PIXIU

Xiao Zhang

Thinking

Now we used financila data and compare the results of different models, and we are thinking:

1. Besides financial forecasting, can we create more financial tasks based on existing data sets? (eg. BigData22) Can **more financial tasks** be completed with current data and model ? Such as **portfolio management.**
https://github.com/stocktweet/stock-tweet

About **portfolio management,** I'm thinking:

1. I think we are doing a predicting work, which means we have in-the sample and out-of sample, and in this way, I think it may not be just an algebra question. And I'm not very sure what this means "I expect the results to degrade quickly as the number of assets increases."

Thinking

portfolio management(contimue):

2. I think we can rank the return or any other index you think better , as long as we make the ranking metric the same, and then choose the TOP ten percent assets to build a portfolio. It is a very simple method but it works.

3. In this process, we may not be able to handle risk and return at the same time. In this case, I think choosing a return index may be better. And we can combine this ranking metric with text information together to find a more proper portfolio. However, here is a question: A portfolio with fifty or more assets can take up a lot of memory.

Thinking

If we are supposed to create more financial tasks, then we should think:

1. Is the existing data set sufficient?

2. Do we need to expand the dataset and **what kind of dataset do we need ?**

3. Now we have dataset BigData22,which contains both price and twitter information, however, twitter text is noisy and sparse, maybe we can **add news text** ? And we can analyse the difference between using numerical data alone and using numerical data and text information together?

Thinking

4. At present, many researches and analyses of assets are done separately, such as spot market, futures market, alternative market, commodity market, securities and stocks.
Can we create a training model that can **analyze products in many different markets**?

5. At present, many models are only trained for one language (one market).
Perhaps we can expand to **domestic and foreign markets** (for example, China and the United States) for training. (Is this feasible? What are the main difficulties in the analysis of different countries?)

6. If our model training works well, **what advantages do we have over traditional analysis ?**
Can we explain that it can capture some special information? (Maybe it can capture information that traditional analysis methods, such as factor analysis, can't capture)