Trading Adviser

Team member: Yi-Bo,Wang、Chen-Xin,Liao、Jia-Hui,Li、Bo,Liang、Qiu,Li (Group 8)
Adviser: Mu-En,Wu、Chen-Yu,Han

Date: 2020/01/02

Introduction

On the basis of Dr. Han's speech and code. We found that Dr. Han did not do the shuffle and split train/test in the beginning. He split the data after labeling the data and creating features. We are curious about that if we do the shuffle and split the data in the beginning, will the small change affect the result?
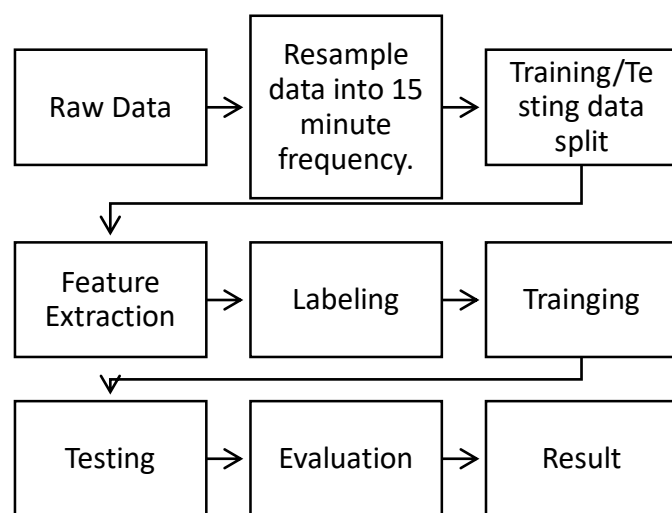
Background

Data Source: Fin Lab
Investment Target: TWSE Index
Duration: 2006/01/01 – 2019/10/09

Mythology

Different from Dr. Han's method, we split the data into two parts, which are training data and testing data(Training Data: 2005/01/01-2015/12/31
Testing Data:2016/01/01-2018/12/31).

Data Preprocessing

Raw Data: Tick Data, including Opening Price, Highest Price, Lowest Price, Closing Price, Volume.
(1)Select Opening Price as our research data and drop Highest Price, Lowest Price, Closing Price and Volume column.
(2)Resample data into 15 minutes frequency.

Feature Extraction

Base on Dr. Han's speech, we use technical index to develop our feature and we have 42 features in total.

Labeling Data

(1)Continuous trading signals
We use the Interpolation Method to computer the label.

$$
y \begin{cases} \dfrac{p_{t+w} - p_{t,t+w}^{min}}{p_{t,t+w}^{max} - p_{t,t+w}^{min}} & if\ p_{t+w} > p_t \\[2em] 0.5 \left(1 - \dfrac{p_{t+w} - p_{t,t+w}^{min}}{p_{t,t+w}^{max} - p_{t,t+w}^{min}} \right) & else \end{cases}
$$

(2)Trading Strategy
When y is better than PR 60 of past ten days summation, we will buy at the opening price and sell it at the next opening price.
(3)Experiment
We tried three different length (30,60,90) of w to label the data.

W=30   W=60   W=90
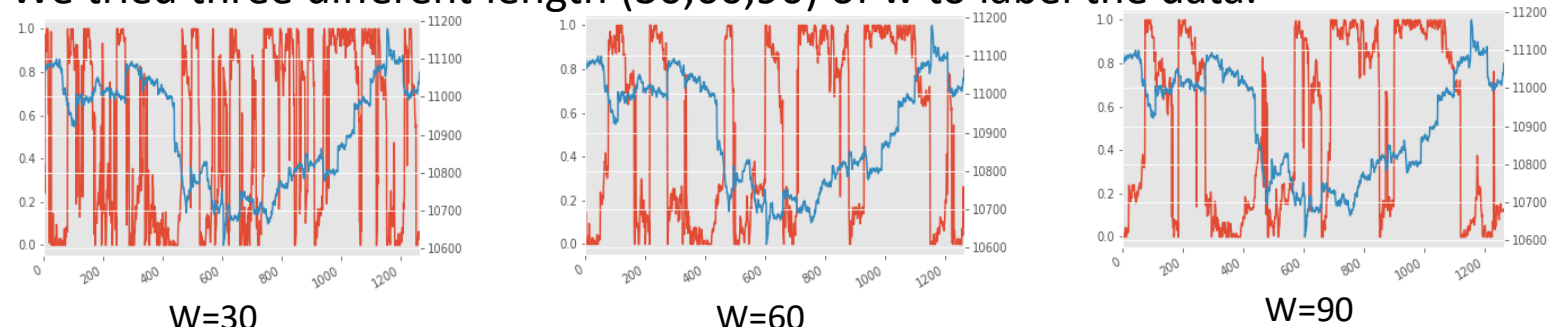
From the three pictures, we found that the less length of w can catch the price wave better.

Training and Testing Model

(1) label 1: w= 30

Training	Testing
Winrate:50.15%	Winrate:50.15%
Odds:-1.31	Odds:-1.14
Profit factor:-1.31	Profit factor:-1.08
Max drawdown: -828.33	Max drawdown: -828.33

(2) label 2: w= 60

Training	Testing
Winrate:49.78%	Winrate:48.75%
Odds:-1.30	Odds:-1.16
Profit factor:-1.28	Profit factor:-1.10
Max drawdown: -813.63	Max drawdown: -997.10

(3) label 3: w= 90

Training	Testing
Winrate:49.31%	Winrate:48.50%
Odds:-1.26	Odds:-1.15
Profit factor:-1.23	Profit factor:-1.08
Max drawdown: -766.79	Max drawdown: -1414.82

Conclusion

We found that in different length of time to give data label will affect the result. The win rates are almost the same. But it affect drawdown a lot. We can see that in label 3, it have the highest drawdown, label 1 and label 2 have no difference.
Compare to Dr. Han's result, in different methods to deal with the data will also affect the result. We can see it that in our method we have much drawdown.