Stock Index and Financial Derivative Prediction Using Neural Networks

Author
Peng Qianyang
5130309751
364443007@qq.com

Abstract

This is a report to give a summary to our group's work this semester. To use neural network in financial data mining, we looked into many related works and tried several different models. In this report I will introduce our exploration process, then the data analysis and the formation of our models. Then I will give analysis to these models and draw the conclusion.

1 Exploration Process

Now I will introduce how our group work is organized. We have 5 group members in all, which is more than most of the other groups. So to match the number of members of our group, we separately tried several different methods in two different datasets to achieve the best result.

At first none of us is having any idea of how to start this work. We tried some simple and traditional methods such as SVM using the database of Wind Financial Terminal. These models are not satisfying and getting an accuracy near 50% in predicting rise and fall, which is close to using a random prediction.

On the second stage, we started to try some neural network approaches. According to the TA, financial portfolio is a good breakthrough point to start our work, so we searched about papers on this topic and did some work about it. Based on a long paper called "On the portfolio based on neural network" by a student in *Tianjing University*, we built our second model. It uses the same database to predict the stock index using neural network. To our surprise, the precision get to 98% even to 99%, which is somehow ridiculous. Ruled out the possibility of program mistakes, we found that the result is in fact so much like a graph translation over time t. That is, such prediction is meaningless. We used the same model but to predict rise and fall rather than stock index, the precision immediately fall to about 50%.

After this fail, we gave up the thought of predicting price using price and decided to find the inner relationship between the data, that is, the inner behaviors of dealers. Listened to a fellow student who is doing financial stuff now introducing concepts as handicap, put order, kill order and turnover, we got to know how can we use the data other than stock price. Based on this we built some new models to deal with the 3-divide problem (rise, flat, fall) and get a over 50% precision, which is much better than the 2-divide models we have built earlier.

Above is our main work this semester. As the leader of our group, my work is mainly on data analysing and model building, and the coding work and some details is mainly implemented by other group members. So in the following parts, data analysing and model building will be more discussed, and you can refer to other members' papers for more details.

2 Group Members

- Peng Qianyang
- Cui Jingyu
- Jia Yuting
- Shi Zhenfeng
- Zhou Chang

3 Data Analysis

From this class we get some raw data of stock exchange, but they are not intuitive and not easy to understand by outsiders. Although you can put all the columns into the neural network but apparently you can never expect a good result. So having a full understanding to the data is important, only

by then we can do data preprocessing to make out model more reasonable.

There are totally 32 columns in the raw data, from column 1 to 29 is what we actually care about. Their meanings are as below:

- line1 InstrumentID:
  Code for the corresponding future. For example *rb1701* stands for Deformed Steel Bar 01.
- line2 TradingDay:
  The date of the indicated time.
- line3 UpdateTime:
  The indicated time, round to second.
- line4 UpdateMillisec:
  The indicated time, round to 500 milliseconds.
- line5 LastPrice:
  The knock down price of last turnover.
- line6 Volume:
  The total number of Standard hands today. A Standard hand consists of 200 units.
- line7 LastVolume:
  The number of Standard hands of the last turnover.
- line8 Turnover:
  The total amount of transactions.
- line9 LastTurnover:
  The amount of transactions in the last trade.
  $t = LastPrice * LastVolume$
  $LastTurnover = 200t$
- line10-14 AskPrice:
  The five lowest price sellers offer.
- line15-19 BidPrice:
  The five highest price from bidder. The lowest AskPrice and the highest BidPrice are called the Handicap.
- line20-24 AskVolume:
  The volume of corresponding AskPrice.
- line25-29 BidVolume:
  The volume of corresponding BidPrice.

After getting these knowledge, we can start to analysis the behaviours of dealers. We can find the behaviour patterns below:

1. If the LastPrice is equal to the last lowest AskPrice, that is a bidder put an order at that price. This somehow means that now people are willing to buy rather than sell, we can somehow expect the price will rise.
2. If the LastPrice is equal to the last lowest BidPrice, that is a seller put an order at that price. This somehow means that now people are willing to sell rather than buy, we can somehow expect the price will fall.
3. If the AskPrice1, the lowest AskPrice is getting lower and lower but without any trade, that means people are wanting to sell their futures at a lower and lower price, but nobody want to buy such a future. We can somehow expect the price will fall.
4. If the BidPrice1, the highest BidPrice is getting higher and higher but without any trade, that means people are wanting to buy such a futures at a higher and higher price, but nobody want to sell their future. We can somehow expect the price will rise.
5. If the AskPrice1, the lowest AskPrice is getting higher and higher, that means sellers are cancelling their orders because the current price cannot match with their expectation. We can somehow expect the price will rise.
6. If the BidPrice1, the highest BidPrice is getting lower and lower, that means bidders are cancelling their orders because the current price cannot match with their expectation. We can somehow expect the price will fall.

From these patterns we can get to know how can we choose the proper features to train in neural network. Generally, the handicap, which means the highest Bidprice and lowest AskPrice, is the most sensitive part of data and has the most close relationship to the final price. Also, user behaviours can be revealed most clearly in the handicap.

Acknowledgments

References

eng Qianyang 2016. *Zou jin ke xue*, volume 1. CCTV