \(i\)th dimension feature should participate, \(d\) represents the total dimensions[10]. The wide part is the logistic regression, it can efficiently learn some representations across feature combinations, but it difficult to learn the representations that do not appear in the training set. Thus, the Deep model is applied. It is a feed-forward network with ReLU[11].

\[
a^{(l+1)} = f(W^{(l)}a^{(l)} + b^{(l)})
\]

where, \(W\) and \(b\) are the hyperparameters[12]. The whole model is training synchronously and the final output is calculated by the logistic loss function. The structure is shown in Fig 3.

\[
P(Y = 1|x) = \sigma(W_{\text{wide}}^{(l)}[x, \phi(x)] + w_{\text{deep}}^{(l)}a^{(l)} + b)
\]

![Structure diagram of the model in this paper](image2)
4. Experiments
This section has mainly represented the results of our experiment. In feature preprocessing, we focus on
the change of factor distribution brought by Box-Cox transformation. The results are shown in the
figures.

As the figures show, the change of data distribution about three groups of factors named a, b and c
respectively. In each group, the first figure is the original distribution, the second is the case after Z-
score transform, and the third is the case after Box-Cox transform. It can be seen that for the factors with
skewed-peak and sharp-peak distribution and not ideal either after Standard transform. But it becomes
better after Box-Cox transform. It is proved that our approach is effective in preprocessing factor data.

After the above work, the data set is randomly divided as training sets and test sets. While modeling,
TensorFlow provides the API of the wide&deep model in the estimator.DNNLinearCombinedClassifier
module[13]. The optimizer of the Wide part is FtrlOptimizer and the learning rate is 0.01; AdamOptimizer[14]
is the optimizer in the Deep part and the learning rate is 0.001. the Loss curve in the training process and the change of Accuracy in the verification process are shown in the figure.
It can be seen that in the training process, the Loss value converges at about 40 steps. When verifying, a relatively optimal model can be obtained at about the seventh Epochs.

In test process, ROC curve and AUC value are commonly used to objectively evaluate the classification effect in classification models[15-17]. The results are given below.

The performance of the model in the test set proves that our method is effective, and the model can achieve relatively ideal classification results.

To compare the prediction ability of our model with other models, three classic models named Naïve Bayes, Decision Tree, and Support Vector Machine(SVM) are chosen. After sufficient training, we directly compared the performance of the four models in the test set.

| Method       | Accuracy | Precision | Recall | AUC  |
|--------------|----------|-----------|--------|------|
| Naïve Bayes  | 0.7574   | 0.7629    | 0.8718 | 0.8312|
| Decision Tree| 0.7712   | 0.8252    | 0.8225 | 0.8280|
| SVM          | 0.7693   | 0.9674    | 0.7636 | 0.8480|
| Wide&Deep    | 0.8352   | 0.8552    | 0.8661 | 0.8733|

It can be seen that our model achieves the optimal results in all the metrics. Therefore, compared with the traditional three models, our model is the best in terms of comprehensive performance.

5. Conclusions
This paper aims to predict the tendency of the high-quality stocks in the A-shares market with deep learning model. At first, a sound investment strategy suitable for small and medium-sized investment is provided based on the stock selection and timing. Then while feature preprocessing, a method based on compound preprocessing is proposed to solve the problem of unnormal distribution. Finally, the
independent distribution of factors is considered in the modeling process, and how to add the cross result of the linear combination of factors into the depth model is studied. To prove the effectiveness of the idea and method, our model is compared with several classical models, and it is proved that our method is effective.

References
[1] Markowitz, Harry. PORTFOLIO SELECTION* [J]. The Journal of Finance, 1952, 7(1):77-91.
[2] Sharpe W F. CAPITAL ASSET PRICES: A THEORY OF MARKET EQUILIBRIUM UNDER CONDITIONS OF RISK* [J]. Journal of Finance, 1964, 19(3):425-442.
[3] Black F, Scholes M S. The Pricing of Options and Corporate Liabilities[J]. Journal of Political Economy, 1973, 81(3):637-654.
[4] Ross S A. The arbitrage theory of capital asset pricing[J]. Journal of Economic Theory, 1976.
[7] https://www.joinquant.com/
[8] https://www.kaggle.com/jack89roberts/top-7-using-elasticnet-with-interactions/notebook
[9] Alice Zheng, Feature Engineering for Machine Learning Models[M], O’Reilly, 2017.
[10] Box G, Cox D R. An Analysis Of Transformations[J]. Journal of the Royal Statistical Society. Series B: Methodological, 1964, 78(2).
[11] Cheng, H. T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhya, H., ... & Anil, R. (2016, September). Wide & deep learning for recommender systems. In Proceedings of the 1st workshop on deep learning for recommender systems (pp. 7-10).
[12] Glorot X, Bordes A, Bengio Y. Deep Sparse Rectifier Neural Networks[C]// Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS). 2011:315-323.
[13] https://tensorflow.google.cn/api_docs/python/tf/estimator/DNNLinearCombinedClassifier.
[14] Kingma D, Ba J. Adam: A Method for Stochastic Optimization[J]. Computer Science, 2014.
[15] Fawcett T. An introduction to ROC analysis[J]. Pattern Recognition Letters, 2005, 27(8):861-874.
[16] K, Z, X, et al. Build the Structure of WFSless AO System Through Deep Reinforcement Learning[J]. IEEE Photonics Technology Letters, 2018, 30(23):2033-2036.
[17] H Xu B, Xu Z, et al. Self-learning control for wavefront sensorless adaptive optics system through deep reinforcement learning[J]. Optik - International Journal for Light and Electron Optics, 2018, 178.