

基于TRA和最优运输学习的多股票交易模式

Original QIML编辑部 量化投资与机器学习 2021-09-10 15:31

收录于合集

#深度研读系列

21个 >



量化投资与机器学习微信公众号，是业内垂直于**量化投资**、**对冲基金**、**Fintech**、**人工智能**、**大数据**等领域的主流自媒体。公众号拥有来自**公募**、**私募**、**券商**、**期货**、**银行**、**保险**、**高校**等行业**20W+**关注者，连续2年被腾讯云+社区评选为“年度最佳作者”。

量化投资与机器学习公众号独家解读

量化投资与机器学习公众号 *QIML Insight*——**深度研读系列**是公众号今年全力打造的一档**深度**、**前沿**、**高水准**栏目。

公众号**遴选**了各大期刊前沿论文，按照理解和提炼的方式为读者呈现每篇论文最精华的部分。QIML希望大家能够读到可以成长的量化文章，愿与你共同进步！

本期遴选论文

来源：KDD 2021

标题：Learning Multiple Stock Trading Patterns with Temporal Routing Adaptor and Optimal Transport

作者：Hengxu Lin、Weiqing Liu、Jiang Bian、Dong Zhou

KDD2021

第27届国际数据挖掘与知识发现大会 KDD 2021 于8月14日-18日在线上正式开幕。本次大会共收到1541篇投稿，其中238篇论文被接收，接收率为15.44%。**其中跟唯一一篇文章跟量化投资投资相关的论文是出自微软研究院的《Learning Multiple Stock Trading Patterns with Temporal Routing Adaptor and Optimal Transport》。**

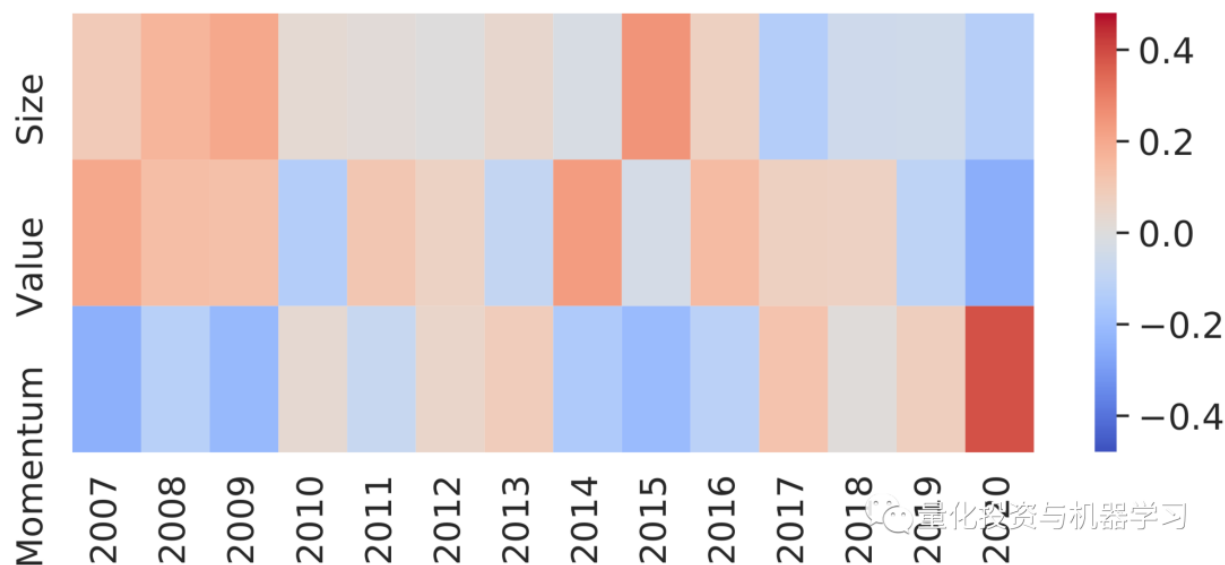
TRA量化模型

首先，整体来介绍一下模型的框架结构：

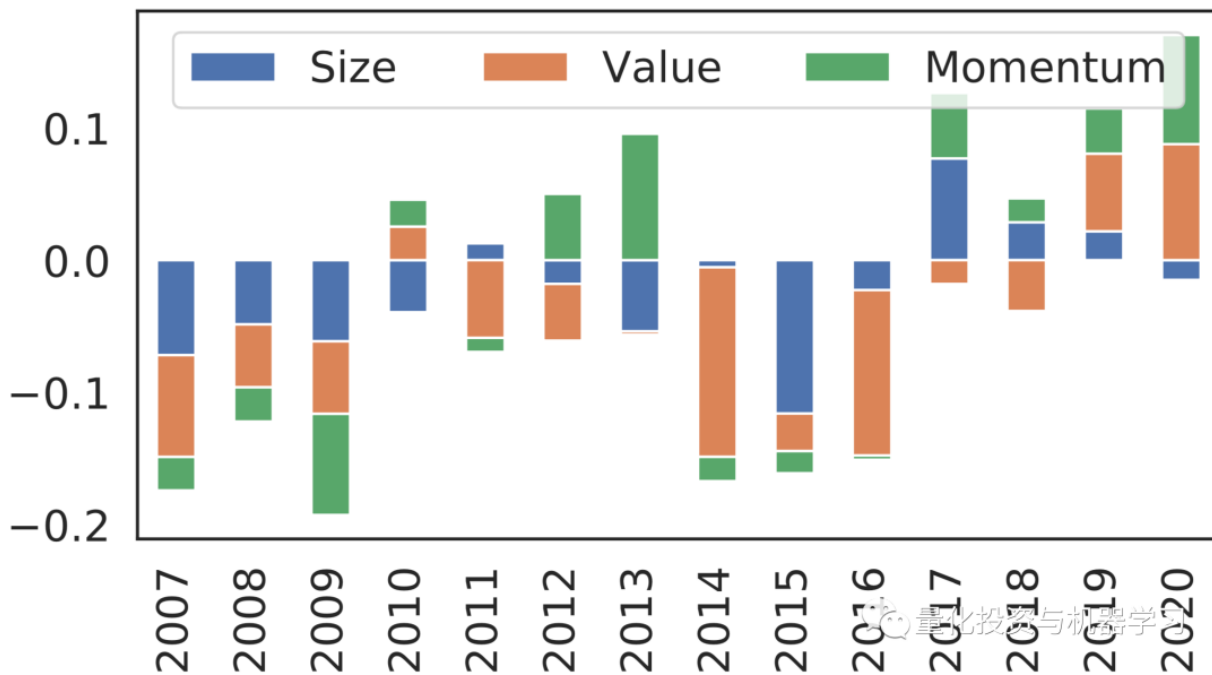
- 股票市场同时存在多种状态，每个时刻都有不同的状态占主导，所以需要根据不同的状态单独构建 Predictor。每只股票，在不同的时刻适合用不同的 Predictor 进行预测，才能达到更优的预测效果。
- 所以，可以对市场状态的数量进行假设，比如有5个状态，那就可以构建5个 Predictor，**但是问题就来了，是对于不同的股票，在每个时刻该选择哪个 Predictor？**
- **作者提出了路由的概念，根据每个 Predictor 历史上预测的表现及股票的特征，匹配出当前时间点最佳的 Predictor，这就是 TRA 核心的思想。**
- 但是问题又来了，如果对于路由不加限制，就会一直选择一个历史上表现最优的 Predictor，而违背了最初的设计，所以作者又引入了最优运输的机制，控制 Router 对 Predictor 的分配。

市场同时存在多种状态

股票预测是量化投资中最为关键的任务。近年来，深度神经网络因其强大的表征学习能力和非线性建模能力，逐渐成为股票预测的主流方法。现有的预测方法均假设股票数据符合独立同分布（IID）且采用单一模型有监督地对股票数据建模。但实际上，股票数据通常会包含多种不同甚至对立的分布（Non-IID），比如动量（历史收益率高的股票未来收益率会高）和反转（历史收益率低的股票未来收益率会高）这两种分布形式同时存在于股票数据中，但是已有的模型并不具备同时学习股票数据中多种分布的能力。



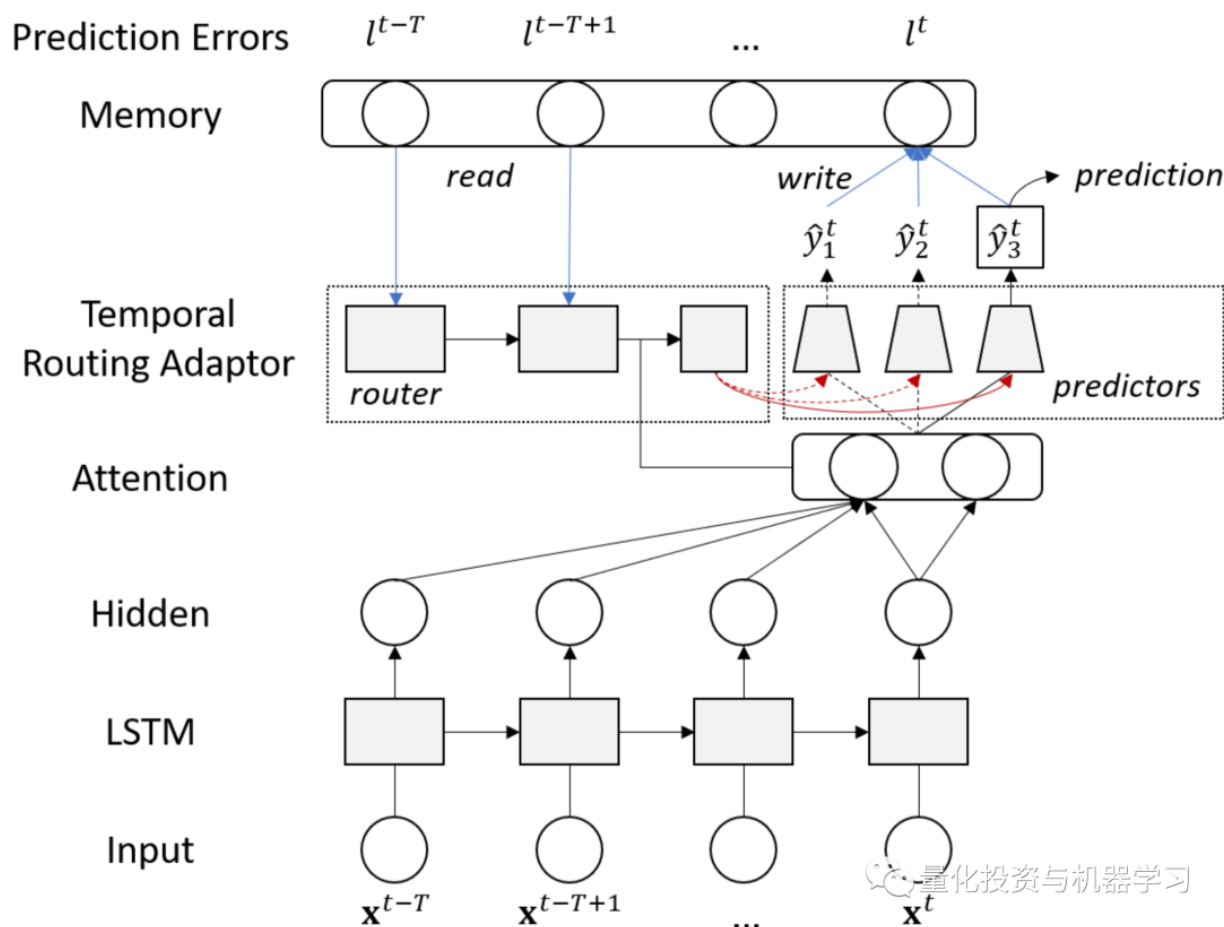
Momentum、Value、Size因子Top组在2007年到2020年间，每年的超额收益



Momentum、Value、Size三个因子截面回归的系数

TRA是什么？

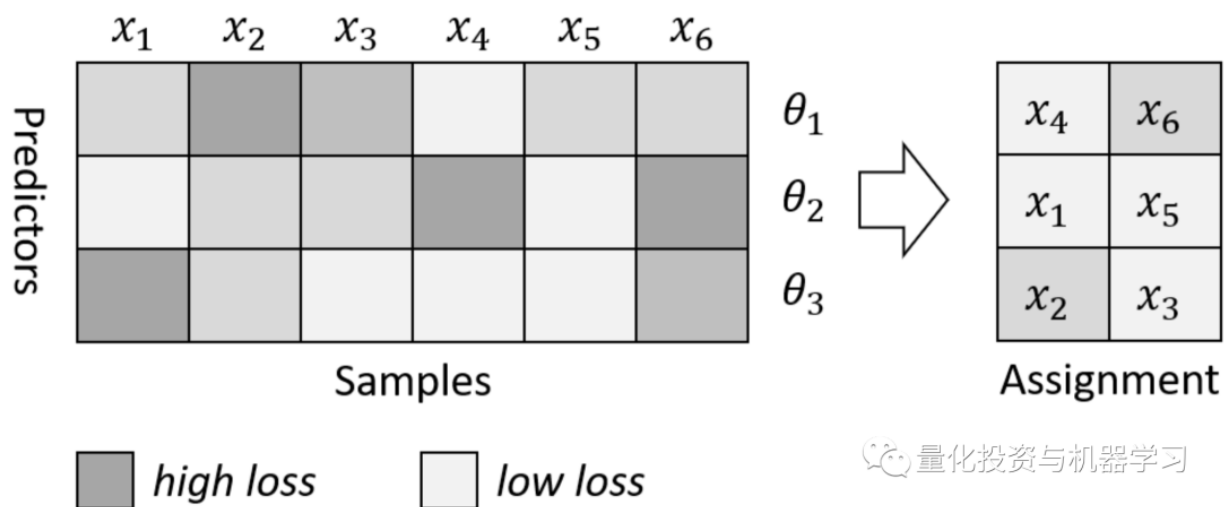
因此，微软亚洲研究院的研究员们提出了 Temporal Routing Adaptor (TRA)，来赋予已有模型学习多种分布的能力。具体而言，TRA 在给定骨干模型的基础上，引入了一组 Predictors 来建模不同分布，和一个 Router 来根据样本的规律 $p(y_t | X_t)$ 将其分配到所属的 Predictor 上进行训练和推理。为了保证 Router 能够预测出样本的规律，研究员们设计并利用了两种与 $p(y_t | X_t)$ 关联的信息作为其输入：1) 利用骨干模型的隐层来表征 $p(y_t | X_t)$ ，2) 利用 Predictor 的历史预测偏差来表征 $p(y_{(< t)} | X_{(< t)})$ 。实验表明，这两种信息对 Router 有能力预测出样本规律起到了重要作用。TRA 的两个主要模块和基于骨干模型的具体实现可以参下图：



针对当前的Sample，根据Attention LSTM的输出及记忆中的temporal prediction errors 共同确定最合适的Predictor

什么是最优运输？

为了有效地训练 TRA 模型，另一个需要解决的问题是，如何保证分配到不同 Predictor 的样本是属于不同规律的。因此，研究员们基于最优运输 (Optimal Transport) 设计了一个迭代优化的算法。最优运输被用来求解在分配的样本满足特定比例约束下，如何分配样本能够最小化整体预测偏差。求解得到的分配方案会用来更新对应的 Predictor，并继续下一轮迭代，直至收敛。



实验表明，TRA 可以稳定提升之前在股票预测中表现最强的基准模型如 Attention LSTM 和Transformer 的预测性能，并取得更高的投资收益，结果见下表。

Method	Ranking Metrics				Portfolio Metrics			
	MSE (↓)	MAE (↓)	IC (↑)	ICIR (↑)	AR (↑)	AV (↓)	SR (↑)	MDD (↓)
Linear	0.163	0.327	0.020	0.132	-3.2%	16.8%	-0.191	32.1%
LightGBM	0.160(0.000)	0.323(0.000)	0.041	0.292	7.8%	15.5%	0.503	25.7%
MLP	0.160 (0.002)	0.323 (0.003)	0.037	0.273	3.7%	15.3%	0.264	26.2%
SFM	0.159 (0.001)	0.321 (0.001)	0.047	0.381	7.1%	14.3%	0.497	22.9%
ALSTM	0.158 (0.001)	0.320 (0.001)	0.053	0.419	12.3%	13.7%	0.897	20.2%
Trans.	0.158 (0.001)	0.322 (0.001)	0.051	0.400	14.5%	14.2%	1.028	22.5%
ALSTM+TS	0.160 (0.002)	0.321 (0.002)	0.039	0.291	6.7%	14.6%	0.480	22.3%
Trans.+TS	0.160 (0.004)	0.324 (0.005)	0.037	0.278	10.4%	14.7%	0.722	23.7%
ALSTM+TRA (Ours)	0.157 (0.000)	0.318 (0.000)	0.059	0.460	12.4%	14.0%	0.885	22.4%
Trans.+TRA (Ours)	0.157 (0.000)	0.320 (0.000)	0.056	0.442	16.1%	14.2%	1.133	23.1%

RA 模型相比于其他基准模型在股票排序预测任务下的性能

实证设计的细节

X: 16个特征

作者选取了中证800股票，以下16个特征market capitalization, price-to-EPS, price-to-book value, price-to-sales, price-to-net cash flow, asset turnover ratio, net profit margin, receivables turnover ratio, EPS growth, asset growth, equity growth, 12 month momentum, 1 month reversal, close to 12 month highest, close to 12 month lowest, maximum return in last month. **每个月末的数据。**

Y: 预测目标变量

未来一个月每个股票月度收益率的截面排序分位数（标准化之后）。

所以每次前向传播的Sample data的数据结构如下，过去60个月，16个因子：

因子

时间	Feature 1					Feature 16
	t-60					
	t-59					
	t					

量化投资与机器学习

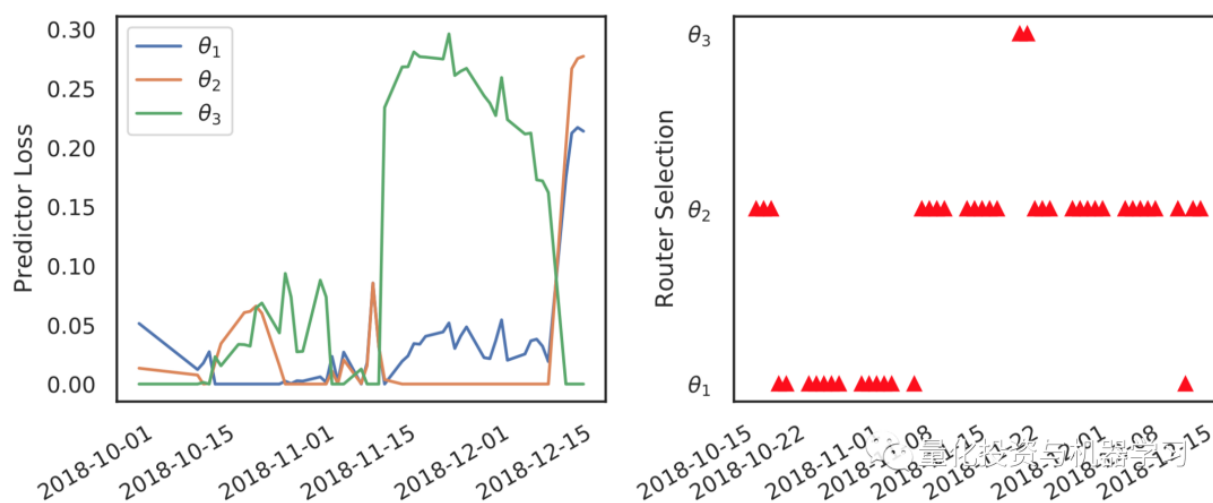
基准模型：

模型	说明
Linear	线性模型
LightGBM	
MLP	多层感知机
SFM	一种将LSTM的隐含表征分解为多频度的改装过的RNN
ALSTM	加入注意力机制的LSTM
Trans	Transformer
ALSTM+TS/Trans.+TS	根据不同的时间段选择对应的Predictor

量化投资与机器学习

核心问题解答

问题1，TRA 能否正确的学到市场存在的多个状态并选择正确的那个用于预测？



在实验中，我们假设市场存在三种状态（对应的有三个Predictor），并随机的选取股票进行预测，图6左侧是不同Predictor的Loss，右侧是对应的那个时间点TRA选择的Predictor，可以看出大部分时间TRA选择的是2号预测器，也就是Loss最小的那个。在2018年下半年，基本上选择的是1号，整体很少会选择3号，因为三号大部分时间Loss都很大。说明TRA能够选择Loss最小的Predictor。

问题2，LSTM输出的隐含表征（hidden state）和Predictor的预测误差对于Router是否都有利用价值？

对于Router的输入，作者做了以下控制：

- Random，随机生成的白噪声作为Router的输入
- LR：只把ALSTM输出的隐含表征（hidden state）作为Router的输入
- TPE：只把Predictor的历史预测误差作为Router的输入
- LR+TPE：LR和TPE都作为Router的输入

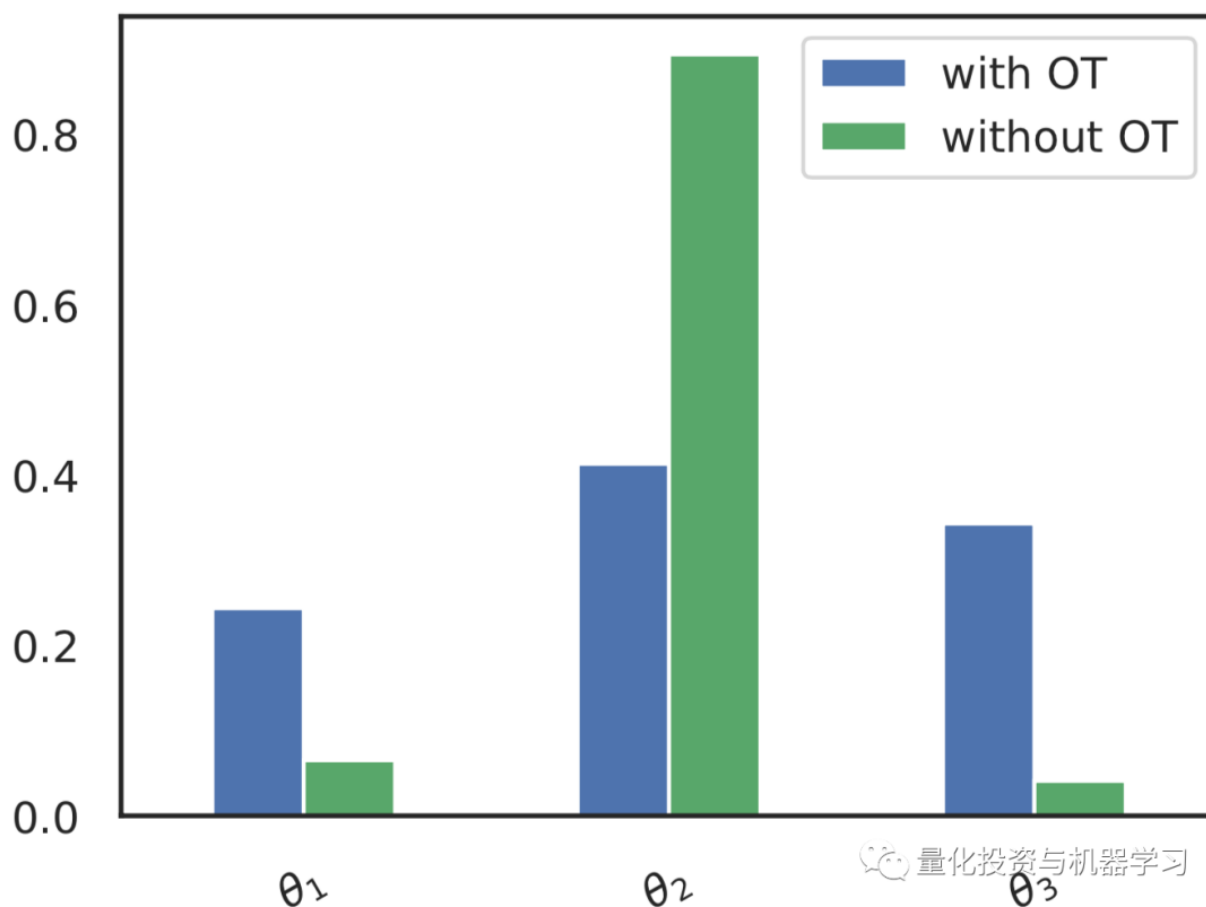
Table 2: TRA's performance when used with different information sources.

Information	MSE (↓)	MAE (↓)	IC (↑)	ICIR (↑)
Random	0.159 (0.001)	0.321 (0.002)	0.048	0.362
LR	0.158 (0.001)	0.320 (0.001)	0.053	0.409
TPE	0.158 (0.001)	0.321 (0.001)	0.049	0.381
LR+TPE	0.157 (0.000)	0.318 (0.000)	0.059	0.450

如上表2可以发现，相对单独输入LR或TPE，LR+TPE的各项指标都得到领先。

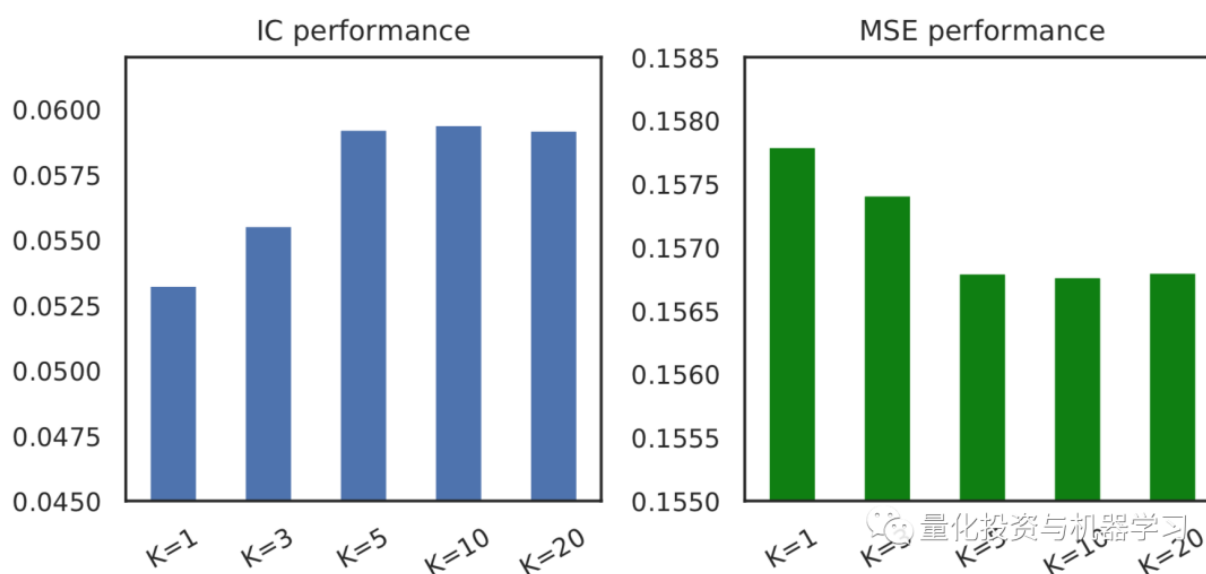
问题3，最优运输规则（Optimal Transport, OT）到底对于TRA有没有帮助？

如图7，作者对比了OT对于模型的影响，在没有OT的时候大部分时间选择的都是Loss相对最少的2号Predictor，而加入OT之后，各个Predictor被选择的比率就比较均匀了。



问题4, TRA模型对于Predictor的数量（也就是假设的市场状态的数量）是否敏感？

从图8可以看出，当K大于5之后，模型的表现趋于稳定，IC维持在0.06左右，MSE也不再降低，说明当Predictor的数量达到一定量，模型对于K就不敏感了。



QLib代码实现

模型实现的代码已经封装在QLib里，关于QLib的介绍可以参考公众号之前的文章：



微软也搞AI量化平台？还是开源的！

如果想要动手复现文中的结果，需要以下几个步骤：

1、Clone QLib到本地，并进行安装；

2、下载特征及标签数据，地址：

<https://drive.google.com/drive/folders/1fMqZYSeLyrHiWmVzygel4sw3vp5Gt8cY?usp=sharing>

3、将数据放入以下文件夹：/qlib/examples/benchmarks/TRA/data

4、修改/qlib/examples/benchmarks/TRA/example.py内的代码，图中红圈部分改为config_alstm_tra_init，然后运行文件。

```
if __name__ == "__main__":  
  
    # set params from cmd  
    parser = argparse.ArgumentParser(allow_abbrev=False)  
    parser.add_argument("--seed", type=int, default=1000, help="random seed")  
    parser.add_argument("--config_file", type=str, default="configs/config_alstm.yamlt", help="config file")  
    args = parser.parse_args()  
    main(**vars(args))
```

量化投资与机器学习

关于模型的细节，可以参考/qlib/examples/benchmarks/TRA/src/model.py

收录于合集 #深度研读系列 21

< 上一篇

波动率目标策略，没有想象的那么简单！

下一篇 >

重构订单簿！基于深度学习的A股Tick级价格变动预测

People who liked this content also liked

北大满哥与奥迪的罗生门

量化投资与机器学习

