Proposal：Timing Model Based on Orderbook

Jiajia Wang; Yuchuan Xu; Yutian Zhou; Masha

Our overall idea is that constructing factors from the three – month orderbook of a stock firstly. The interval of factors is three seconds. Second step is to screen factors. And third, we use SVM model to predict the change of stock. Also, we use other stocks as the verification set to verify the robustness of the model.

Database

Because of the large amount of data, we choose to use database to help process and calculate factors. Because our data are structured data, we choose MySQL as the database.

Features Construction

1. Motivation

The price and quantity information on the order book can reflect the attitude and trading ability of the trader to the stock, so as to judge the stock trend. So we build factors from the information in the order book. In addition, considering the effectiveness and collinearity of the factors, we also need to preliminarily screen the factors

2. Factors

Our factors include *bid ask price spread, bid ask volume spread , (bid price1+ ask price1)/2, (bid volume1+ ask volume1)/2, bid price5 - bid price1, ask price5 - ask price1, ave(ask price) , ave(bid price), ave(ask volume) , ave(bid volume), log*($bid\ price1_t/bid\ price1_{t-1}$), *log*($ask\ price1_t/ask\ price1_{t-1}$), *log*($bid\ volume1_t/bid\ volume1_{t-1}$), *log*($ask\ volume1_t/ask\ volume1_{t-1}$), *log*($(volume_t - volume_{t-1})/(volume_{t-1} - volume_{t-2})$) , $\sum_i(ask\ volumei(last\ price/(ask\ pricei - last\ price))/\sum_i last\ price/(ask\ pricei - last\ price))$, $\sum_i(bid\ volumei(last\ price/(last\ price - bid\ pricei))/\sum_i last\ price/(last\ price - bid\ pricei))$. *If those functions are not effective, we may include factors:* $\lambda_{ask\ deal}$, $\lambda_{bid\ deal}$, $\lambda_{ask\ limit}$, $\lambda_{bid\ limit}$, *large ask deal order volumes(maybe >2000)/ total ask deal order volumes, large bid deal order volumes(maybe >2000)/ total bid deal order volumes, large ask limit order volumes(maybe >2000)/ total ask limit order volumes, large bid limit order volumes(maybe >2000)/ total bid limit deal order volumes.*

We will select the more effective factors according to the linear single factor time series regression, and remove the factors with high correlation.

Feature selection based on LASSO

1. Motivation

After factor construction, we use the LASSO to do the feature selection. As we construct the features, it is possible that many features are highly related as some of them are the linear transformation of the original tick data.

2. Model

Input:

All the features

Parameter tuning:

Use cross validation to choose the optimal \lambda\

Output:

Features selected by LASSO

High frequency trading strategy based on SVM

1. Motivation

Essentially, the strategy we want to do is actually a classification problem. SVM algorithm has been widely used in classification so we decided to build our model based on SVM. SVM with hard margin and SVM with soft margin can handle linear divisibility and linear indivisibility separately.

2. Model

Input:

X: selected features and Y: the three classes (Price goes up, down and unchanged)

Kernel selection:

Using the Cross-Validation method assuming we use Hinge Loss Function, that is, when selecting kernel, different kernels (Polynomial Kernel, Gaussian Kernel, Linear Kernel, Laplace Kernel, Sigmoid Kernel, RBF Kernel) are tried separately, the one with the least inductive error is the best kernel. For the parameters of different kernels, cross-validation is also used to do the parameter tuning.

Loss Function selection:

Assume that investors are risk averse, define the loss function as following:

$$L_i = \sum_{j \neq j_i} \max(0, s_j - s_{y_i} + \Delta)$$

$$L = \frac{1}{N} \sum_{i=1}^{N} L_i$$

As the investors are risk averse, so the weights are not equal. The penalty for misclassifying the price going down as price going up will be much sever.

Parameter tuning:

Except for the parameters of different kernels,

The weights of the loss function can be the key parameter of tuning.

The number of classes can be another key parameter of tuning.

Output:

Predicted class for the next time

3. Robustness analysis

Apply the model to another two stocks, to see whether it also have good performances. If it is, we will conclude that our model is robust.