

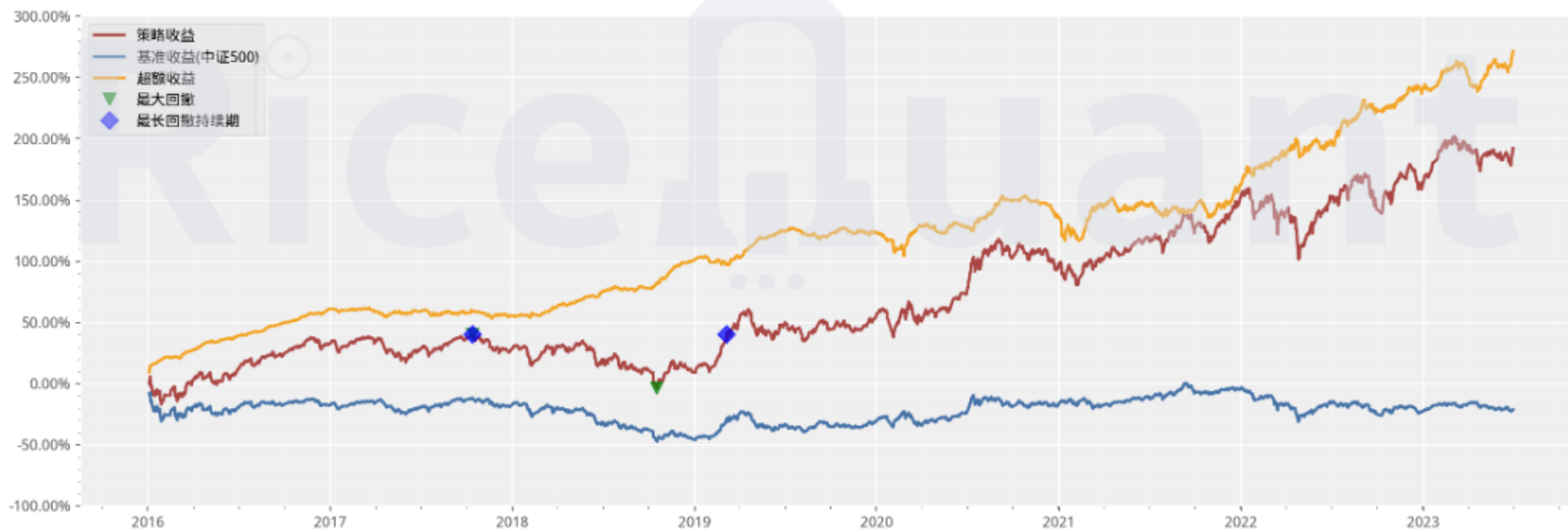
端到端深度学习因子挖掘

卖方研报梳理

一、为什么要做深度学习端到端因子挖掘

300个人工挖掘量价类因子进行因子组合

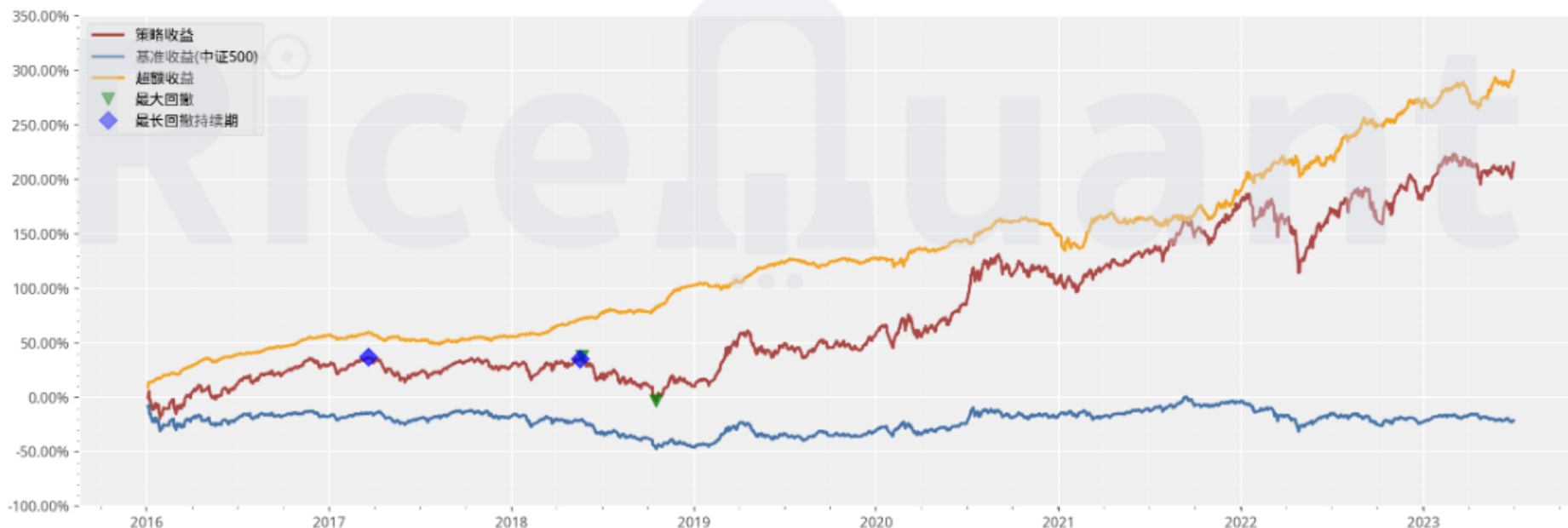
策略收益率	策略年化收益率	阿尔法	贝塔	夏普率	索提诺比率	周度累计回撤深度
192.470%	16.011%	0.1861	0.8937	0.6654	0.8622	8.497
基准收益率	基准年化收益率	波动率	跟踪误差	下行风险	信息比率	周度累计回撤夏普率
-21.253%	-3.252%	21.616%	8.692%	0.1668	2.262	0.162
超额收益率（算术）	胜率	周度胜率	盈亏比	最大回撤	最大回撤/最长回撤持续期	周度超额累计回撤深度
213.723%	56.892%	58.639%	0.8654	31.465%	MaxDD 2017-10-13~2018-10-18, 370 days MaxDD 2017-10-13~2019-03-06, 509 days	2.92
超额收益率（几何）	年化超额收益率（几何）	超额夏普率	超额收益波动率	超额收益最大回撤（几何）	超额收益最大回撤/最长回撤持续期	周度超额累计回撤夏普率
271.403%	19.911%	2.072	8.692%	14.494%	MaxDD 2016-10-23~2021-01-13, 187 days MaxDD 2020-10-23~2021-12-16, 419 days	0.6517



一、为什么要做深度学习端到端因子挖掘

深度学习端到端挖掘因子

策略收益率	策略年化收益率	阿尔法	贝塔	夏普率	索提诺比率	周度累计回撤深度
215.053%	17.211%	0.2004	0.9311	0.6992	0.9055	7.917
基准收益率	基准年化收益率	波动率	跟踪误差	下行风险	信息比率	周度累计回撤夏普率
-21.253%	-3.252%	22.246%	8.124%	0.1718	2.537	0.1971
超额收益率（算术）	胜率	周度胜率	盈亏比	最大回撤	最大回撤/最长回撤持续期	周度超额累计回撤深度
236.305%	57.825%	58.115%	0.8367	29.802%	MaxDD 2018-05-22~2019-10-18, 148 days MaxDD 2017-03-01~2018-05-18, 423 days	2.194
超额收益率（几何）	年化超额收益率（几何）	超额夏普率	超额收益波动率	超额收益最大回撤（几何）	超额收益最大回撤/最长回撤持续期	周度超额累计回撤夏普率
300.080%	21.151%	2.361	8.124%	11.547%	MaxDD 2020-10-27~2021-01-13, 78 days MaxDD 2017-03-21~2018-03-12, 356 days	0.9857



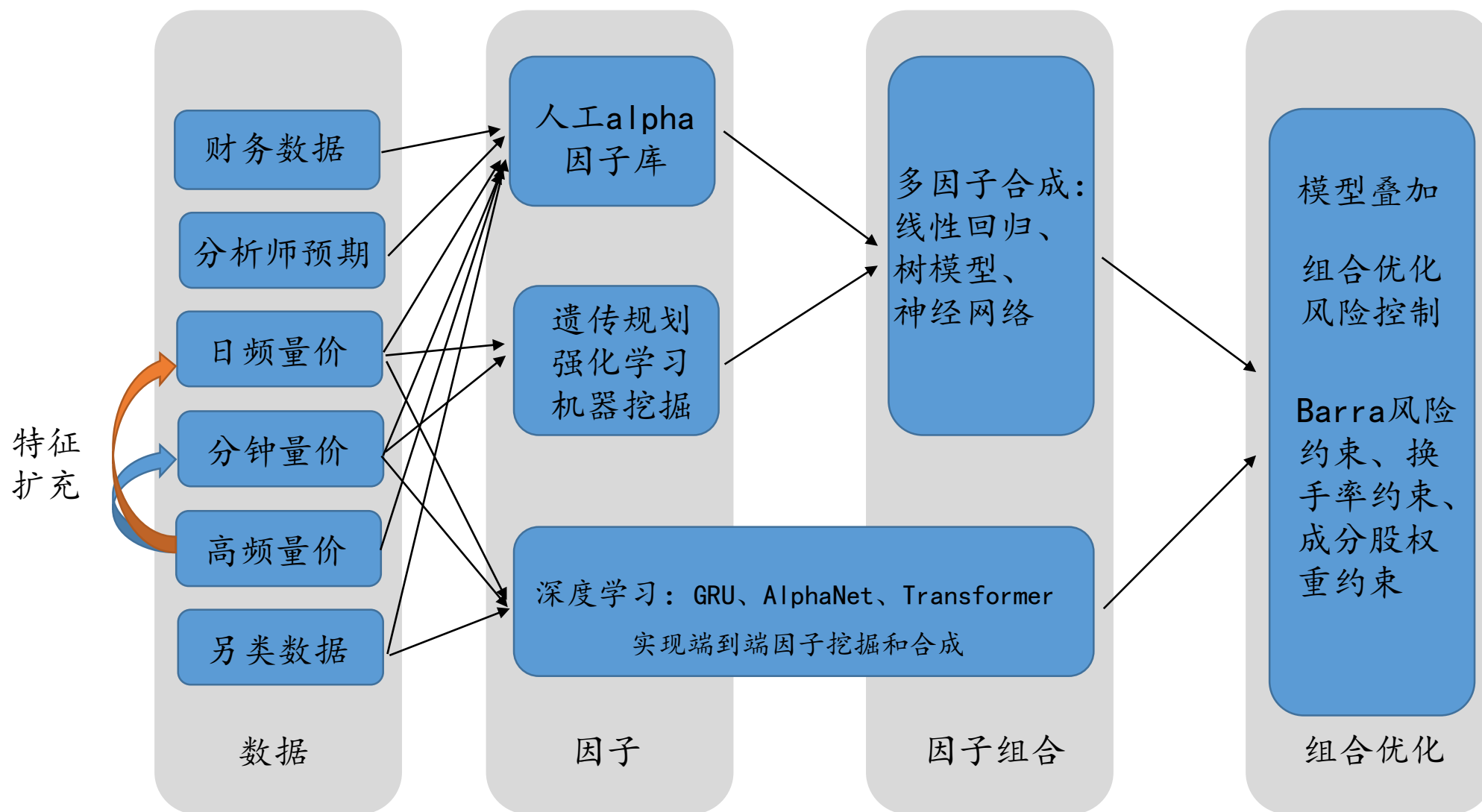
一、为什么要做深度学习端到端因子挖掘

人工挖掘合成因子+深度学习因子等权叠加



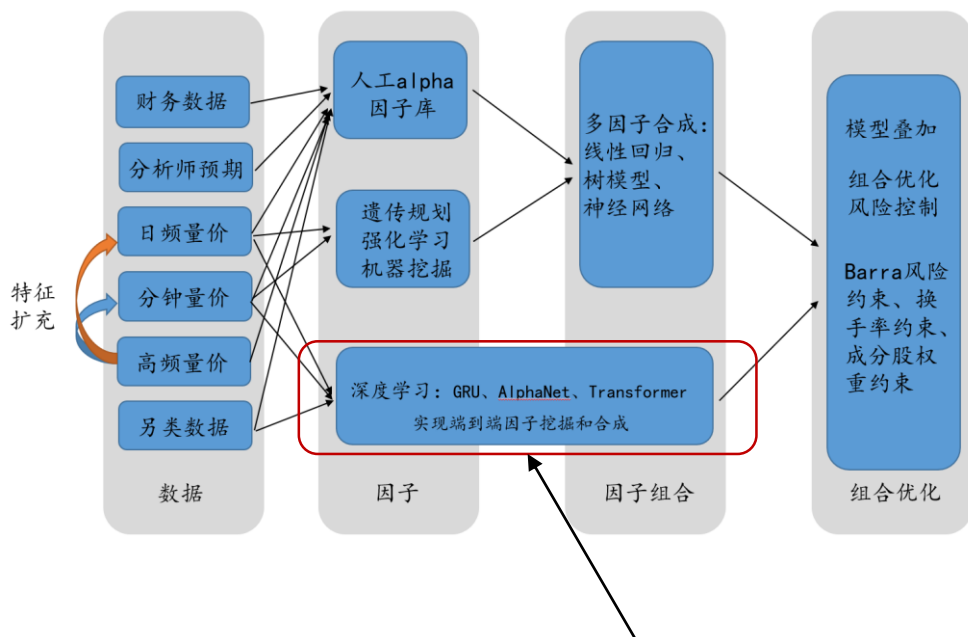
二、投研架构

人工智能融入多因子选股体系



二、投研架构

资深券商研究内容



(1) 华泰证券：之前的机器学习以及深度学习研究主要侧重点是多因子合成领域，近期开始布局端到端的深度学习因子挖掘领域。

(2) 海通证券：本身机器学习和深度学习相关的研报较少，近期的深度学习研报主要是端到端的深度学习因子挖掘领域，同时更偏向于利用高频数据进行模型训练也是布局最早的券商。

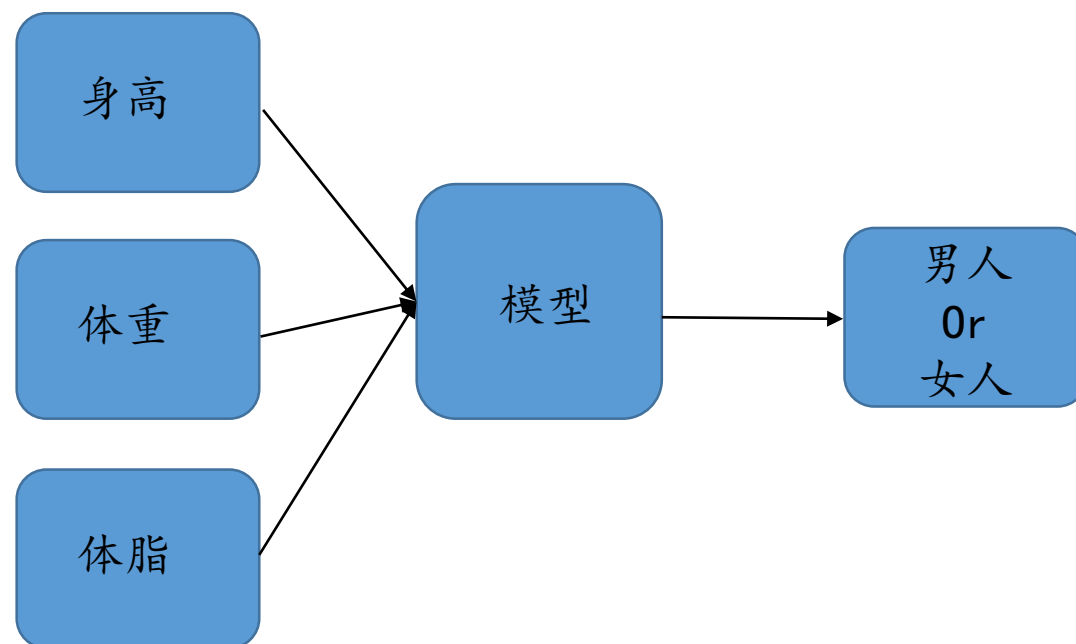
(3) 东方证券：主要侧重点有两块，一块是端到端的端到端的深度学习因子挖掘领域，另一块就是机器挖掘单因子这块，使用遗传算法和强化学习等方法论来基于原始数据构建单因子表达式。

本次汇报内容主要是端到端的深度学习因子挖掘，端到端指的是从原始数据端直接到合成因子端，横跨了因子和因子组合这两部分。因为该方法付出相同的努力对于组合的表现提升来的更加直接，性价比高。

三、端到端因子和传统因子组合深度学习模型区别

传统因子组合深度学习模型类比

◆ 打个比方，我们现在要去根据一个人（20岁）的身高、体重和体脂去预测这个人是男人还是女人



样本A: 身高 185cm 体重 75KG 体脂 18%

样本B: 身高 168cm 体重 50KG 体脂 28%

三、端到端因子和传统因子组合深度学习模型区别

传统因子组合深度学习模型类比

- ◆ 很显然，样本A是男生，样本B是女生。我们可以给出以下的公式作为我们的模型

$$Y = \text{身高} + \text{体重} - \text{体脂} * 100$$

If $y > 200$ 为男生,

Else 为女生

$$y_A = 185 + 75 - 0.18 * 100 = 242$$

$$y_B = 168 + 50 - 0.28 * 100 = 190$$

- ◆ 这和传统因子组合深度学习模型类似，我们采用当期的截面因子值（动量、波动率、流动性）输入进深度学习模型，得到未来收益率的预测值。

三、端到端因子和传统因子组合深度学习模型区别

传统因子组合深度学习模型类比

- ◆ 现在，我们有一个新的样本C，这是一个20岁的女性。

样本C：身高 173cm 体重 55KG 体脂 24%

$$y_C = 173 + 55 - 0.24 * 100 = 204$$

- ◆ 但是模型却告诉我们样本C是一个男性，造成误判的原因一方面可能是因为模型不够复杂，另一方面可能是单纯通过身高、体重和体脂来判断性别并不是很准确。

三、端到端因子和传统因子组合深度学习模型区别

端到端深度学习因子挖掘模型类比

样本A

年龄	身高	体重	体脂
11	145cm	36KG	25%
12	150cm	39kg	23%
13	157cm	42kg	23%
14	163cm	45kg	22%
15	172cm	51kg	20%
16	178cm	62kg	19%
17	180cm	65kg	20%
18	182cm	68kg	20%
19	184cm	72kg	19%
20	185cm	75KG	18%

发育期

时序
数据A

截面数据A

样本C

年龄	身高	体重	体脂
11	146cm	37KG	29%
12	154cm	40kg	29%
13	164cm	43kg	28%
14	167cm	46kg	27%
15	168cm	46kg	27%
16	169cn	48kg	27%
17	170cn	49kg	26%
18	171cm	51kg	25%
19	172cm	52kg	25%
20	173cm	55kg	24%

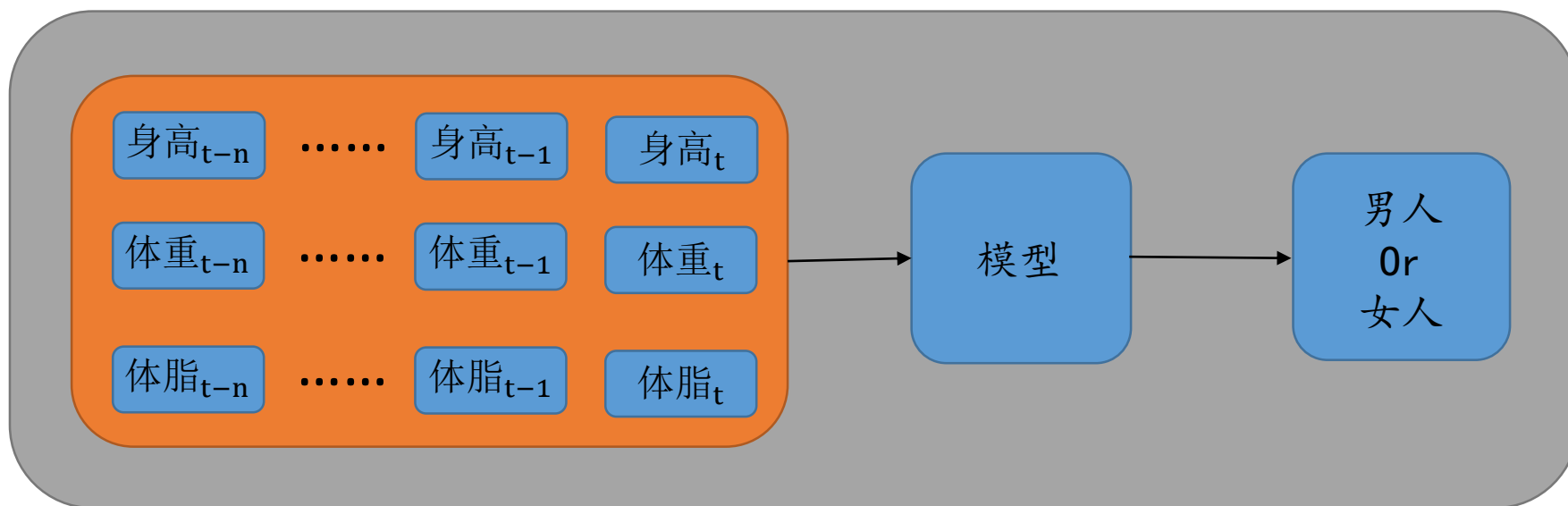
发育期

时序
数据C

截面数据C

三、端到端因子和传统因子组合深度学习模型区别

端到端深度学习因子挖掘模型类比

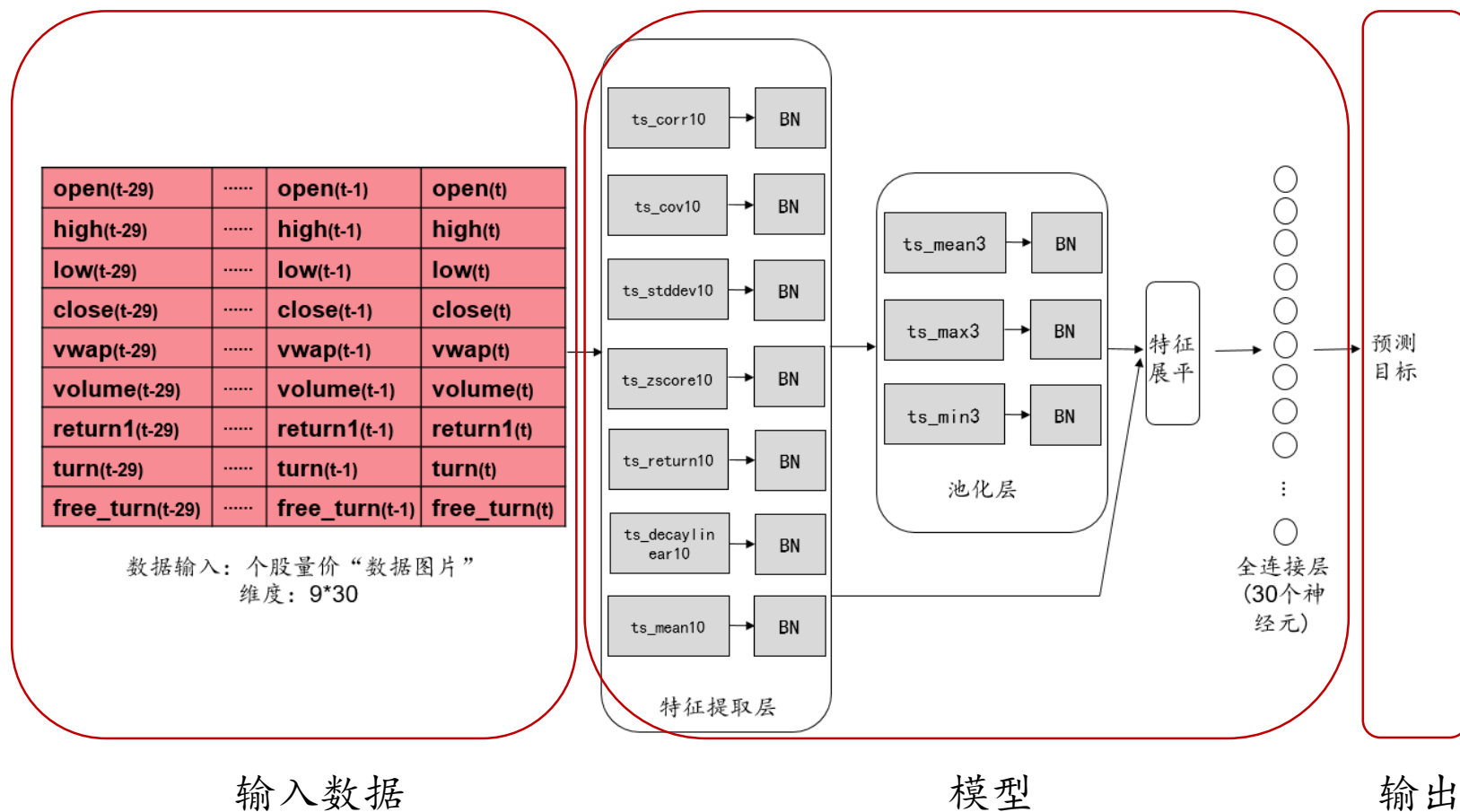


- ◆ 在端到端深度学习因子挖掘模型里面这边的时序身高、体重和体脂就被替换成了一段时间的分钟序列特征或者日频序列特征，或者是两者的混频数据。

四、华泰金工深度学习研报解析

AlphaNet系列 AlphaNet_V1

- ◆ 《20200614-华泰证券-华泰人工智能系列三十二： AlphaNet，因子挖掘神经网络》流程图



四、华泰金工深度学习研报解析

AlphaNet系列 AlphaNet_V1

◆ 《20200614-华泰证券-华泰人工智能系列三十二： AlphaNet，因子挖掘神经网络》模型

名称	定义
ts_corr(X, Y, d)	过去 d 天 X 值构成的时序数列和 Y 值构成的时序数列的相关系数。
ts_cov(X, Y, d)	过去 d 天 X 值构成的时序数列和 Y 值构成的时序数列的协方差。
ts_stddev(X, d)	过去 d 天 X 值构成的时序数列的标准差。
ts_zscore(X, d)	过去 d 天 X 值构成的时序数列的平均值除以标准差。
ts_return(X, d)	$(X - \text{delay}(X, d)) / \text{delay}(X, d) - 1$, $\text{delay}(X, d)$ 为 X 在 d 天前的取值。
ts_decaylinear(X, d)	过去 d 天 X 值构成的时序数列的加权平均值，权数为 $d, d - 1, \dots, 1$ (权数之和应为 1，需进行归一化处理)，其中离现在越近的日子权数越大。
ts_min(X, d)	过去 d 天 X 值构成的时序数列中最小值。
ts_max(X, d)	过去 d 天 X 值构成的时序数列中最大值。
ts_sum(X, d)	过去 d 天 X 值构成的时序数列之和。
BN	批标准化层(Batch Normalization)，用来标准化以上自定义网络层提取的特征，使得不同特征在量纲上趋于一致。

◆ 过去20日的量价相关性为 `ts_corr(close, volume, 20)`。

四、华泰金工深度学习研报解析

AlphaNet系列 AlphaNet_V1

◆ 《20200614-华泰证券-华泰人工智能系列三十二： AlphaNet，因子挖掘神经网络》模型

在时间维度上步进，与CNN类似，步进大小stride是可调参数

open(t-4)	open(t-3)	open(t-2)	open(t-1)	open(t)
high(t-4)	high(t-3)	high(t-2)	high(t-1)	high(t)
low(t-4)	low(t-3)	low(t-2)	low(t-1)	low(t)
close(t-4)	close(t-3)	close(t-2)	close(t-1)	close(t)
vwap(t-4)	vwap(t-3)	vwap(t-2)	vwap(t-1)	vwap(t)
volume(t-4)	volume(t-3)	volume(t-2)	volume(t-1)	volume(t)
return1(t-4)	return1(t-3)	return1(t-2)	return1(t-1)	return1(t)
turn(t-4)	turn(t-3)	turn(t-2)	turn(t-1)	turn(t)
free_turn(t-4)	free_turn(t-3)	free_turn(t-2)	free_turn(t-1)	free_turn(t)

open(t-4)	open(t-3)	open(t-2)	open(t-1)	open(t)
high(t-4)	high(t-3)	high(t-2)	high(t-1)	high(t)
low(t-4)	low(t-3)	low(t-2)	low(t-1)	low(t)
close(t-4)	close(t-3)	close(t-2)	close(t-1)	close(t)
vwap(t-4)	vwap(t-3)	vwap(t-2)	vwap(t-1)	vwap(t)
volume(t-4)	volume(t-3)	volume(t-2)	volume(t-1)	volume(t)
return1(t-4)	return1(t-3)	return1(t-2)	return1(t-1)	return1(t)
turn(t-4)	turn(t-3)	turn(t-2)	turn(t-1)	turn(t)
free_turn(t-4)	free_turn(t-3)	free_turn(t-2)	free_turn(t-1)	free_turn(t)

ts_corr(X,Y,3)

F1	
F2	
.....	

运算的结果：
“特征图片”

在特征维度上
遍历，与CNN
卷积不同，不
是局部感知

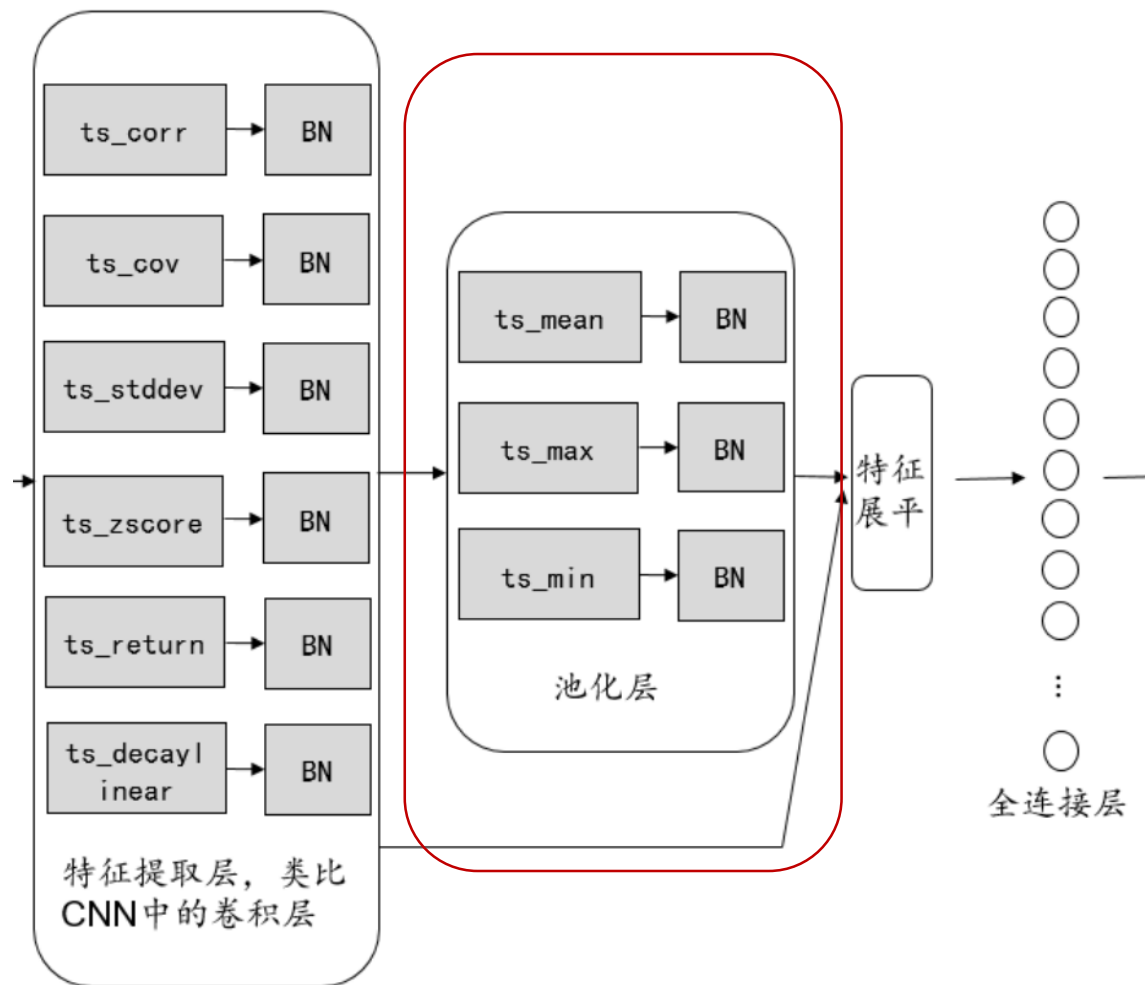
(1) 对于过去30个交易日的量价数据，取 $d=10$ ， $\text{stride}=10$ 。对于ts_corr这类需要X和Y的算子， $9*30$ 维度的量价图片数据会转换成 $36*3$ 维度的降维数据。横轴维度36是由于两两特征算相关性， $C_9^2 = 36$ 。纵轴维度3是由于总共30个交易日，步长为10意味着对于close和volume的相关性，我们会用 $t \sim t-9$ 、 $t-10 \sim t-19$ 、 $t-20 \sim t-29$ 这三个区间计算三个相关系数。

(2) 对于ts_zscore这类只需要X的算子， $9*30$ 维度的量价图片数据会转换成 $9*3$ 维度的降维数据。

四、华泰金工深度学习研报解析

AlphaNet系列 AlphaNet_V1

◆ 《20200614-华泰证券-华泰人工智能系列三十二： AlphaNet，因子挖掘神经网络》模型



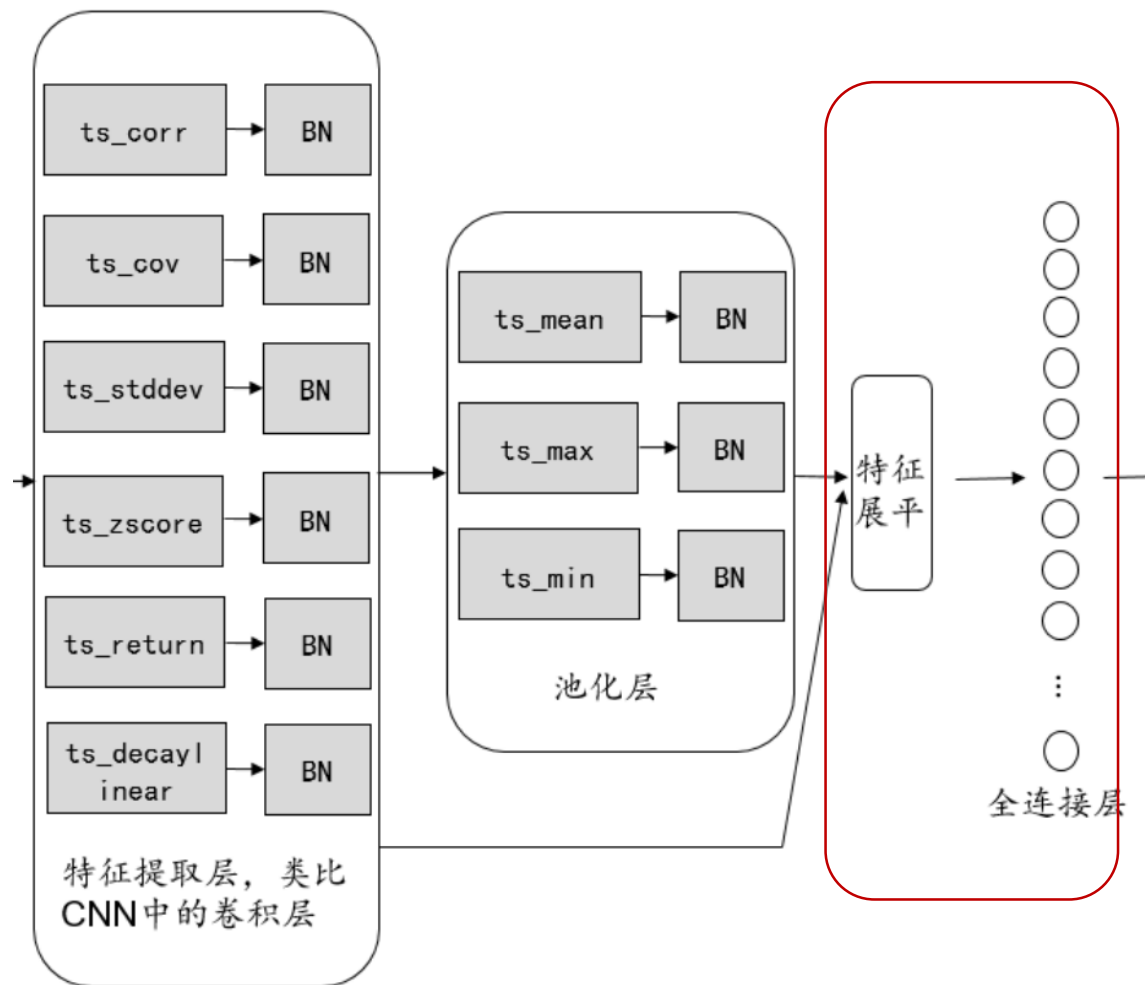
(1) 池化, 也叫Pooling, 其本质其实就是采样, 池化对于输入的图片, 选择某种方式对其进行压缩, 以加快神经网络的运算速度。这里说的某种方式, 其实就是池化的算法, 比如最大池化或平均池化。

(2) 池化层会将我们上面计算的`ts_corr`这类算子得到的 36×3 维度的矩阵和`ts_zscore`这类算子得到的 9×3 维度矩阵沿着横轴分别求均值、最大值和最小值, 36×3 维度的矩阵变成 36×1 , 以此类推。

四、华泰金工深度学习研报解析

AlphaNet系列 AlphaNet_V1

◆ 《20200614-华泰证券-华泰人工智能系列三十二： AlphaNet，因子挖掘神经网络》模型



(1) 最后模型将第一个特征提取层得到的特征向量和池化层得到的特征向量全部展平成一维向量输入进全连接层，输出对于股票的预测值。

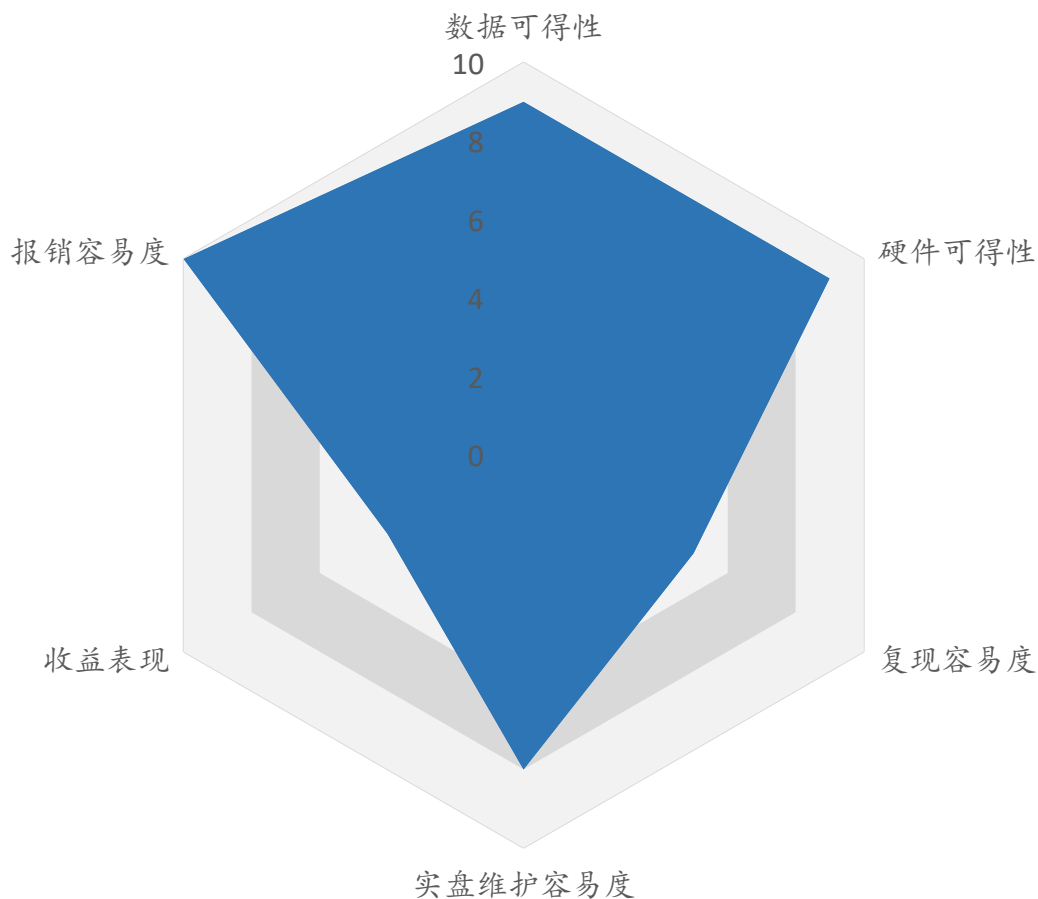
(2) 之所以将信息提取层和池化层出来的所以特征展平堆叠是因为我们不想遗漏任何有用的信息。

(3) 全连接层能够对展平的特征实现加权求和，和线性回归类似。

四、华泰金工深度学习研报解析

AlphaNet系列 AlphaNet_V1

◆ 《20200614-华泰证券-华泰人工智能系列三十二： AlphaNet，因子挖掘神经网络》雷达图



(1) 数据可得性 9分 日频开高收低量等数据获取非常容易。

(2) 硬件可得性 9分 模型不复杂，基本上正常的笔记本电脑都能跑起来。

(3) 复现容易度 5分 相比于传统神经网络，这里的AlphaNet是自定义的不能直接调包，复现麻烦。

(4) 收益表现 4分 由于数据和模型均不复杂，因此表现也只能算一般

(5) 实盘维护容易度 9分 数据盘后增量更新容易，每季度训练新模型开销不大。

(6) 报销容易度9分 基本上不存在额外的支出，当前公募拥有的资源就能跑起来。

四、华泰金工深度学习研报解析

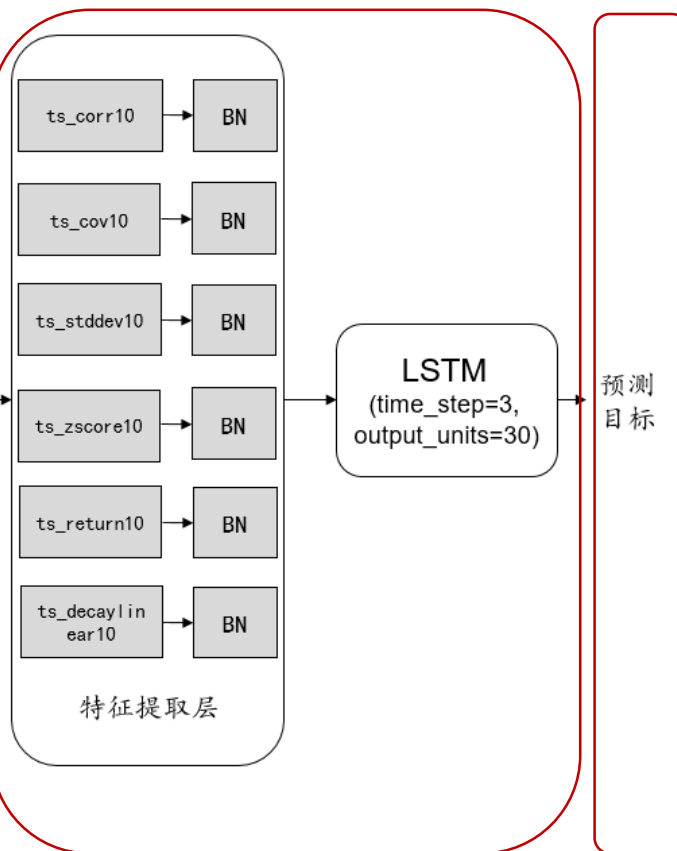
AlphaNet系列 AlphaNet_V2

◆ 《20200824-华泰证券-华泰人工智能系列之三十四：再探 AlphaNet，结构和特征优化》

open(t-29)		open(t-1)	open(t)
high(t-29)		high(t-1)	high(t)
low(t-29)		low(t-1)	low(t)
close(t-29)		close(t-1)	close(t)
vwap(t-29)		vwap(t-1)	vwap(t)
volume(t-29)		volume(t-1)	volume(t)
return1(t-29)		return1(t-1)	return1(t)
turn(t-29)		turn(t-1)	turn(t)
free_turn(t-29)		free_turn(t-1)	free_turn(t)
close(t-29)/free_turn(t-29)		close(t-1)/free_turn(t-1)	close(t)/free_turn(t)
open(t-29)/turn(t-29)		open(t-1)/turn(t-1)	open(t)/turn(t)
volume(t-29)/low(t-29)		volume(t-1)/low(t-1)	volume(t)/low(t)
vwap(t-29)/high(t-29)		vwap(t-1)/high(t-1)	vwap(t)/high(t)
low(t-29)/high(t-29)		low(t-1)/high(t-1)	low(t)/high(t)
vwap(t-29)/close(t-29)		vwap(t-1)/close(t-1)	vwap(t)/close(t)

数据输入：个股量价“数据图片”
维度：15*30

输入数据



模型

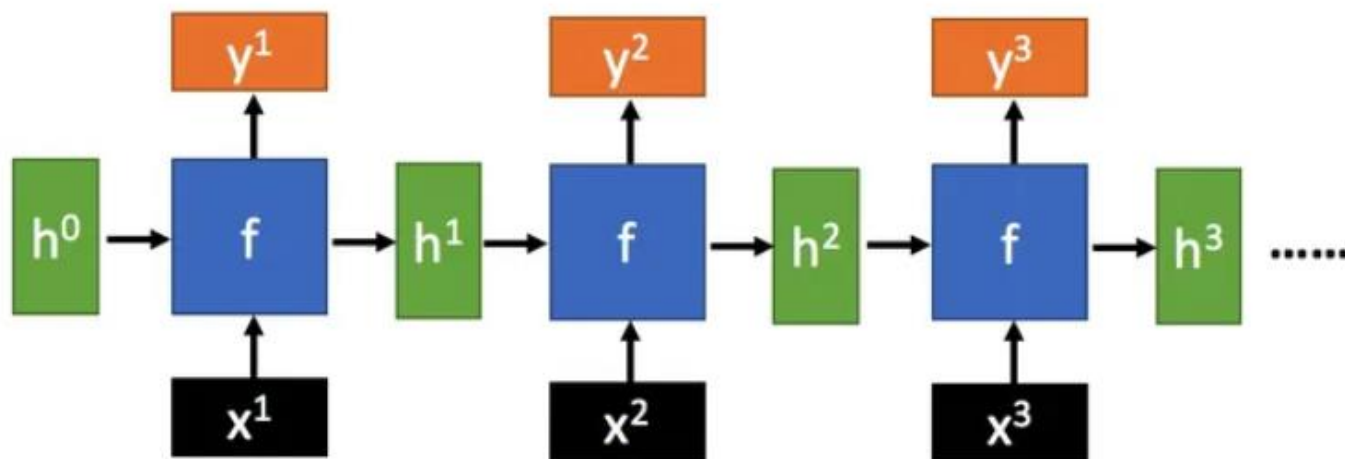
输出

- (1) V2相比于V1增量了6个比率特征。
- (2) 将池化层替换成LSTM来提取信息。

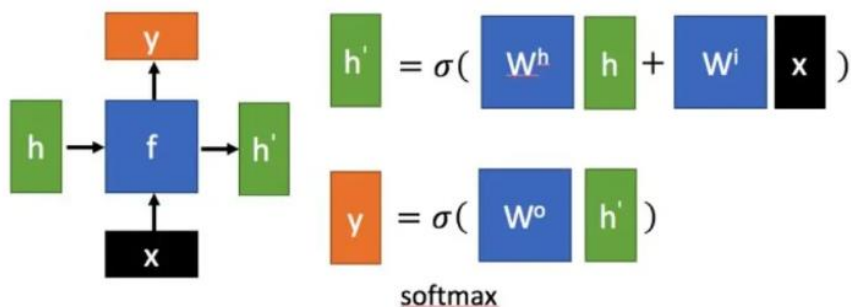
四、华泰金工深度学习研报解析

循环神经网络介绍

◆ RNN模型



RNN序列



RNN节点

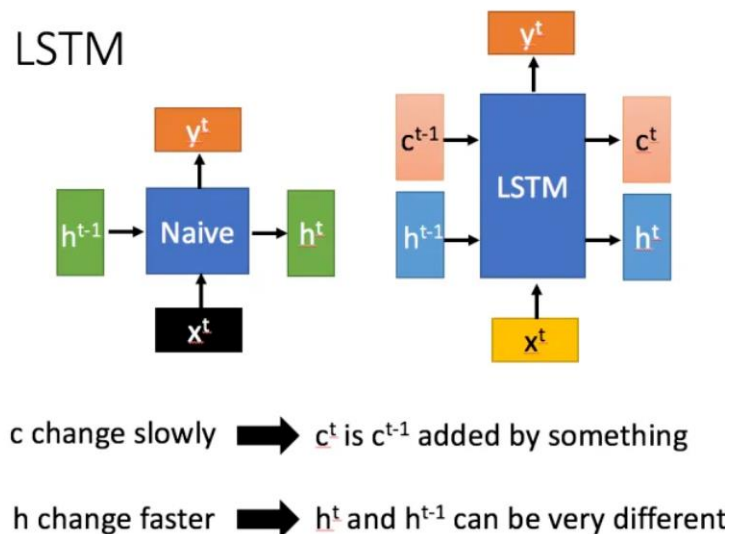
(1) RNN是神经网络的一种，对具有序列特性的数据非常有效，它能挖掘数据中的时序信息以及语义信息，利用了RNN的这种能力，使深度学习模型在解决语音识别、语言模型、机器翻译以及时序分析等NLP领域的问题时有所突破。

(2) 输入的 X^t 表示 t 期的特征（开高收低量等）。

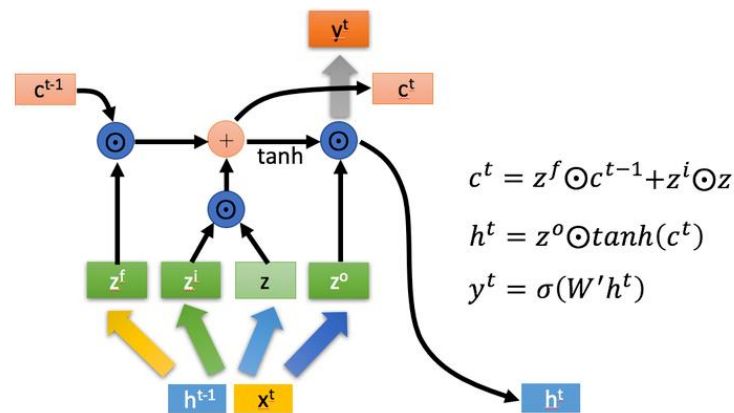
四、华泰金工深度学习研报解析

长短期记忆神经网络（LSTM）介绍

◆ LSTM模型



LSTM与普通RNN节点对比



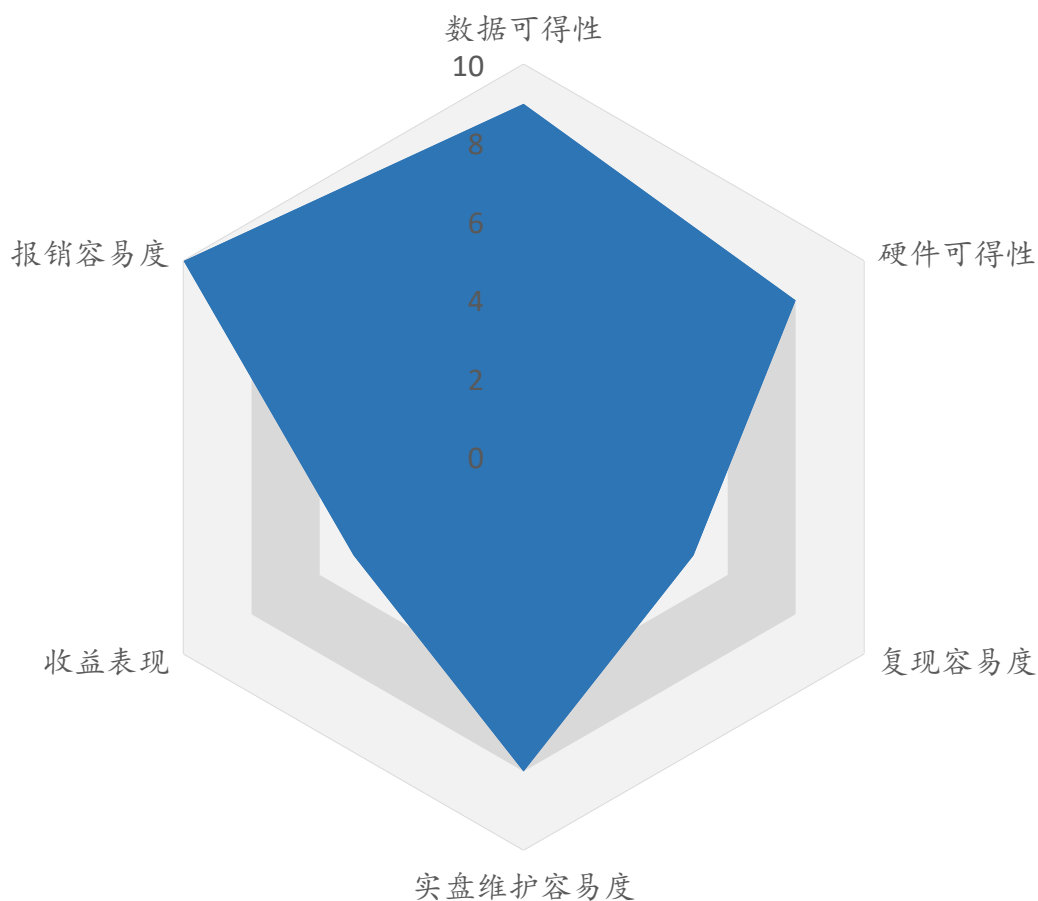
(1) 长短期记忆（Long short-term memory, LSTM）是一种特殊的RNN，主要是为了解决长序列训练过程中的梯度消失和梯度爆炸问题。简单来说，就是相比普通的RNN，LSTM能够在更长的序列中有更好的表现。

(2) 引入输入门、遗忘门和输出门，有选择的记忆有效信息，避免序列过长导致的信息遗忘。

四、华泰金工深度学习研报解析

AlphaNet系列 AlphaNet_V2

◆ 《20200824-华泰证券-华泰人工智能系列之三十四：再探 AlphaNet，结构和特征优化》雷达图



(1) 数据可得性 9分 日频开高收低量等数据获取非常容易。

(2) 硬件可得性 8分 模型不复杂，基本上正常的笔记本电脑都能跑起来，引入LSTM训练时间增加较多，但CPU依然能跑。

(3) 复现容易度 5分 和AlphaNet_V1类似，但是如果完成了AlphaNet_V1的复现，V2基本上没有额外的难度增加。

(4) 收益表现 5分 相较于V1略好

(5) 实盘维护容易度 9分 数据盘后增量更新容易，每季度训练新模型开销不大。

(6) 报销容易度9分 基本上不存在额外的支出，当前公募拥有的资源就能跑起来。

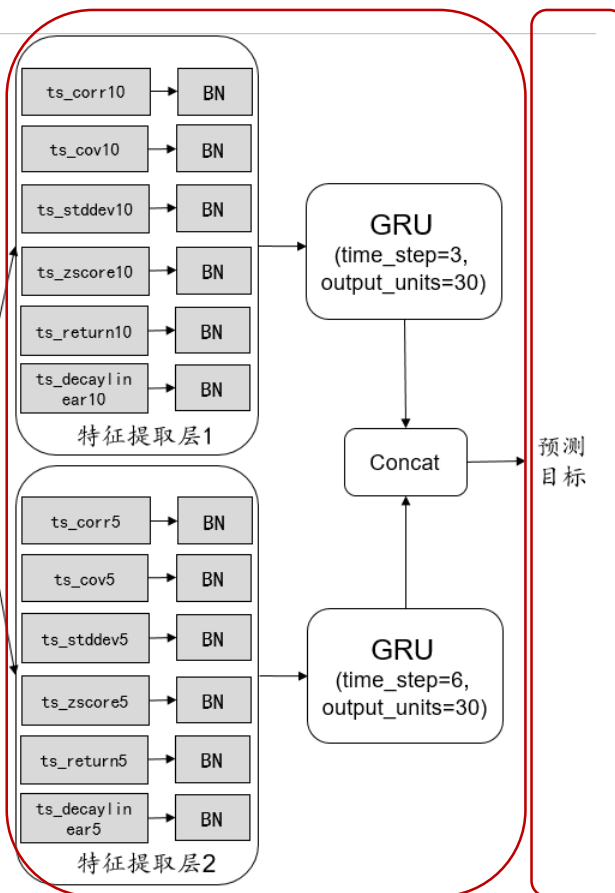
四、华泰金工深度学习研报解析

AlphaNet系列 AlphaNet_V3

◆ 《20200824-华泰证券-华泰人工智能系列之三十四：再探 AlphaNet，结构和特征优化》

open(t-29)		open(t-1)	open(t)
high(t-29)		high(t-1)	high(t)
low(t-29)		low(t-1)	low(t)
close(t-29)		close(t-1)	close(t)
vwap(t-29)		vwap(t-1)	vwap(t)
volume(t-29)		volume(t-1)	volume(t)
return1(t-29)		return1(t-1)	return1(t)
turn(t-29)		turn(t-1)	turn(t)
free_turn(t-29)		free_turn(t-1)	free_turn(t)
close(t-29)/free_turn(t-29)		close(t-1)/free_turn(t-1)	close(t)/free_turn(t)
open(t-29)/turn(t-29)		open(t-1)/turn(t-1)	open(t)/turn(t)
volume(t-29)/low(t-29)		volume(t-1)/low(t-1)	volume(t)/low(t)
vwap(t-29)/high(t-29)		vwap(t-1)/high(t-1)	vwap(t)/high(t)
low(t-29)/high(t-29)		low(t-1)/high(t-1)	low(t)/high(t)
vwap(t-29)/close(t-29)		vwap(t-1)/close(t-1)	vwap(t)/close(t)

数据输入：个股量价“数据图片”
维度：15*30



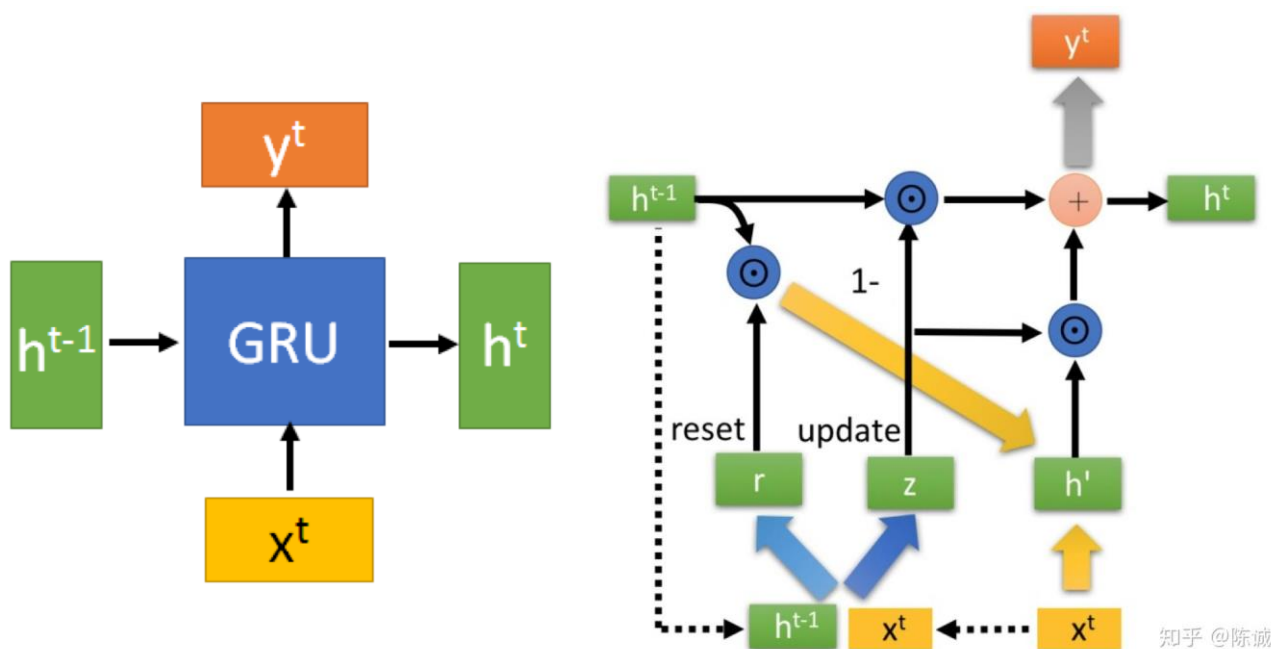
(1) V3和V2在特征上保持一致。

(2) 新增加一个5日窗口的特征提取层。
(3) 将LSTM更换成GRU模型，并将5日和10日的叠加得到最终预测目标。

四、华泰金工深度学习研报解析

门控循环神经网络 (GRU) 介绍

◆ GRU模型

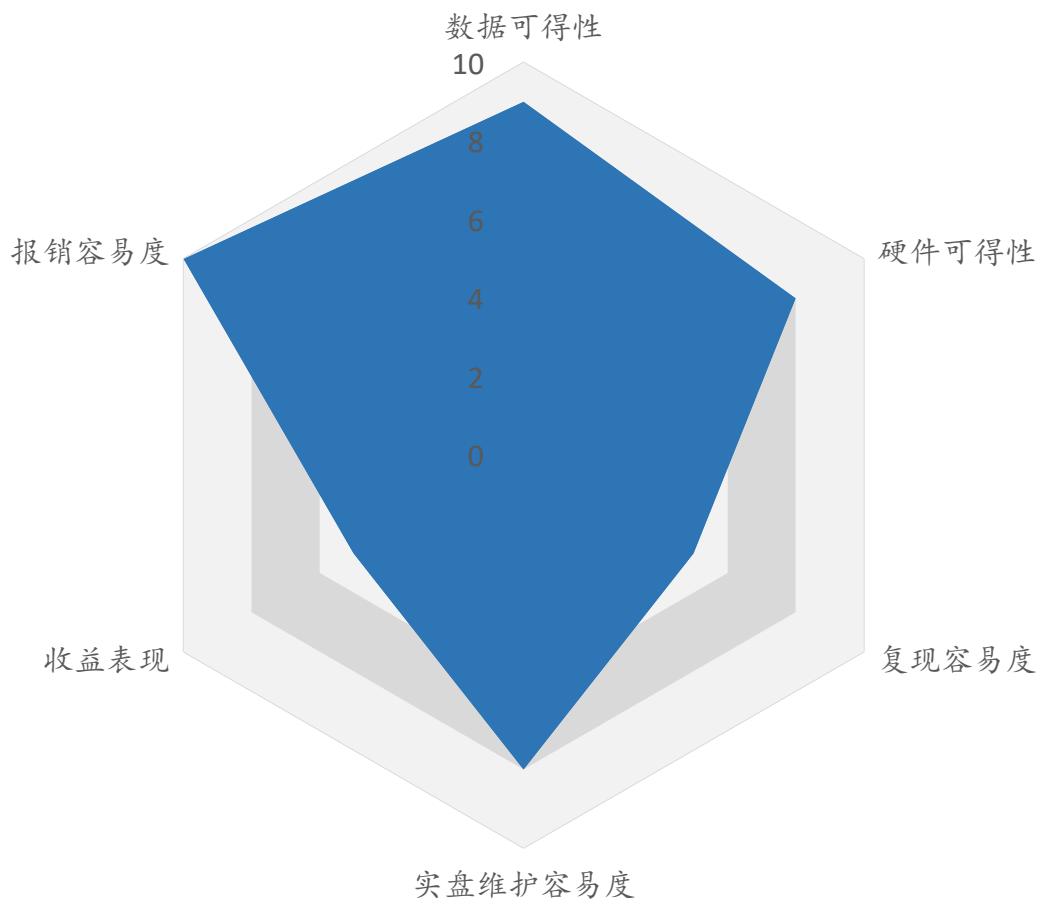


(1) GRU (Gate Recurrent Unit) 是循环神经网络 (Recurrent Neural Network, RNN) 的一种。和LSTM (Long-Short Term Memory) 一样，也是为了解决长期记忆和反向传播中的梯度等问题而提出来的。

四、华泰金工深度学习研报解析

AlphaNet系列 AlphaNet_V3

◆ 《20200824-华泰证券-华泰人工智能系列之三十四：再探 AlphaNet，结构和特征优化》雷达图



(1) 数据可得性 9分 日频开高收低量等数据获取非常容易。

(2) 硬件可得性 8分 模型不复杂，基本上正常的笔记本电脑都能跑起来，引入GRU训练时间增加较多，但CPU依然能跑。

(3) 复现容易度 5分 在V2的基础上无需过多的修改

(4) 收益表现 5分 相较于V1略好，和V2表现类似，换手率会有一定增加。

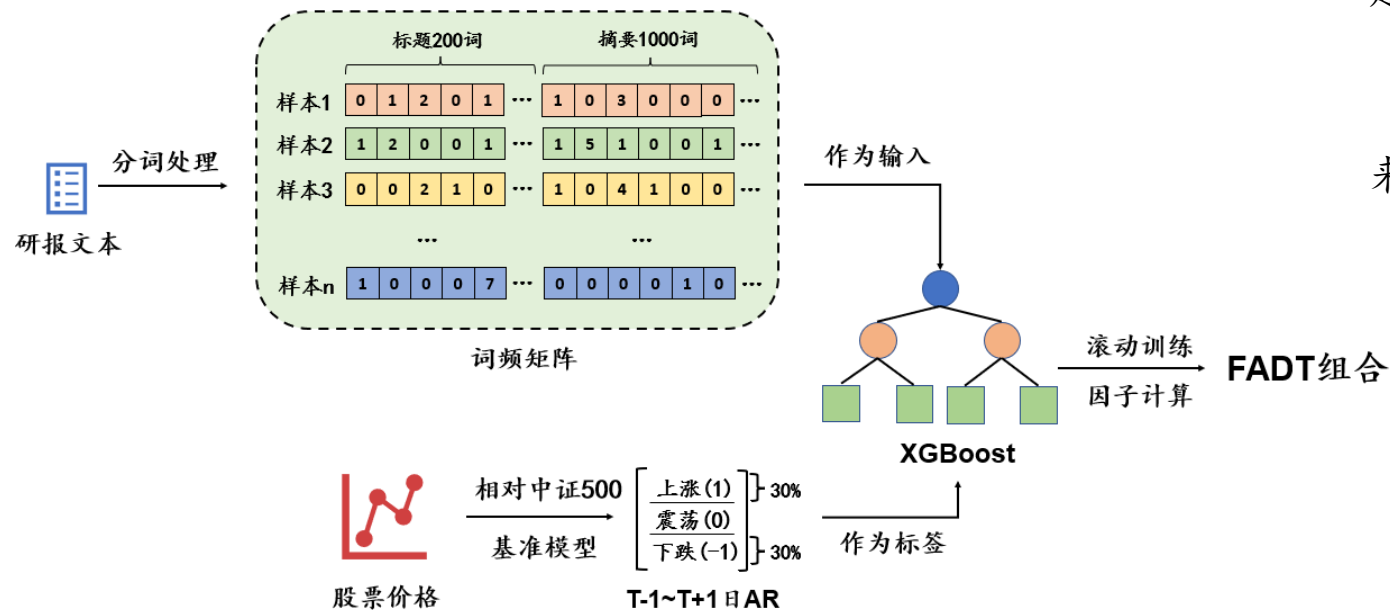
(5) 实盘维护容易度 9分 数据盘后增量更新容易，每季度训练新模型开销不大。

(6) 报销容易度9分 基本上不存在额外的支出，当前公募拥有的资源就能跑起来。

四、华泰金工深度学习研报解析

分析师研报文本Bert模型

◆ 《20221028-华泰证券-人工智能63：再探文本FADT选股》模型



(1) 在华泰早起分析研报文本因子研究中，通过统计标题和文本当中关键词的个数来提取文本信息。

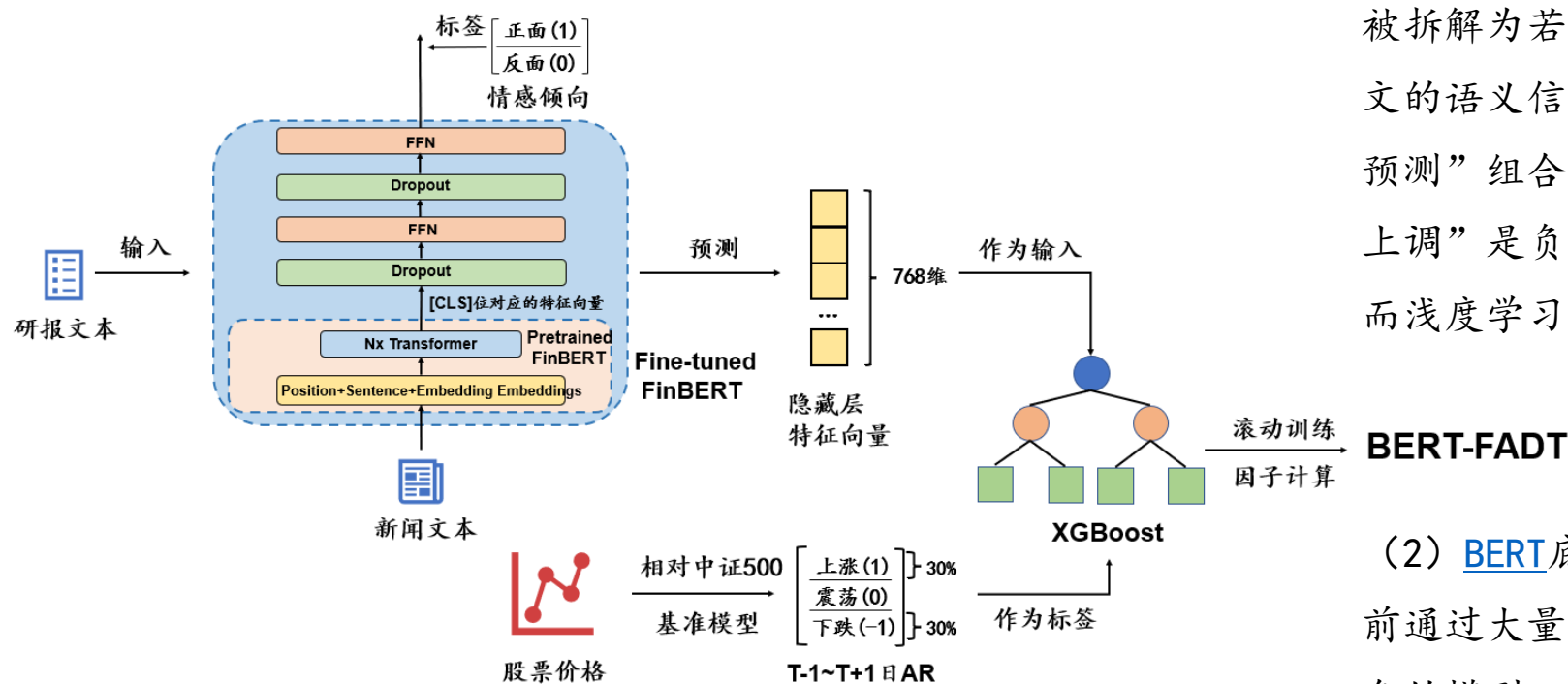
(2) 将提取的词频向量输入到XGBoost模型当中，以未来一段时间个股收益为预测目标，进行滚动训练。

基于词频向量+XGBoost的分析师研报文本挖掘示意图

四、华泰金工深度学习研报解析

分析师研报文本Bert模型

◆ 《20221028-华泰证券-人工智能63：再探文本FADT选股》模型



基于FinBERT+XGBoost的分析师研报文本挖掘示意图

(1) 在原始的词频向量提取信息时，一句完整的话最后被拆解为若干词语的词频，在这个过程中容易损失上下文的语义信息，如“上调”这个词和“成本”或“盈利预测”组合在一起时将表达出完全相反的含义（“成本上调”是负向含义，“盈利预测上调”是正向含义），而浅度学习模型似乎无法学习这种词语组合。

(2) **BERT**底层基于Transformer的网络结构，是一个提前通过大量文本语料数据库训练出来的通用的解决NLP任务的模型。

四、华泰金工深度学习研报解析

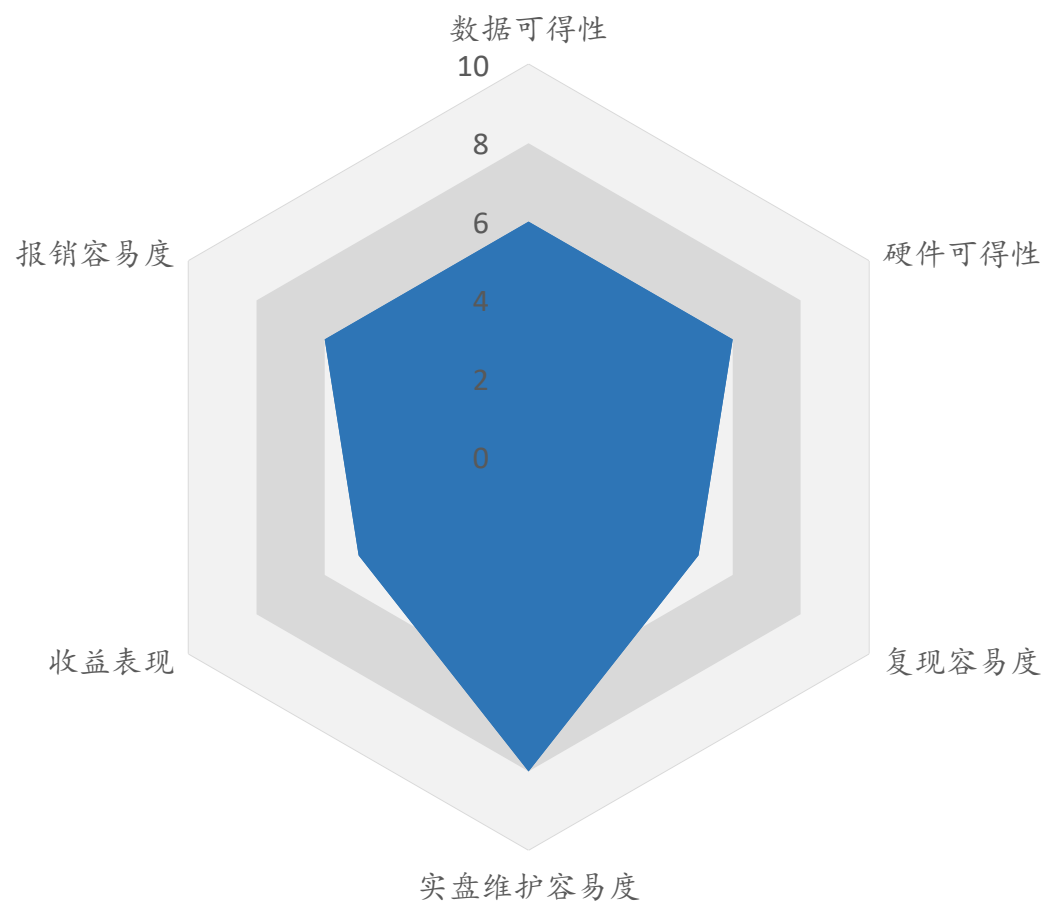
■ 分析师研报文本Bert模型

- ◆ 微调预训练模型：使用带标注的新闻舆情数据对FinBERT进行微调，使得FinBERT可以在测试集上获得较高的预测准确率；
- ◆ 对研报文本进行语义编码：采用上述训练好的FinBERT对研报文本进行编码，输出FinBERT的CLS层768 维向量，替代原版的词频向量；
- ◆ 特征向量二次训练：将上述得到的特征向量作为特征输入给XGBoost模型，标签与原版标签相同，因子构建方式也与原版相同，训练XGBoost模型并构建因子值。

四、华泰金工深度学习研报解析

分析师研报文本Bert模型

◆ 《20221028-华泰证券-人工智能63：再探文本FADT选股》雷达图



(1) 数据可得性 6分 需要采购朝阳永续的分析师文本数据。

(2) 硬件可得性 6分 Bert模型哪怕微调也需要用到不小的算力，需要一块显存在12GB以上的GPU才能微调模型。

(3) 复现容易度 5分 整个数据预处理以及模型训练流程步骤较多且繁琐。

(4) 收益表现 5分 相模型多头表现不错，但是从22年下旬之后模型超额表现不佳，和基本面因子普遍失效有关。

(5) 实盘维护容易度 8分 数据盘后增量更新容易，每季度训练新模型开销不大。

(6) 报销容易度 6分 基需要采购朝阳永续数据，同时还需要GPU服务器，对CPU和存储要求不高。

四、华泰金工深度学习研报解析

■ 华泰报告总结

- ◆ 华泰的端到端深度学习因子挖掘在数据上并没有创新，而是在模型上引入了AlphaNet这类网络结构，V1是AlphaNet进行信息提取加上池化层聚合信息，V2是将池化层改用LSTM来聚合信息，V3在AlphaNet回看10日的基础上增加了回看5日的部分，并改用GRU来聚合信息。
- ◆ 华泰的分析师研报文本Bert模型将原来通过词频向量提取文本信息改用Bert模型提取文本信息，将CLS分类层特征向量利用XGBoost模型二次训练用以预测股票未来收益。

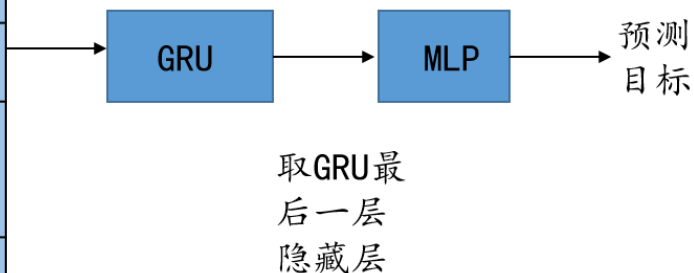
五、海通金工深度学习研报解析

海通高频GRU_V1

- ◆ 《20211228-海通证券-选股因子系列研究（七十六）：基于深度学习的高频因子挖掘》模型

Open(t-n)		Open(t-1)	Open(t)
close(t-n)		close(t-1)	close(t)
vol(t-n)		vol(t-1)	vol(t)
OI(t-n)		OI(t-1)	OI(t)
.	.	.	.
.	.	.	.
.	.	.	.
Mf_netflow(t-n)		Mf_netflow(t-1)	Mf_netflow(t)

数据输入：个股30分钟价量数据，
序列长度160（20天）



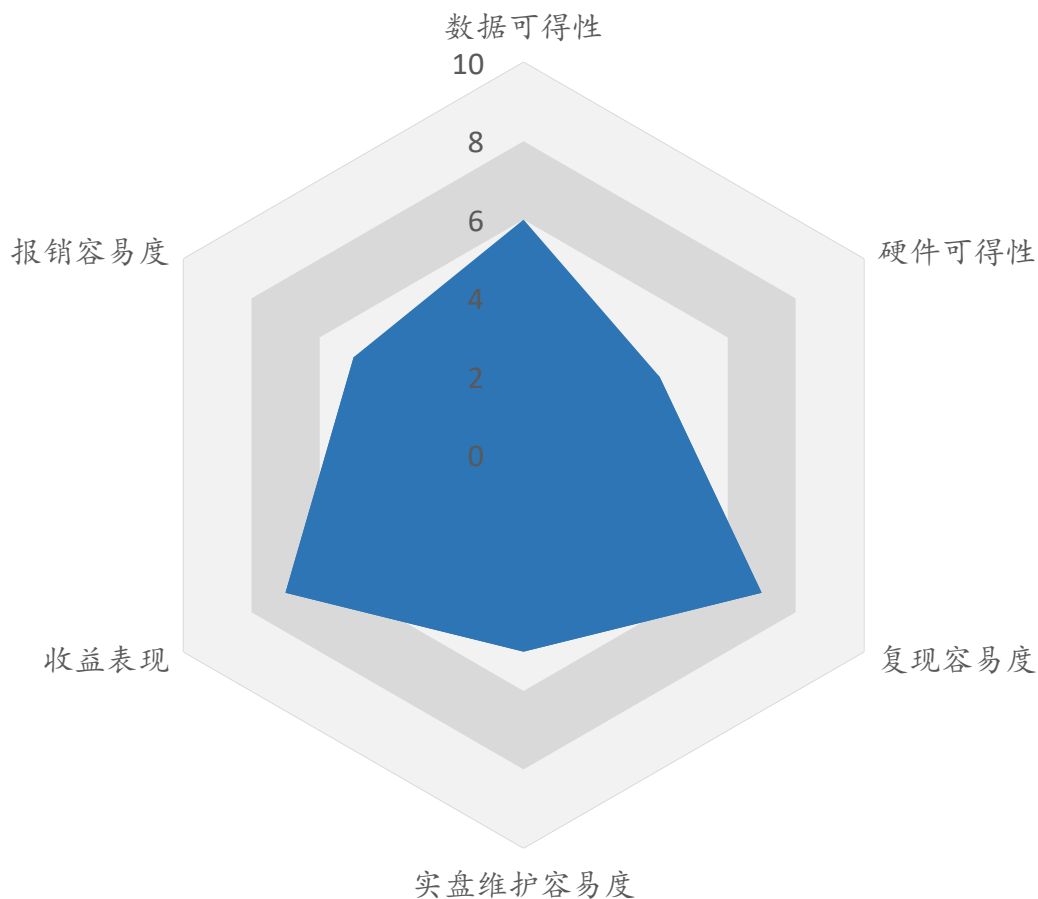
（1）在将日频的序列特征升频到30分钟，特征数量扩展至164个，序列长度由原来的30扩展到160。

（2）相比于AlphaNet去掉了特征提取层，直接改用更强大的GRU来提取信息。

五、海通金工深度学习研报解析

海通高频GRU_V1

◆ 《20211228-海通证券-选股因子系列研究（七十六）：基于深度学习的高频因子挖掘》雷达图



(1) 数据可得性 5分 需要逐笔成交等高频数据来做分钟特征扩充。

(2) 硬件可得性 4分 分钟序列较长使得GRU训练数据也相应增加，存储和处理逐笔数据需要高性能的CPU和大容量的固态硬盘（20TB为佳），模型训练将数据读入内存所需空间大。

(3) 复现容易度 7分 网络实现并不复杂，主要是分钟特征的制作，当然分钟开高收低等原始特征稍加处理也能直接进模型

(4) 收益表现 7分 更高频的数据和更多的特征哪怕在网络并不复杂的情况下也能带来不小的收益提升。

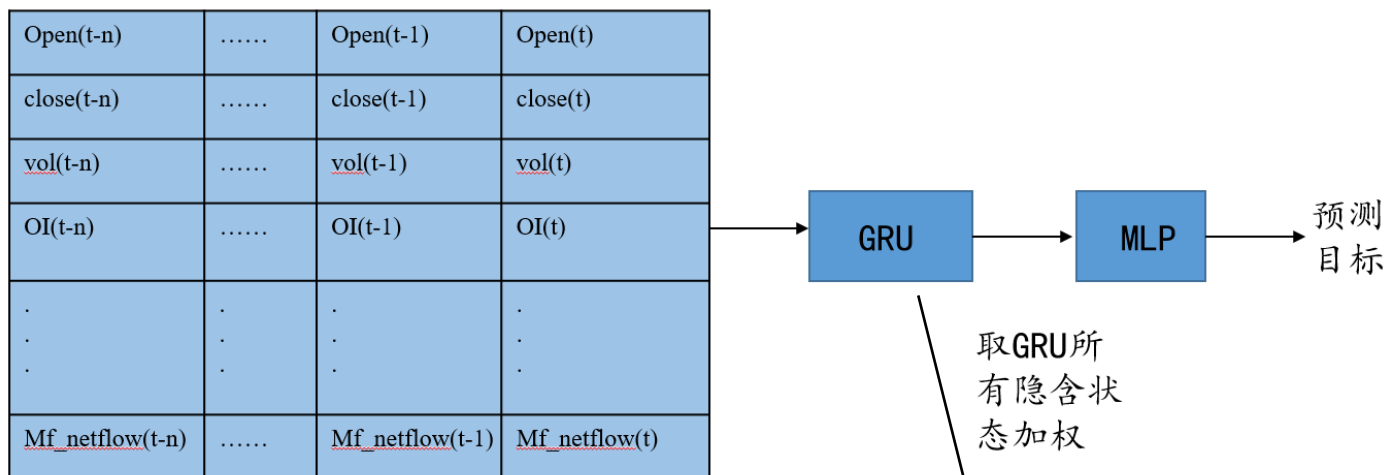
(5) 实盘维护容易度 5分 盘后需要维护逐笔数据以及通过逐笔数据扩充的上百个分钟特征，流程较为繁琐。

(6) 报销容易度 5分 逐笔数据和硬件开销均较大，通联逐笔数据盘后10w+一年，服务器采购费用6w+。

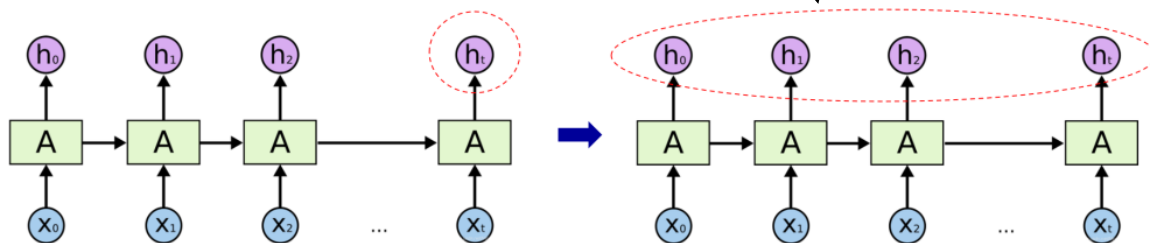
五、海通金工深度学习研报解析

海通高频GRU_V2

◆ 《20220531-海通证券-选股因子系列研究（七十九）：用注意力机制优化深度学习高频因子》模型



数据输入：个股10分钟价量数据，
序列长度480（20天）



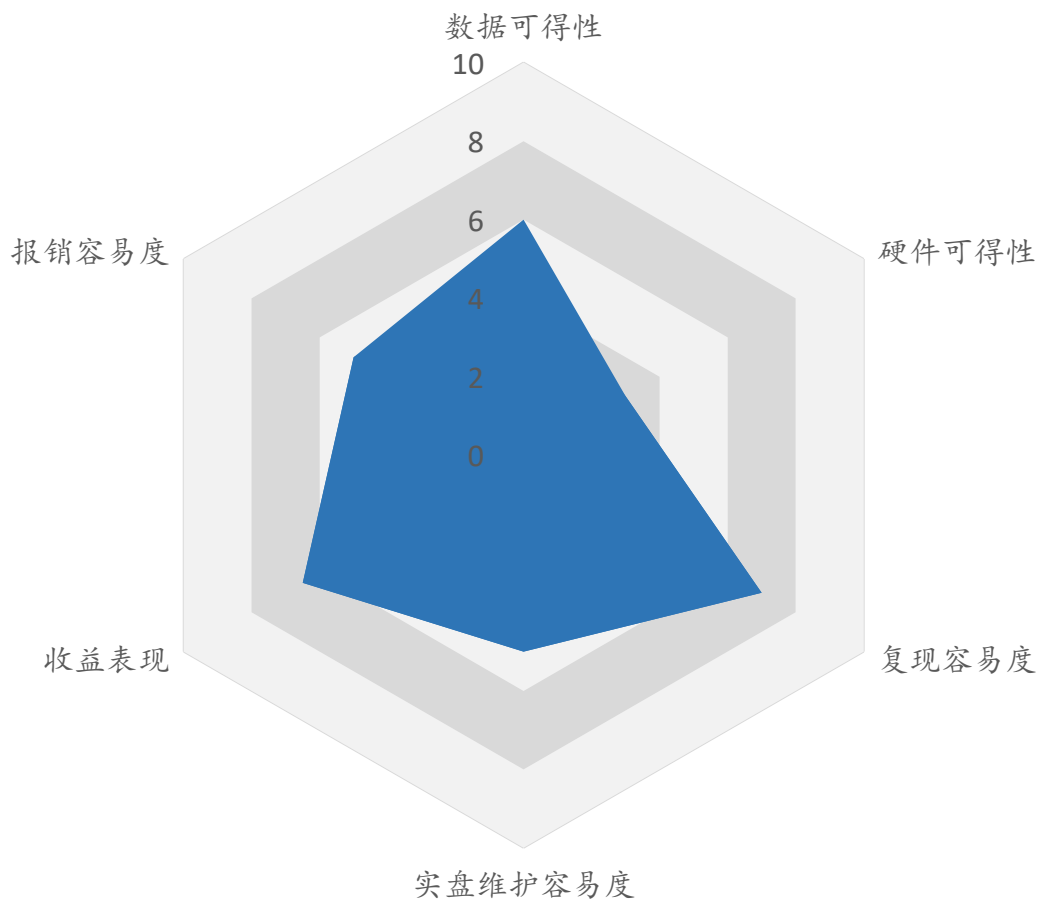
(1) 将分钟数据频率由30分钟提升到10分钟会造成序列长度过长导致模型遗忘前面的信息，反映在最后组合上是换手率增加，因为模型只关注近期的序列信息。

(2) 我们将只取最后一层隐含状态变成取所有时候隐含状态并进行加权，将加权后的特征输入到全连接神经网络中。

五、海通金工深度学习研报解析

海通高频GRU_V2

◆ 《20220531-海通证券-选股因子系列研究（七十九）：用注意力机制优化深度学习高频因子》雷达图



(1) 数据可得性 5分 需要逐笔成交等高频数据来做分钟特征扩充。

(2) 硬件可得性 3分 相比于GRU_V1，更高频的分钟特征也意味着更多的硬件性能消耗。

(3) 复现容易度 7分 网络实现并不复杂，主要是分钟特征的制作，当然分钟开高收低等原始特征稍加处理也能直接进模型

(4) 收益表现 6.5分 相比于30分钟频率的模型，10分钟频率反而会导致一定的收益表现下降，但是和30分钟的相关性降低也能带来新信息。

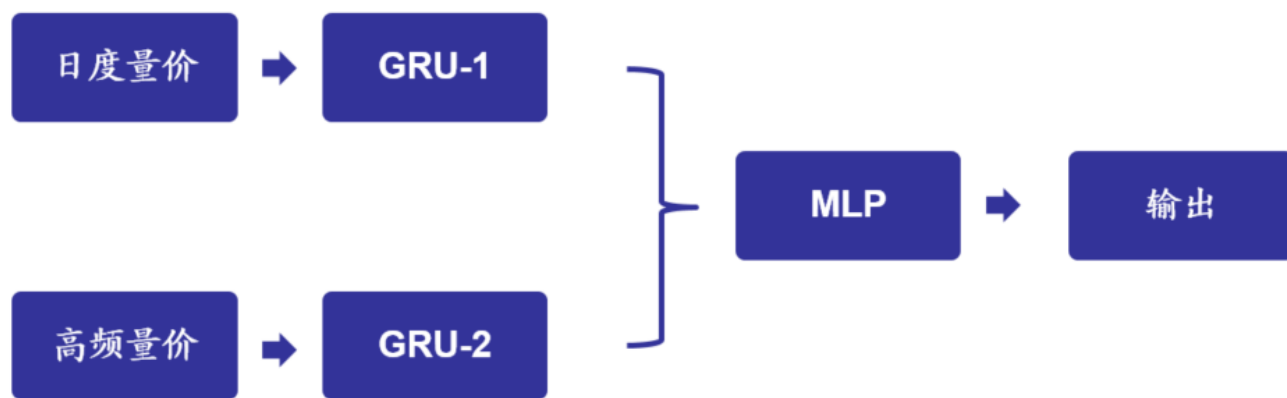
(5) 实盘维护容易度 5分 盘后需要维护逐笔数据以及通过逐笔数据扩充的上百个分钟特征，流程较为繁琐。

(6) 报销容易度 5分 逐笔数据和硬件开销均较大，通联逐笔数据盘后10w+一年，服务器采购费用6w+。

五、海通金工深度学习研报解析

海通混频GRU_V1

- ◆ 《20230513-海通证券-选股因子系列研究（八十七）：高频与日度量价数据混合的深度学习因子》模型



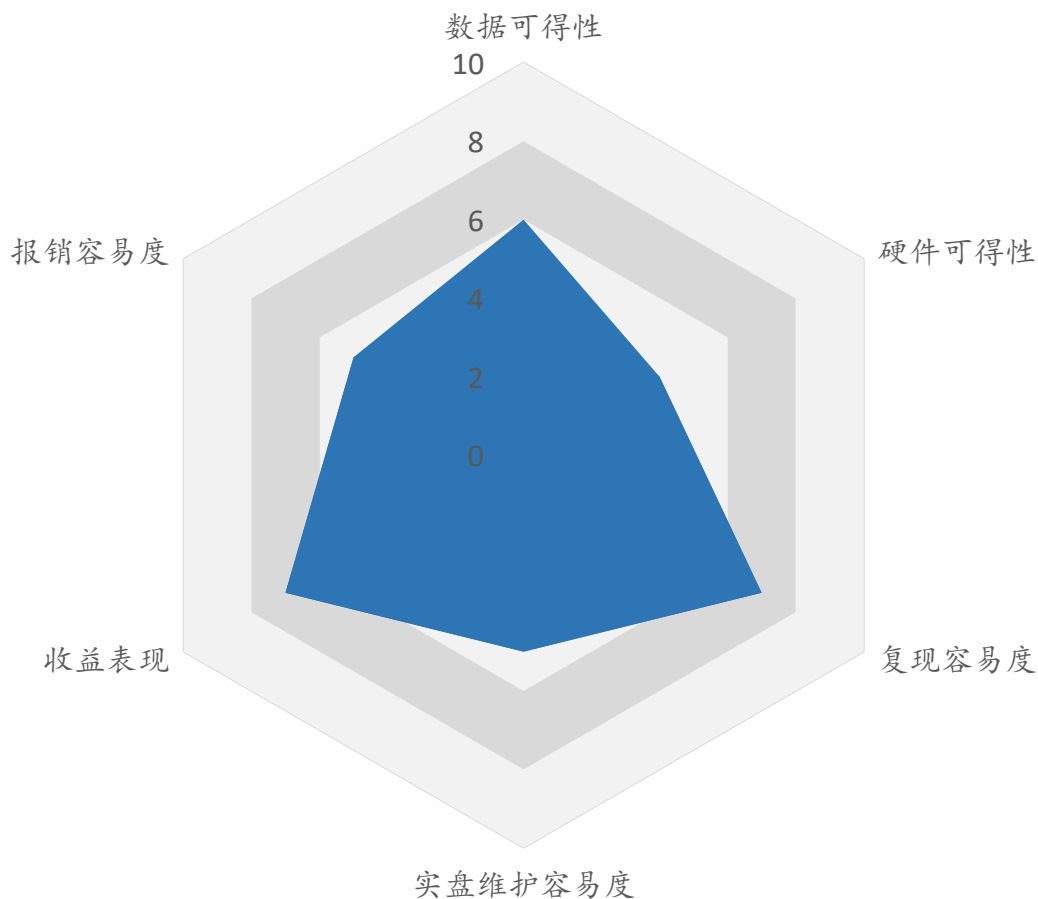
(1) 将过去 60 个交易日的日频量价与 60 分钟频高频量价序列分别输入到两个GRU模型中，然后取两个GRU最后一个隐含状态堆叠进入到全连接神经网络当中。

(2) 混频数据在训练过程当中往往会让预测更偏向于分钟数据，同时日度信息也已经蕴含在分钟的信息当中，因此这种方法可能并不能特别均衡的学习两种数据。

五、海通金工深度学习研报解析

海通混频GRU_V1

◆ 《20230513-海通证券-选股因子系列研究（八十七）：高频与日度量价数据混合的深度学习因子》雷达图



(1) 数据可得性 5分 需要逐笔成交等高频数据来做分钟特征扩充。

(2) 硬件可得性 4分 相比于GRU_V1，对硬件资源消耗类似。

(3) 复现容易度 7分 网络实现并不复杂，主要是分钟特征的制作，当然分钟开高收低等原始特征稍加处理也能直接进模型

(4) 收益表现 7分 相比于30分钟频率的模型，混频模型并不能很好的提升收益表现。

(5) 实盘维护容易度 5分 盘后需要维护逐笔数据以及通过逐笔数据扩充的上百个分钟特征，流程较为繁琐。

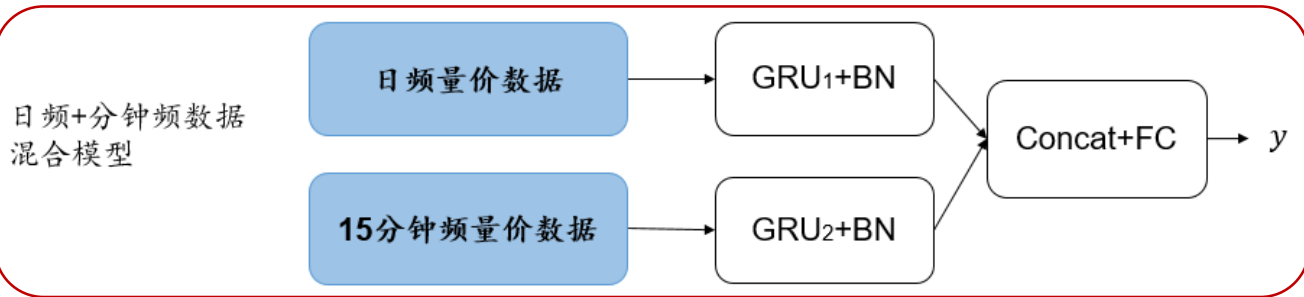
(6) 报销容易度 5分 逐笔数据和硬件开销均较大，通联逐笔数据盘后10w+一年，服务器采购费用6w+。

五、海通金工深度学习研报解析

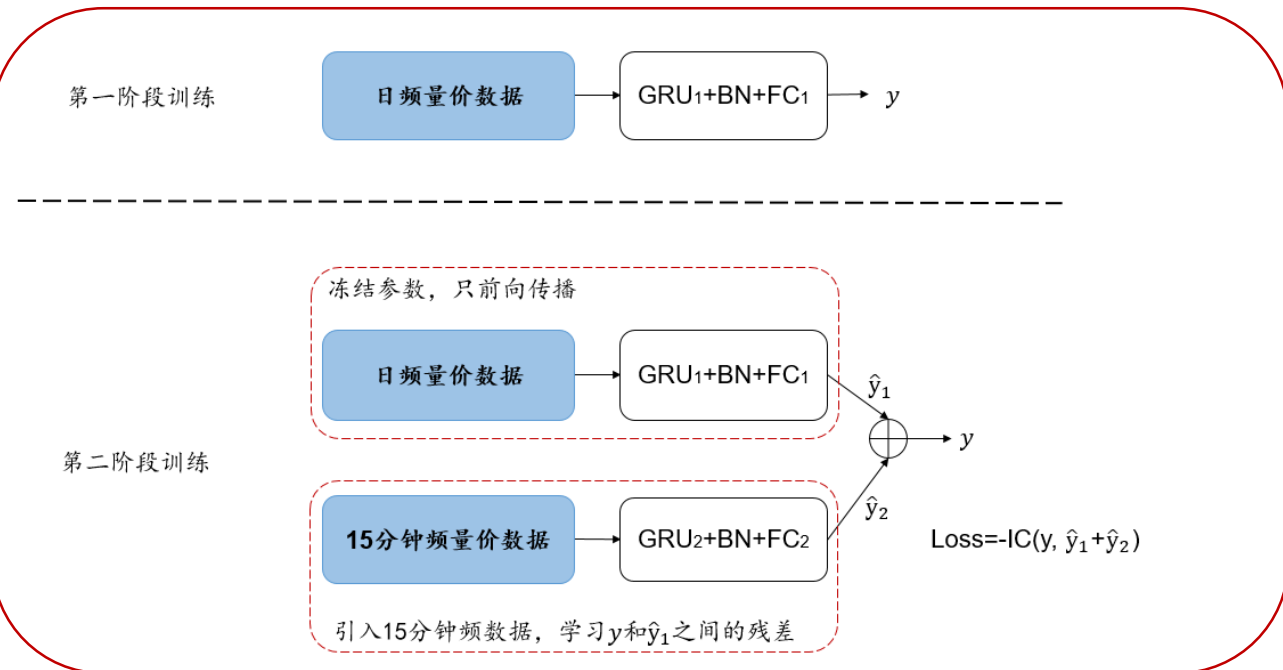
相关报告对比：华泰混频GRU_V1

◆ 《20230511-华泰证券-金工：神经网络多频率因子挖掘模型》模型

模型



训练方式



相比海通的混频模型，华泰的混频模型有以下两点不同

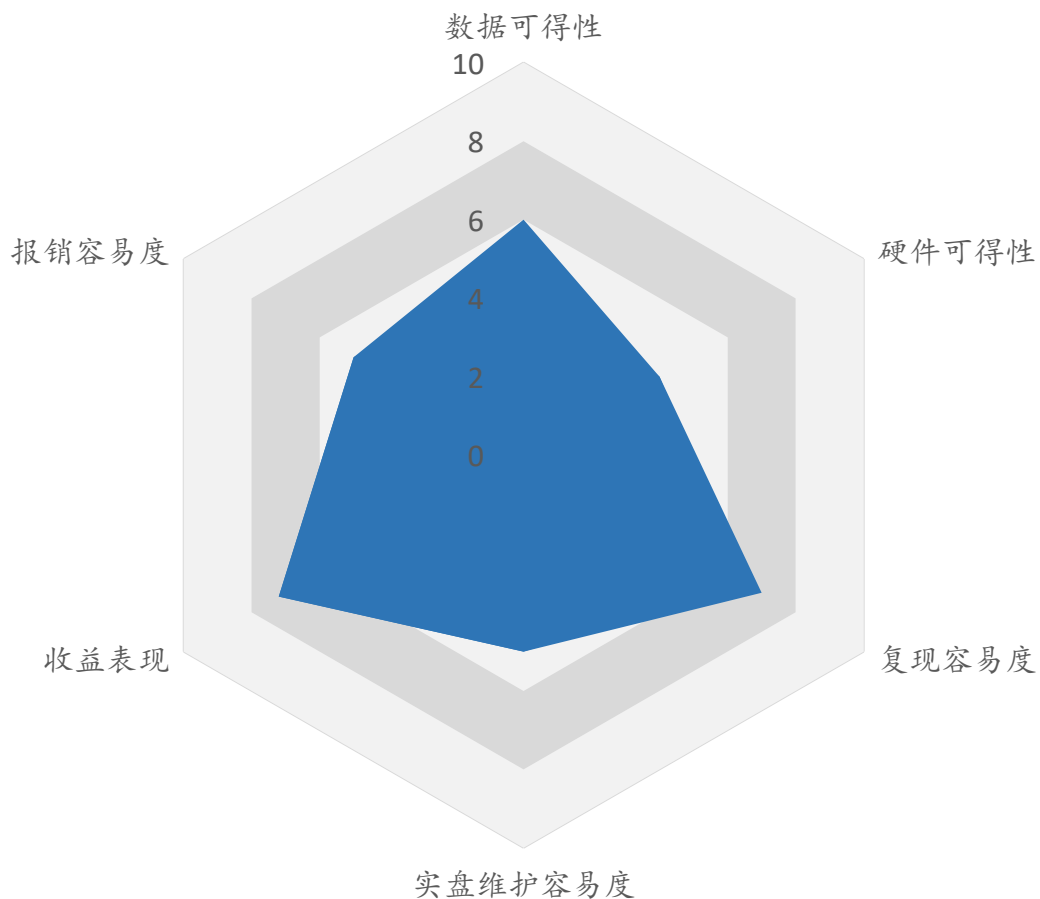
(1) 数据上回看过去40个交易日的日频数据和15分钟频的数据。

(2) 训练方式上，海通混频模型是将日频和分钟频率的GRU同时进行训练。而华泰是先训练日频的GRU模型然后冻结日频模型的参数，引入15分钟的GRU模型训练日频模型的残差收益率，最后将日频和分钟模型进行叠加。

五、海通金工深度学习研报解析

相关报告对比：华泰混频GRU_V1

◆ 《20230511-华泰证券-金工：神经网络多频率因子挖掘模型》雷达图



(1) 数据可得性 5分 需要逐笔成交等高频数据来做分钟特征扩充。

(2) 硬件可得性 4分 相比于GRU_V1，对硬件资源消耗类似。

(3) 复现容易度 6.5分 相比于GRU_V1训练代码稍微复杂

(4) 收益表现 7.2分 相比于30分钟频率的模型，新的混频模型训练方式对收益有一定提升。

(5) 实盘维护容易度 5分 盘后需要维护逐笔数据以及通过逐笔数据扩充的上百个分钟特征，流程较为繁琐。

(6) 报销容易度 5分 逐笔数据和硬件开销均较大，通联逐笔数据盘后10w+一年，服务器采购费用6w+。

五、海通金工深度学习研报解析

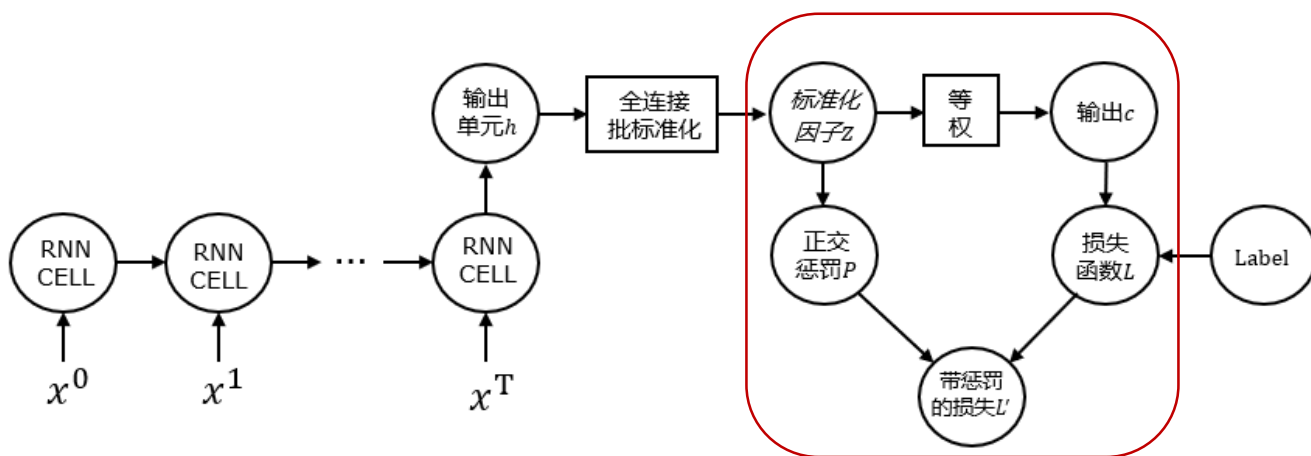
海通报告总结

- ◆ 海通高频GRU_V1利用利用L2数据构建164个30分钟序列特征直接输入GRU模型当中提取信息，V2则是将数据频率提升到10分钟，但是由于序列过长导致GRU模型对于早起信息遗忘，因此加入attention结构来更好记忆早期信息。
- ◆ 海通混频GRU_V1将分钟数据和日频数据输入到两个GRU模型中一起训练，然后将训练的结构叠加，该方法不能很好平衡日频和分钟数据的信息权重；华泰混频GRU_V1则是先训练日频GRU，然后再冻结参数后，以日频的残差收益率训练分钟GRU，两者再进行叠加，对于混频数据信息权重的平衡有所改善。

六、东方金工深度学习研报解析

东方日频GRU_V1

◆ 《20210311-东方证券-因子选股系列之七十四：神经网络日频alpha模型初步实践》模型



与前述模型的主要差异：

华泰、海通：最后输出的是一个合成因子，模型的训练目标既各个子因子如何挖掘、也包含因子之间如何合成，才能使最后合成因子的效果最好；

东方：最后输出的是多个单因子，模型的训练目标为，在等权合成的情况下、如何构建这些单因子能让得到的合成因子效果最好，同时也要保证这些单因子之间的相关性足够低。

(1) 输入数据为日频原始数据

(2) 模型上依然采用LSTM或者GRU这类序列模型。

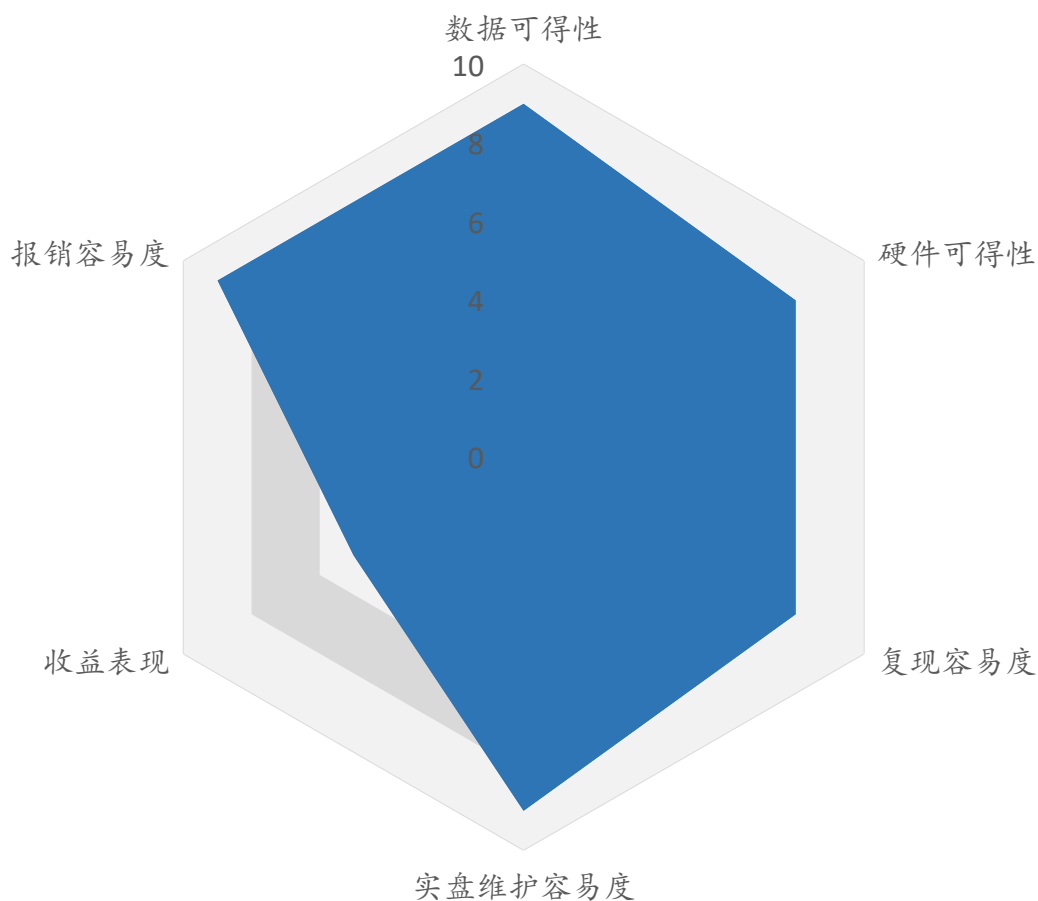
(3) 训练模式上，GRU的输出不在经过全连接层，而是直接等权组合进入损失函数（mse或者IC），并且对损失函数叠加特征相关性的惩罚，以此保证GRU的输出特征之间相关性较低。

(4) 将输出的低相关性的特征和原来因子库的特征相结合一起进行因子组合。

六、东方金工深度学习研报解析

东方日频GRU_V1

◆ 《20210311-东方证券-因子选股系列之七十四：神经网络日频alpha模型初步实践》雷达图



(1) 数据可得性 9分 日频开高收低量等数据获取非常容易。

(2) 硬件可得性 8分 模型不复杂，基本上正常的笔记本电脑都能跑起来。

(3) 复现容易度 8分 相比模型和数据构建均不复杂，除了需要将生成的因子进行二次组合相比于一般模型复杂。

(4) 收益表现 5分 由于数据和模型均不复杂，因此表现也只能算一般

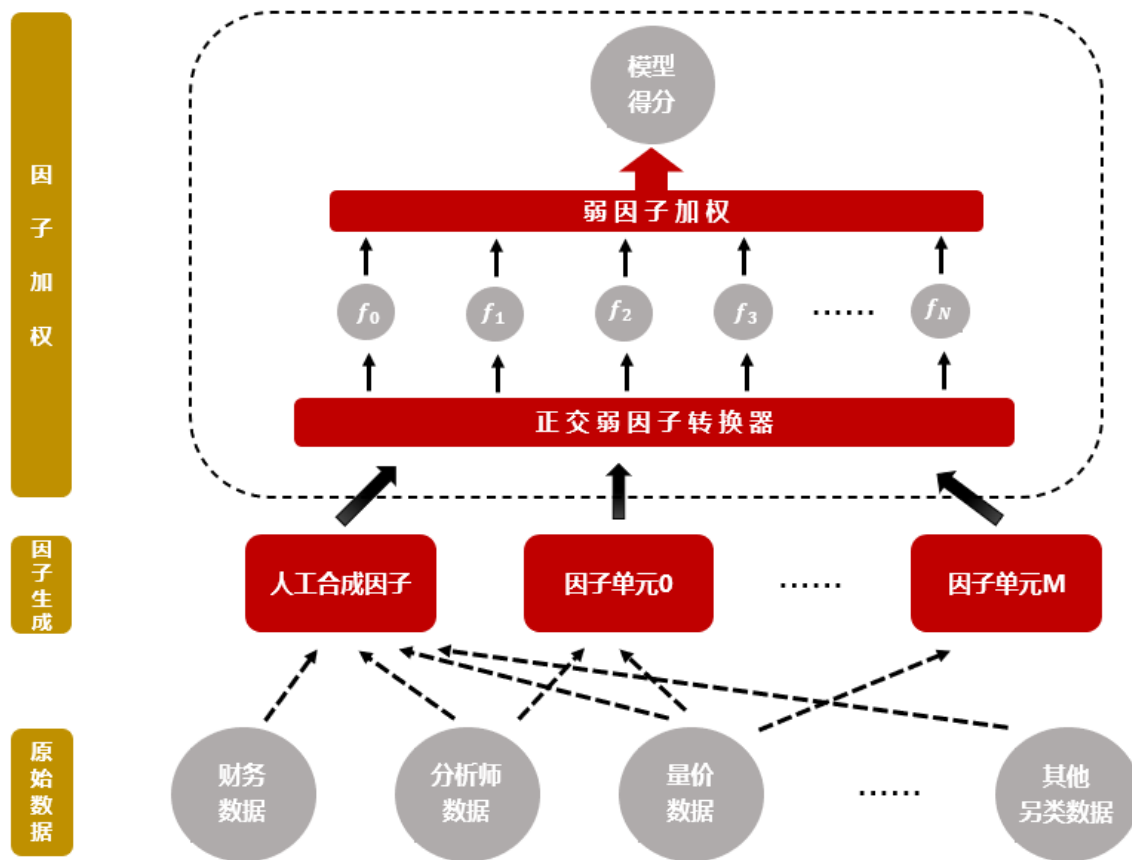
(5) 实盘维护容易度 9分 数据盘后增量更新容易，每季度训练新模型开销不大。

(6) 报销容易度9分 基本上不存在额外的支出，当前公募拥有的资源就能跑起来。

六、东方金工深度学习研报解析

东方混频多模型_V1

◆ 《20220328-东方证券-因子选股系列之八十一：周频量价指增模型》模型



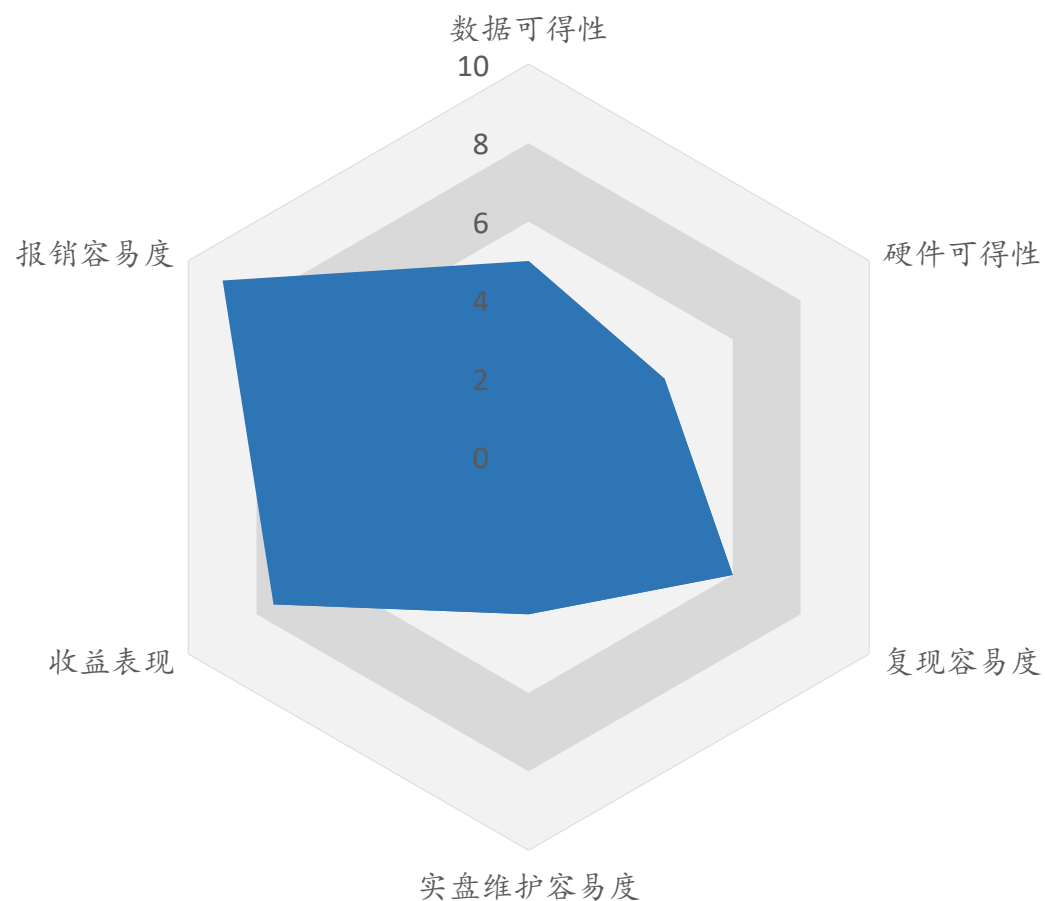
(1) 在之前模型的基础上更改输入数据频率和模型类型从而构建多个因子但愿

(2) 将多个因子单元产生的因子进行叠加一起进入因子组合模型进行训练。

六、东方金工深度学习研报解析

东方混频多模型_V1

◆ 《20220328-东方证券-因子选股系列之八十一：周频量价指增模型》雷达图



(1) 数据可得性 5分 需要逐笔成交等高频数据来做分钟特征扩充。

(2) 硬件可得性 4分 相比于GRU_V1，对硬件资源消耗类似。

(3) 复现容易度 6分 相比于GRU_V1训练代码稍微复杂，不光需要构建分钟特征，同时还要训练多个模型叠加

(4) 收益表现 7.5分 相比于30分钟频率的模型，由于是通过多模型叠加而成，表现也必然有一定提升。

(5) 实盘维护容易度 4分 盘后需要维护逐笔数据以及通过逐笔数据扩充的上百个分钟特征，再次基础上还得维护多个模型的训练。

(6) 报销容易度 5分 逐笔数据和硬件开销均较大，通联逐笔数据盘后10w+一年，服务器采购费用6w+。

六、东方金工深度学习研报解析

东方报告总结

- ◆ 东方日频GRU_V1将日频数据序列输入到GRU或者LSTM这列序列模型中提取信息，得到30个提取出来的特征，这些特征不在合成而是直接作为因子输出，同时在训练过程中对30个特征之间的相关性进行惩罚，因此这30个特征之间也是低相关的。
- ◆ 东方混频多模型V1在上面日频GRU_V1的基础上，通过修改输入数据类型和模型结构来构建多个因子挖掘单元，最后将多个模型出来的因子进行组合得到最终的合成因子。

七、总结

模型比较

模型名称	数据类型	网络结构	数据可得性	硬件可得性	复现容易度	收益表现	实盘维护容易度	报销容易度
华泰alphanet_v1	日频	alphanet	9	9	5	4	9	9
华泰alphanet_v2	日频	alphanet+LSTM	9	8	5	5	9	9
华泰alphanet_v3	日频	alphanet+GRU	9	8	5	5	9	9
分析师研报Bert	分析师研报文本	Bert+XGBoost	6	6	5	5	8	6
华泰混频GRU_V1	日频+分钟	GRU+冻结参数	5	4	6.5	7.2	5	5
海通高频GRU_V1	分钟	GRU	5	4	7	7	5	5
海通高频GRU_V2	分钟	GRU+Attention	5	3	7	6.5	5	5
海通混频GRU_V1	日频+分钟	GRU	5	4	7	7	5	5
东方日频GRU_V1	日频	GRU、LSTM等	9	8	8	5	9	9
东方混频多模型_V1	日频+分钟	GRU、LSTM、Transformer	5	4	6	7.5	4	5

七、总结

开展深度学习因子挖掘所需资源

(1) 高性能服务器:

CPU 64核为佳 (AMD 第三代epyc 7T83 7763等)

内存 512GB或者1TB

存储 至少10TB往上的固态硬盘存储空间, 30TB为佳

显卡 至少一张4090 有大模型需要可采购A800等专业显卡

硬件条件

(2) 数据:

每日盘后的分钟数据、3秒快照数据、逐笔成交和逐笔委托
(通联、wind、金数源)

分析师研报摘要等文本数据 (朝阳永续)

软件条件