

Paper research : short-term market timing strategy based on boosting ML algos

1. Proposal ready to go (ddl: 3/3)
  - a. Trista: boosting method detail
  - b. Alfred: Single method research
  - c. Evan: preprocess
  - d. Lorelei: Data collection part
2. Data collection: in the paper they use 51 factors for input data, and use 50etf (and wind A stock) return for predicting target:

图表 2: 51 种数据因子定义

1	修正后美国多利指数和标普500: 1 天	27	修正后美国多利指数
2	修正后美国标普500: 1 天: 修正后标普500 的 1 天	28	修正后美国标普500: 1 天: 修正后标普500 的 1 天
3	修正后美国标普500: 1 天: 修正后标普500 的 1 天	29	修正后美国标普500: 1 天: 修正后标普500 的 1 天
4	修正后美国标普500: 1 天: 修正后标普500 的 1 天	30	修正后美国标普500: 1 天: 修正后标普500 的 1 天
5	修正后美国标普500: 1 天: 修正后标普500 的 1 天	31	修正后美国标普500: 1 天: 修正后标普500 的 1 天
6	修正后美国标普500: 1 天: 修正后标普500 的 1 天	32	修正后美国标普500: 1 天: 修正后标普500 的 1 天
7	修正后美国标普500: 1 天: 修正后标普500 的 1 天	33	修正后美国标普500: 1 天: 修正后标普500 的 1 天
8	修正后美国标普500: 1 天: 修正后标普500 的 1 天	34	修正后美国标普500: 1 天: 修正后标普500 的 1 天
9	修正后美国标普500: 1 天: 修正后标普500 的 1 天	35	修正后美国标普500: 1 天: 修正后标普500 的 1 天
10	修正后美国标普500: 1 天: 修正后标普500 的 1 天	36	修正后美国标普500: 1 天: 修正后标普500 的 1 天
11	修正后美国标普500: 1 天: 修正后标普500 的 1 天	37	修正后美国标普500: 1 天: 修正后标普500 的 1 天
12	修正后美国标普500: 1 天: 修正后标普500 的 1 天	38	修正后美国标普500: 1 天: 修正后标普500 的 1 天
13	修正后美国标普500: 1 天: 修正后标普500 的 1 天	39	修正后美国标普500: 1 天: 修正后标普500 的 1 天
14	修正后美国标普500: 1 天: 修正后标普500 的 1 天	40	修正后美国标普500: 1 天: 修正后标普500 的 1 天
15	修正后美国标普500: 1 天: 修正后标普500 的 1 天	41	修正后美国标普500: 1 天: 修正后标普500 的 1 天
16	修正后美国标普500: 1 天: 修正后标普500 的 1 天	42	修正后美国标普500: 1 天: 修正后标普500 的 1 天
17	修正后美国标普500: 1 天: 修正后标普500 的 1 天	43	修正后美国标普500: 1 天: 修正后标普500 的 1 天
18	修正后美国标普500: 1 天: 修正后标普500 的 1 天	44	修正后美国标普500: 1 天: 修正后标普500 的 1 天
19	修正后美国标普500: 1 天: 修正后标普500 的 1 天	45	修正后美国标普500: 1 天: 修正后标普500 的 1 天
20	修正后美国标普500: 1 天: 修正后标普500 的 1 天	46	修正后美国标普500: 1 天: 修正后标普500 的 1 天

3.

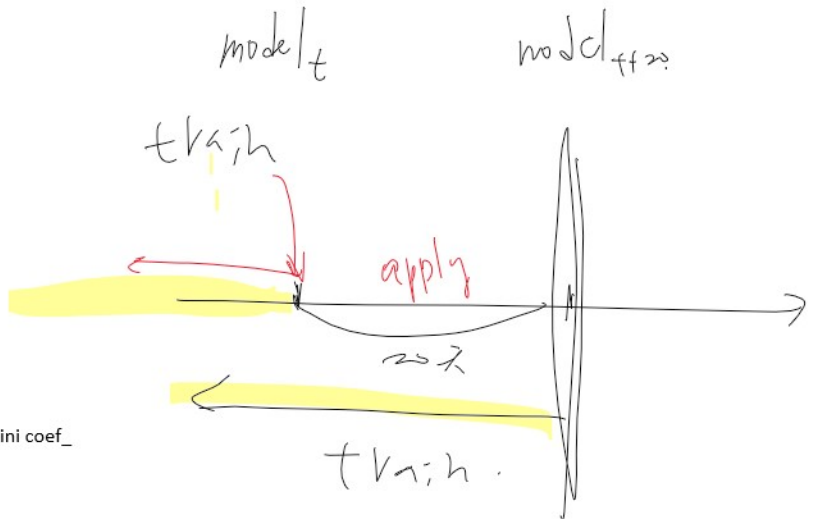
定量研究专题报告



11	修正后美国标普500: 1 天: 修正后标普500 的 1 天	47	修正后美国标普500: 1 天: 修正后标普500 的 1 天
12	修正后美国标普500: 1 天: 修正后标普500 的 1 天	48	修正后美国标普500: 1 天: 修正后标普500 的 1 天
13	修正后美国标普500: 1 天: 修正后标普500 的 1 天	49	修正后美国标普500: 1 天: 修正后标普500 的 1 天
14	修正后美国标普500: 1 天: 修正后标普500 的 1 天	50	修正后美国标普500: 1 天: 修正后标普500 的 1 天
15	修正后美国标普500: 1 天: 修正后标普500 的 1 天	51	修正后美国标普500: 1 天: 修正后标普500 的 1 天
16	修正后美国标普500: 1 天: 修正后标普500 的 1 天		

4. Data cleaning:

- a. Data formats, structures (good to use)
  - b. Data normalization, extreme value process ??
5. Eda
- a. 可视化, (畫圖)
  - b. 降維, PCA? LDA? Lasso? ??
  - c. See see (康康) cross entropy? Information gain, gini coef\_



(some software engineering work to do)

6. Phase1: Single model (decision treeeeee)
  - a. in the paper they used the decision treeeeee for prediction  
Classify 漲跌 或 跌跌, 根據預測結果買進賣出 T000000
  - b. 滾動每 20 天重新訓練一次模型, 用於 未來 20 天的預測
  - c. 5 fold cross validation for hyper para (depth ect.)
  - d. Maybe try other model? (logistic, ect.)
7. Phase2: Boooosting method
  - a. We mainly try to use adaboost method as our boosting algo to optimize our result  
 具体来说, Boosting 是集成学习方法 (Ensemble Method) 的一种, 在实践中有着广泛的应用, 而 Adaboost 模型则是其中最流行的一种实现. 最早由 Yoav Freund 和 Robert Schapire 在 1995 年提出. Adaboost 模型是一种弱分类器, 它通过组合多个弱分类器的输出, 来提高分类性能.  
 在 Adaboost 模型中, 弱分类器通常是基于决策树的. 它的训练过程是迭代的, 每次迭代都会训练一个新的弱分类器, 并将其与之前的弱分类器组合起来. 这个过程会一直持续到弱分类器的性能不再提高为止. 最终, 所有弱分类器的输出会被加权并组合起来, 形成最终的强分类器. 由于弱分类器的性能通常较差, 因此通过组合多个弱分类器, 可以显著提高模型的分类性能.
  - b. Try with diff base classifier to test
  - c. We also use expanding window to trian our model and predict trading signal
8. Xgboost????? Doctor fang will explain