



挑战组

赛题一：基金产品的长周期申购和赎回预测

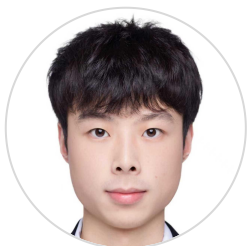
基金申赎智能预测系统 基于多源特征融合与注意力机制增强LSTM

队名：晨曦队

队员：黄浚珂、王俊松、沈琦明

机构：南京大学、云南财经大学

团队成员



黄浚珂
南京大学

研究方向：金融工程、人工智能
负责工作：方案设计、模型设计与训练、特征工程开发



王俊松
云南财大

研究方向：计算机视觉、垂直大模型
负责工作：方案设计、特征工程开发、数据采集与预处理



沈琦明
云南财大

研究方向：区块链
负责工作：数据采集与预处理、流程自动化与部署



徐兴贵 | 指导老师
云南财大

信息学院副院长、副教授
“兴滇计划”青年人才



第十五届“工行杯”全国大学生金融科技创新大赛江苏省一等奖

美国大学生数据建模竞赛F奖

第二十届“花旗杯”金融创新应用大赛全国冠军

第二十七届中国机器人及人工智能大赛云南省一等奖

C O N T E N T

目 录

1 方案综述

2 核心模型

3 特征构建

4 训练优化

5 总结

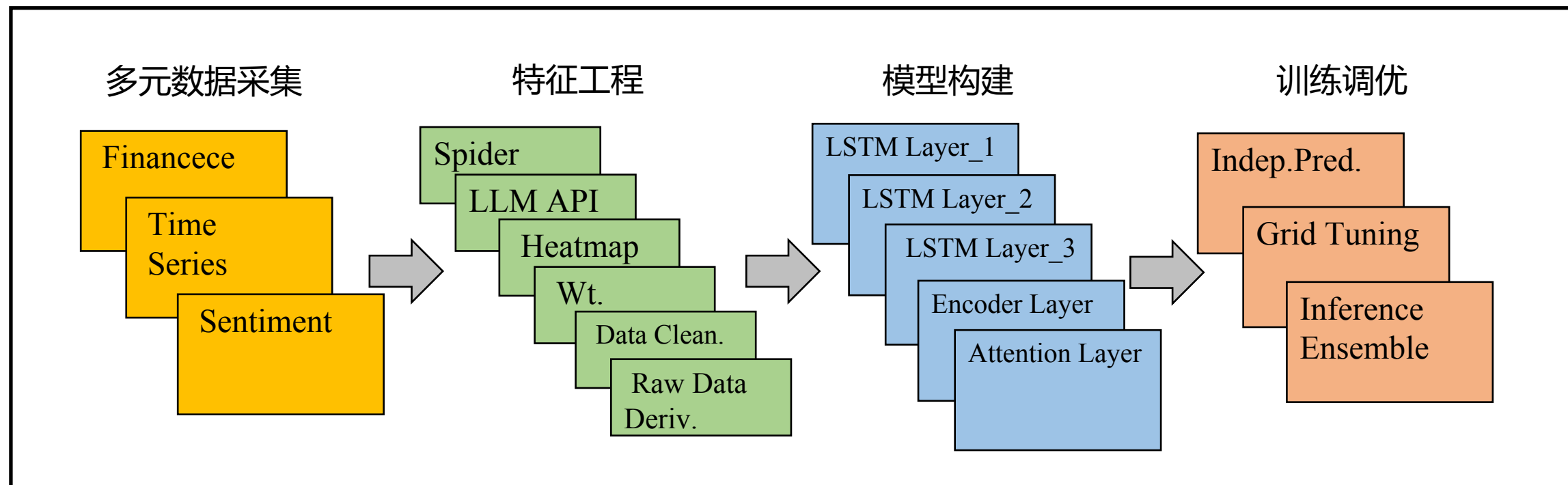


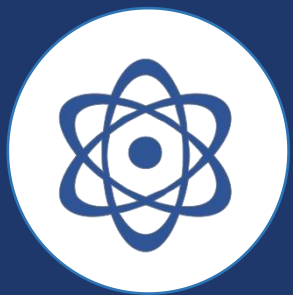
01 方案综述

总体方案



以 20 支基金历史数据为基础，从**多源数据采集**、**特征工程**、**模型构建**、**训练调优**四大环节出发，通过多维度特征挖掘、高效特征验证、深度学习建模及集成调优，构建了一套**精准、通用、解释性强**的基金申赎行为预测框架：





02

核心模型

模型选择



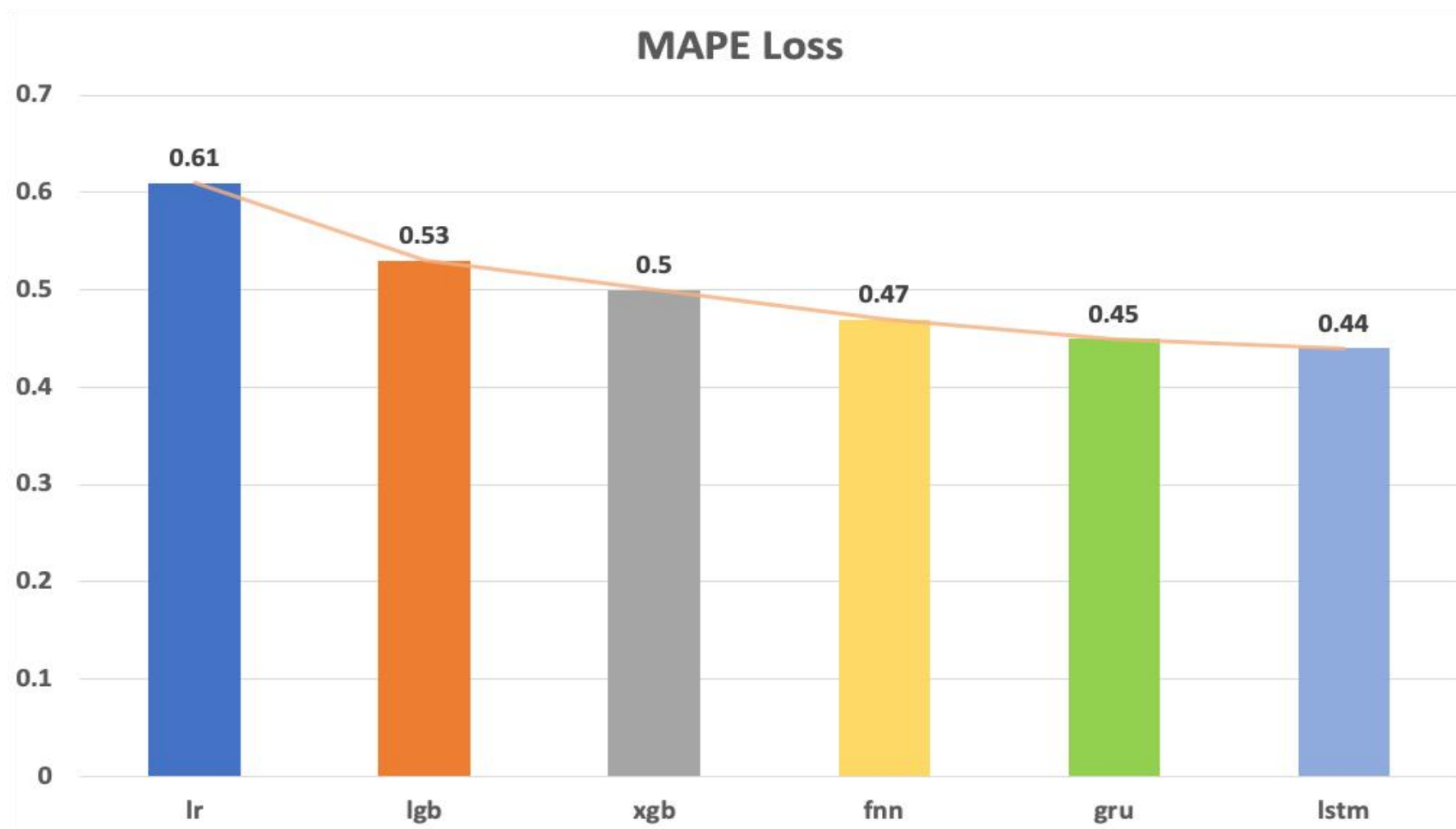
多模型 VS LSTM

模型类型	局限性	LSTM优势
树模型 (如随机森林)	难建模时间顺序依赖, 对长期趋势敏感度低	显式学习序列动态, 捕捉非线性时序规律
基础时序模型 (如 ARIMA)	依赖线性假设与平稳性, 需复杂特征工程	直接处理非平稳数据, 自动学习复杂模式 (如行情突变)
传统神经网络 (如 RNN/MLP)	忽略时间维度, 仅适合静态特征分析	时序感知结构, 动态更新记忆状态
朴素贝叶斯	假设特征独立, 无法捕捉时序依赖关系	建模序列间长短期依赖, 保留时序特征关联性
支持向量机 (SVM)	处理高维时序数据时计算复杂, 泛化能力受限	高效处理高维时序输入, 对噪声数据具有更强鲁棒性

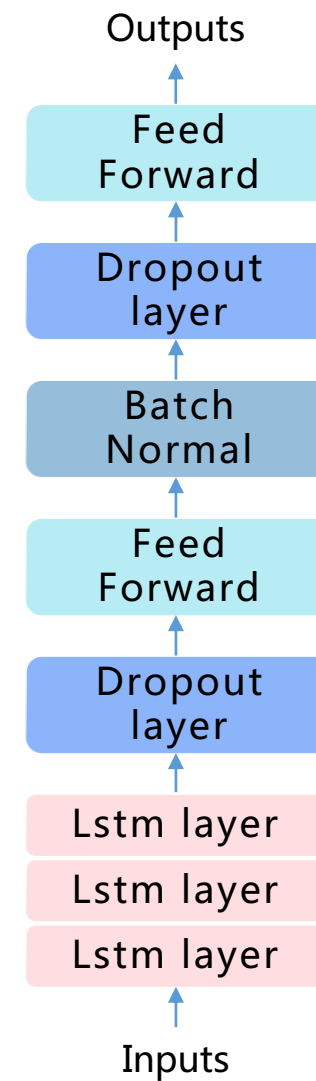
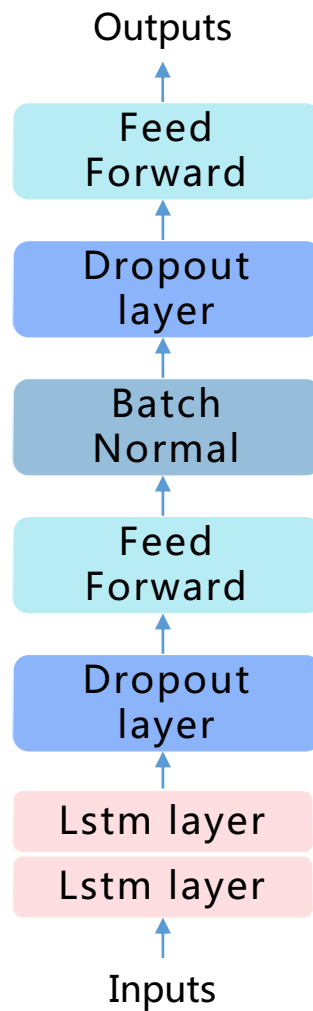
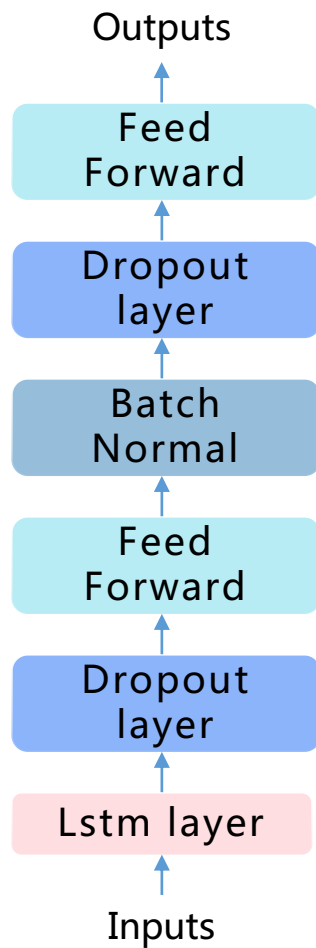
效果对比



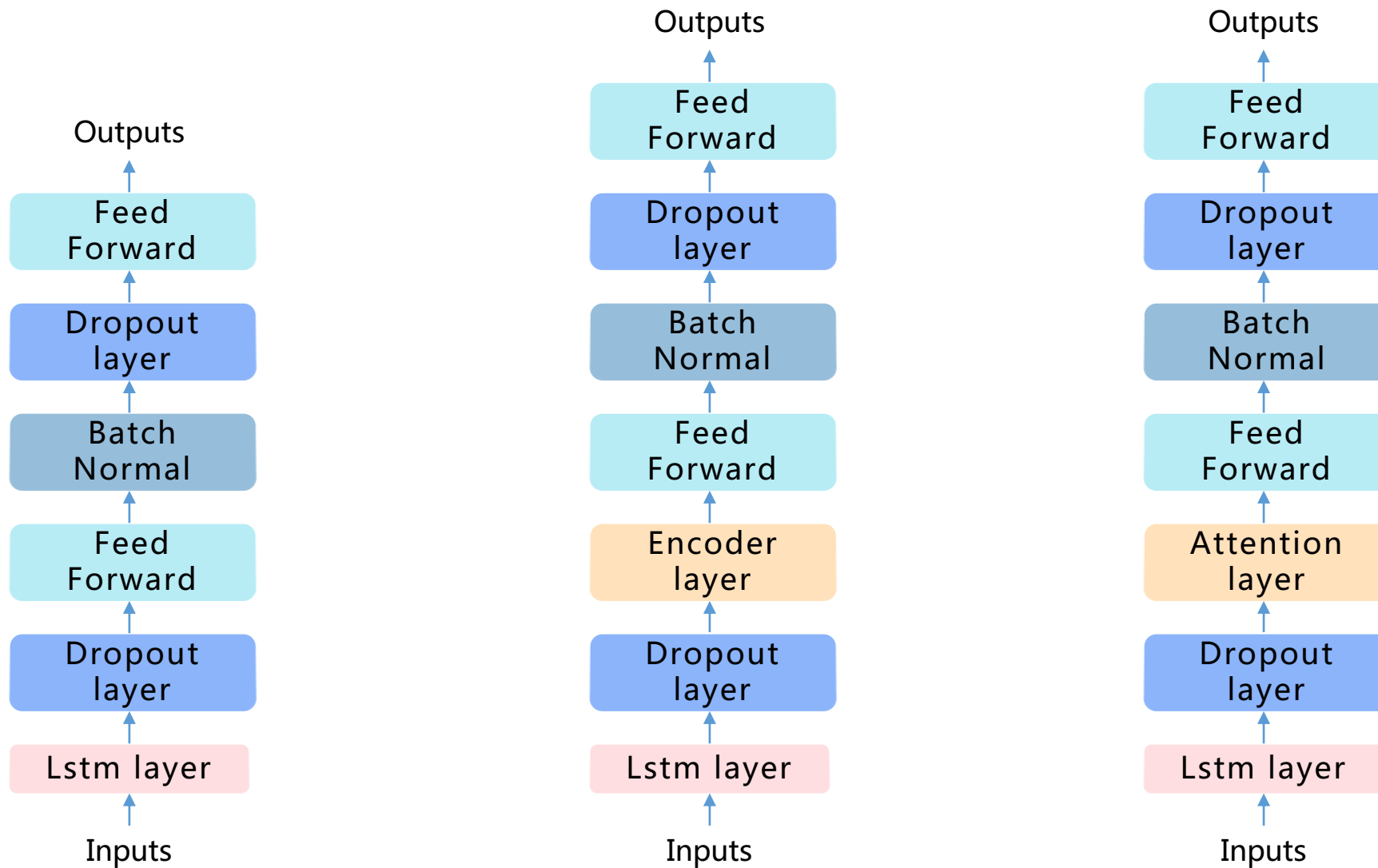
对比不同模型，LSTM 在基金申赎预测任务中展现出**更优的时序建模能力**



优化方向

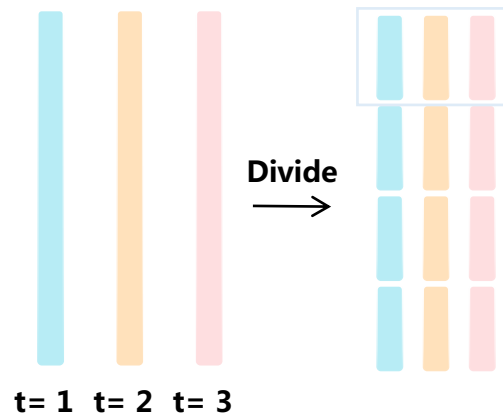
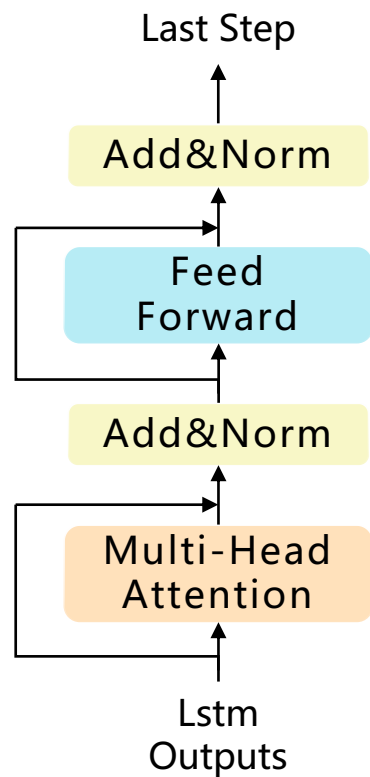
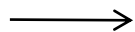
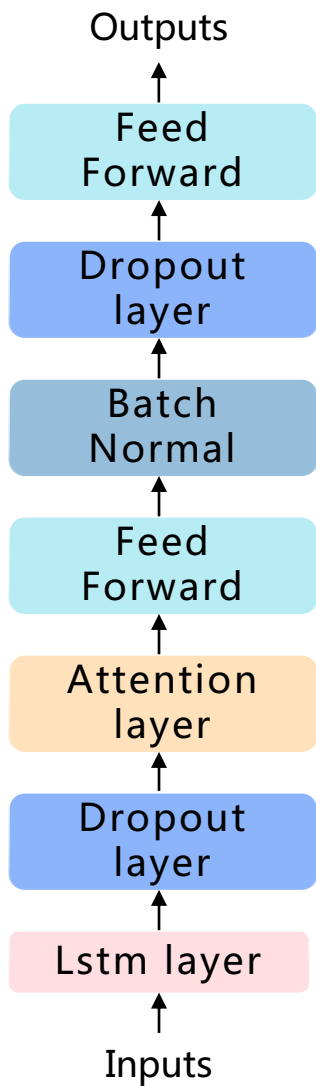


优化方向





模型架构

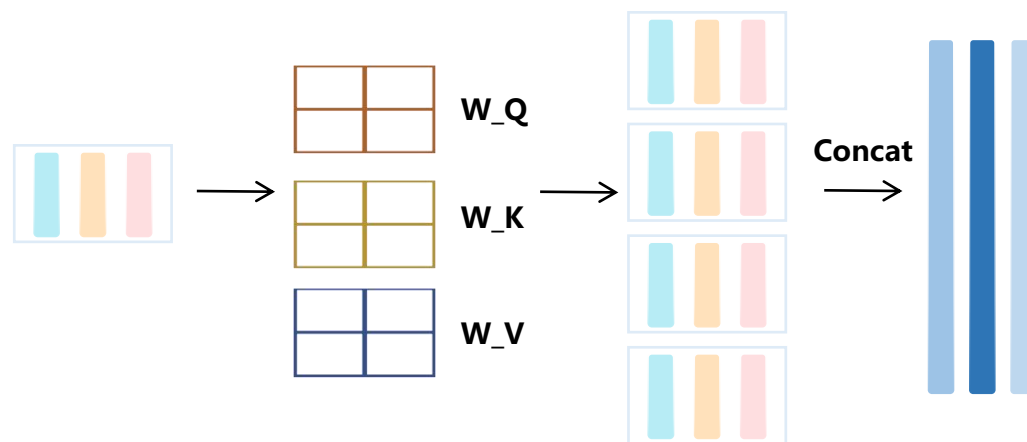


head1 $d \in (0,64)$

head2 $d \in (64,128)$

head3 $d \in (128,192)$

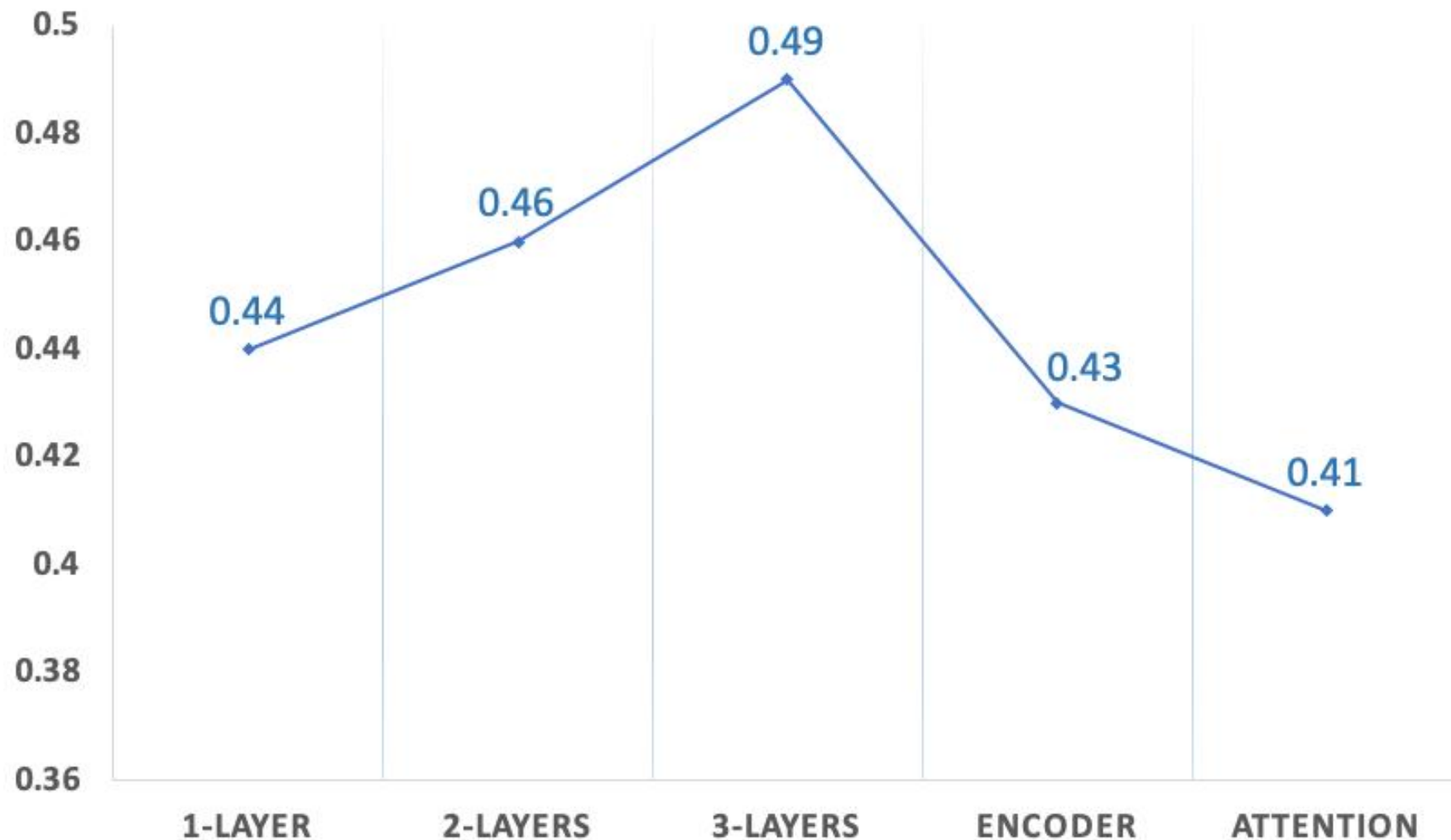
head4 $d \in (192,1256)$



模型演变结果



MAPE LOSS



- 过度堆叠基础层可能引发**过拟合问题**
- **Encoder** 对特征的规整、**Attention** 对关键时序信息的聚焦有效**提升预测精度**



03

特征构建

特征总结



申购相关特征	赎回相关特征
<ul style="list-style-type: none"> ● 申购金额 ● 赎回金额 ● UV1曝光 ● UV3曝光 	<ul style="list-style-type: none"> ● 申购金额 ● 赎回金额 ● UV1曝光 ● UV3曝光
<ul style="list-style-type: none"> ● 星期 ● 节假日 ● 发薪日 	<ul style="list-style-type: none"> ● 星期 ● 节假日
<ul style="list-style-type: none"> ● 市场热度 ● 百度指数 	

我们根据申购赎回两种场景**分别构建**了不同的特征工程。

UV1/3曝光、申购/赎回金额、星期、节假日、发薪日、市场热度、百度指数是我们获取剔除的原始数据，对部分时段无波动的行为特征以及金融特征的有机结合。通过均值占比计算量化周期规律，对比传统编码更贴合实际交易节奏；节假日特征通过大模型精准识别调休，确保交易时限判断无偏差。

投资者情绪指标



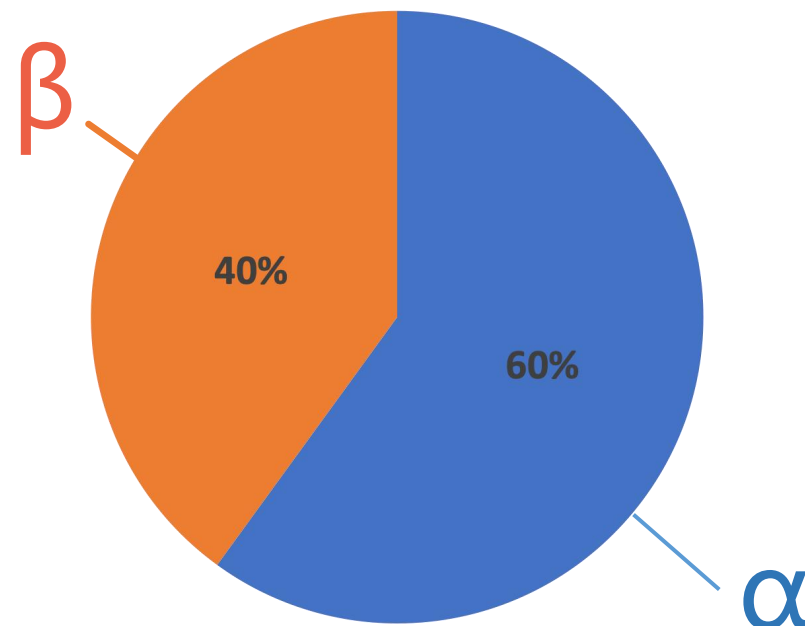
我们通过爬取天天基金网的对应**基金评论**，用大模型对每条评论的**情绪进行打分**量化并结合阅读量、评论量等参数的**加权**得每日该基金的情绪分，以此**量化情绪指标**对申赎的影响并纳入建模。

单帖权重公式

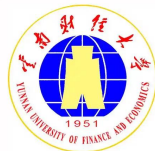
$$w_i = \alpha \cdot \log(1 + \text{Views}_i) + \beta \cdot \left(\frac{\text{Comments}_i}{\text{Views}_i} \right)$$

阅读量贡献

互动率贡献



最优参数设置



投资者情绪指标

我们选取东方财富网的对应基金评论，用大模型对每条评论的情绪进行打分量化并

结合阅读	评论ID	评论内容	用户名	正向	负向	情绪得分
0.1 “\$富国中证 指数(LOF)/ 工太垃圾恶 超级超级”	1338	今天终于不是股债双杀类型了权益市场表现不佳，但是债	真真理财记	58	0	0.3
	1338	，跌的比别人多，涨的比别人少	Oscarlee	39	0	0.3
	1338	医药板块异动！创新药在政策支持下表现强劲，同泰大健	唐佑超	87	0	0.7
	1338	涨涨跌跌是常态，只是希望看到涨多一点呀	昵称一百斤	60	0	0.7
	1338	拉垮的半年啊，选股不行，连持有的债券都选不准了	gdhdjm8888	61	0	0.1
	1338	这周也该好好表现一下啊，努努力冲吧	萤火之辉	52	0	0.7
	1338	最近太拉夸了！偶尔微跌，就是不长！	时尚的游淼洋	197	3	0.3
	1338	昨晚差点飘红，抗跌不错啊。今天继续努力来个翻红，	萤火之辉	44	0	0.7
	1338	放眼全球，这是是要拉起来的节奏吧期待中。	知足常乐就美	43	0	0.7
	1338	行情来了。连续拉涨，表现可以的，继续冲吧	萤火之辉	66	0	0.9
	1338	告别调整期了，接下来就有的好涨了	深思熟虑的米	46	0	0.9
	1338	稳定跑赢基准的基金投资策略有哪些？	焦易圈	205	2	0.5
	1338	在政策利好的推动下，金融板块迎来戴维斯双击。同泰金	唐佑超	61	0	0.7
	1338	适应市场变化，选好基金就是成功的关键了	昵称一百斤	55	0	0.7
	1338	昨晚大涨行情，表现不错，今天继续加油冲	萤火之辉	77	0	0.7



0.9

来了。连
，表现可
继续冲吧”

市场热度



大模型核心元素

💡 大模型prompt | 核心元素提取图

角色

金融数据
校准专家



核心任务

构造市场热度



调整原则

- 🎯 行业指数主体
- 🔒 情绪得分辅助
- 🔧 基金属性背景
- 📈 前5天均值趋势

不同基金代码

基于基金属性区分基金类型
如货币/权益/债基等

行业指数

市场热度计算逻辑

前5天申赎金额均值

趋势参考

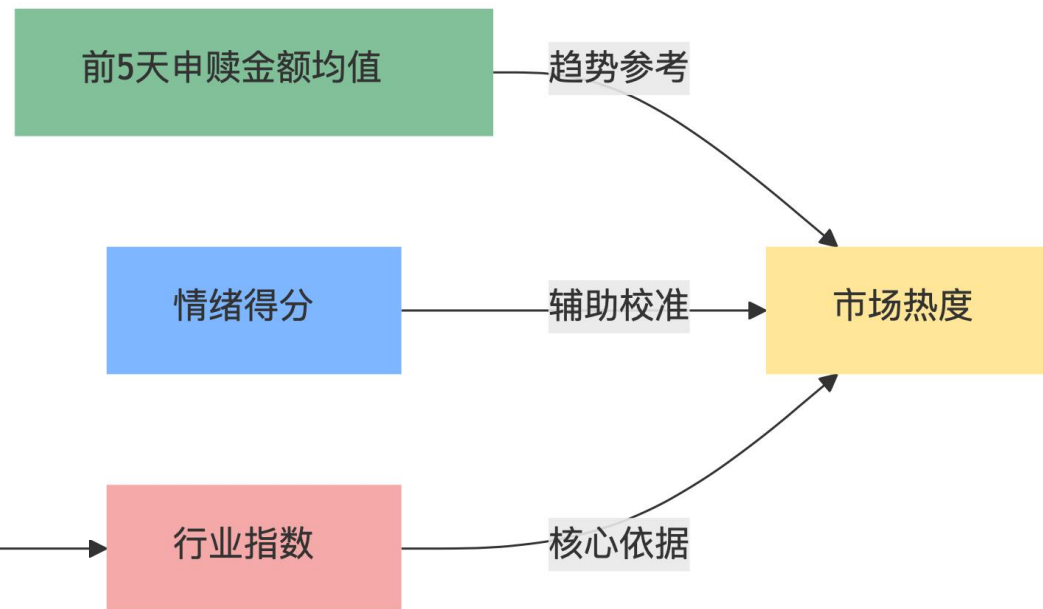
情绪得分

辅助校准

市场热度

核心依据

市场热度

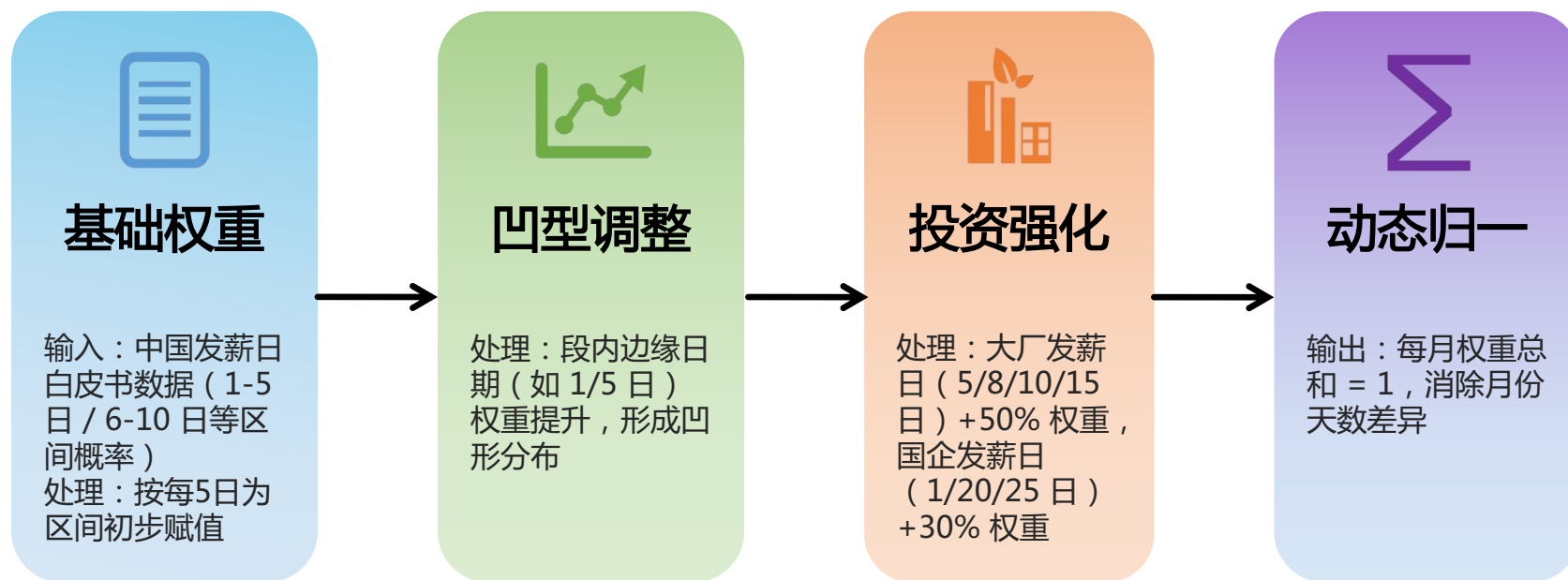


发薪日



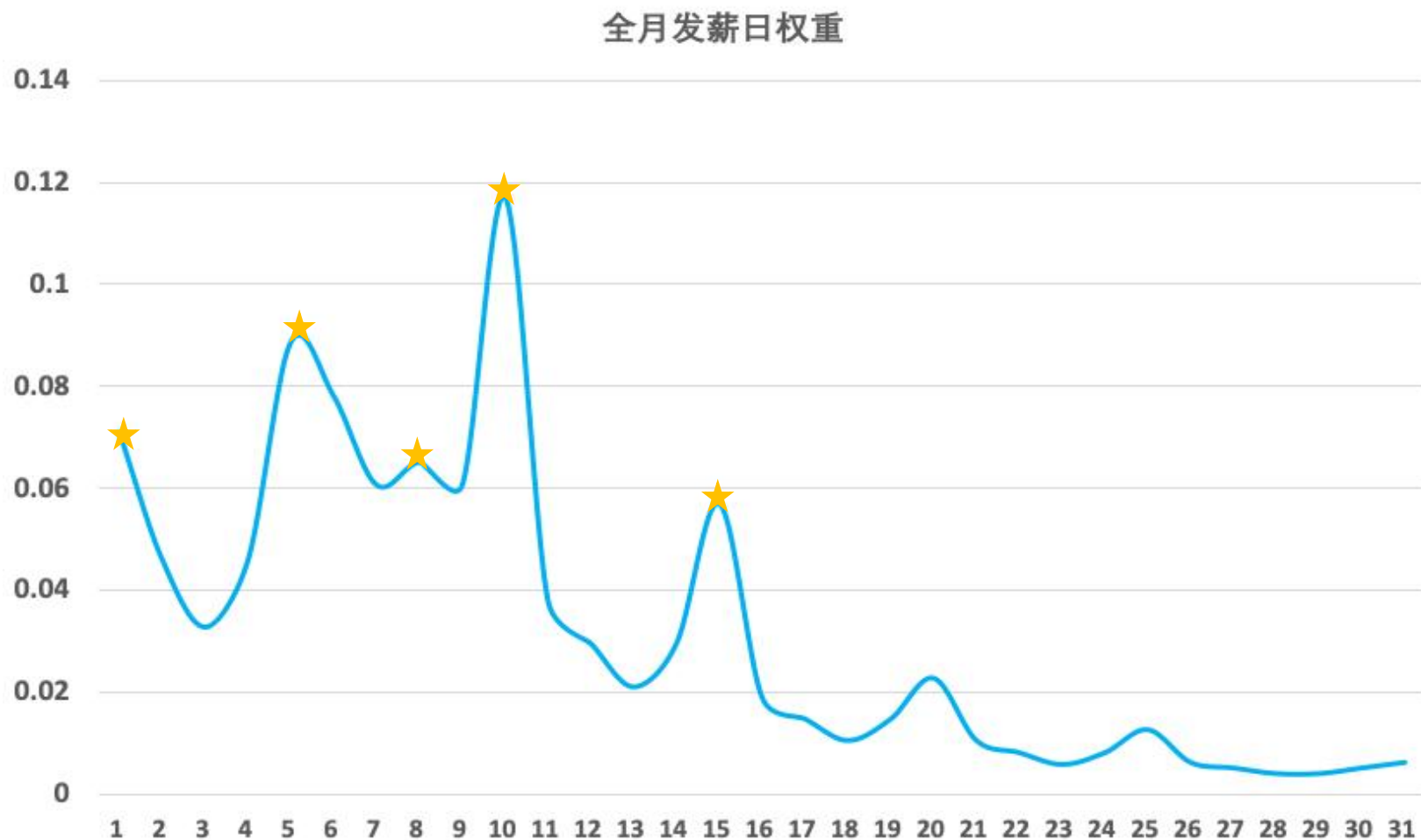
发薪日与投资者**资金流动性**及基金申赎行为有较强的相关性，我们依据中国企业发薪日分布规律，用**多阶段权重分配法**构建月份中各日发薪概率分布的发薪日特征，以此捕捉资金流动性对申赎的作用并纳入建模。

多阶段权重分配法

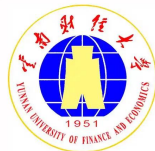


发薪日

全月权重分布图

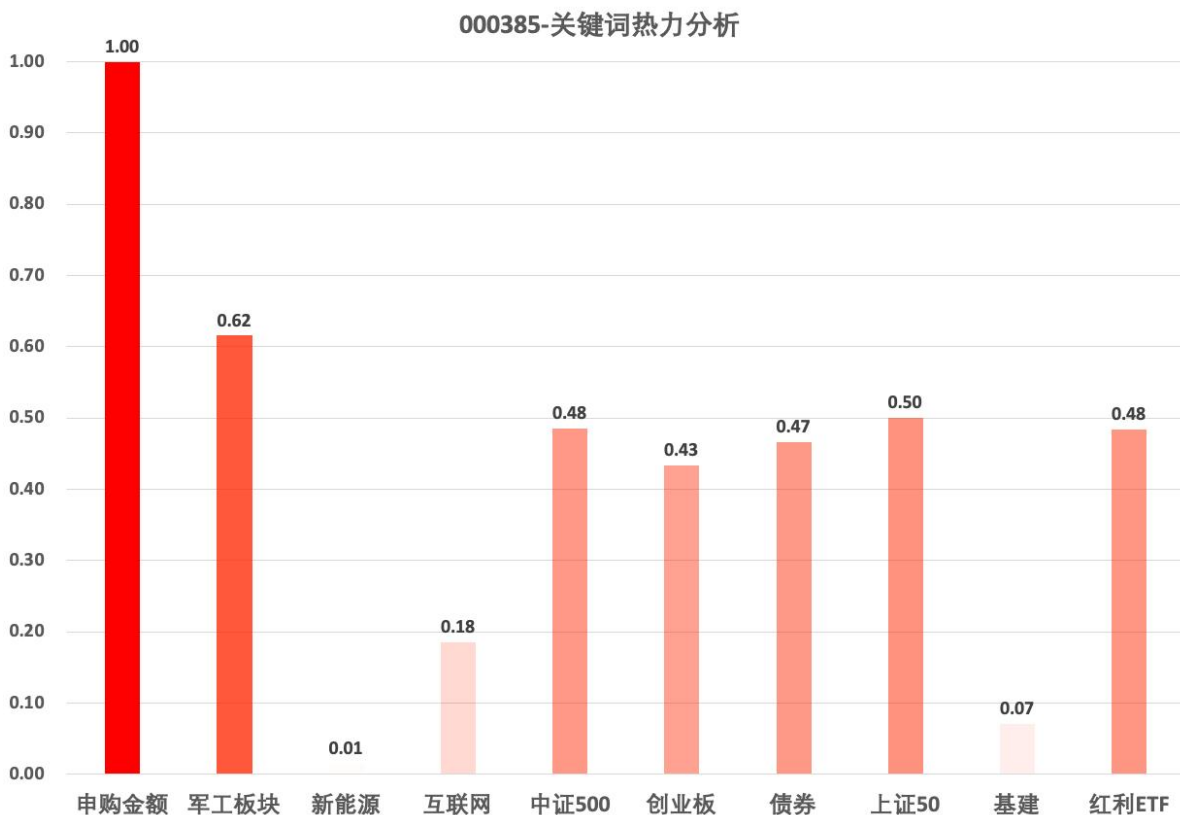


- 发薪日对投资者资金流动性的影响
- 从抽象概念转化为可量化、可解释的日度数据，为模型捕捉申赎行为规律提供直接输入
- 填补了“单纯时序特征无法体现资金流动性差异”的建模空白



百度指数

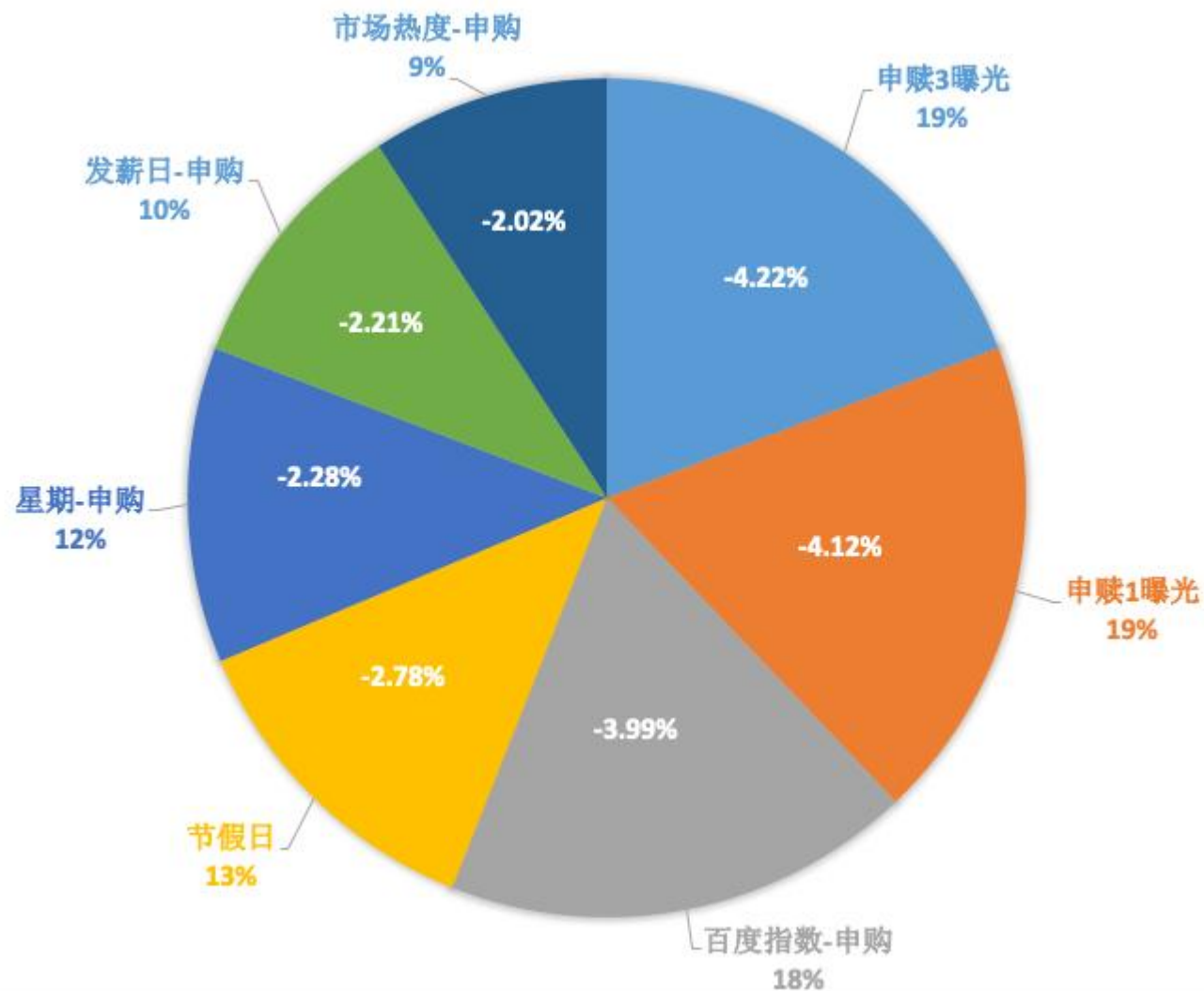
百度指数以网民在百度的搜索量为数据基础，以关键词为统计对象，分析并计算出各个关键词在百度网页搜索中搜索频次的加权。基于基金代码申赎数据做热力图分析，我们爬取并过滤出最适合20支基金的8个百度指数关键词，将该信息作为我们的百度指数特征。



特征效果汇总



申购金额相关特征效果占比及精度提升

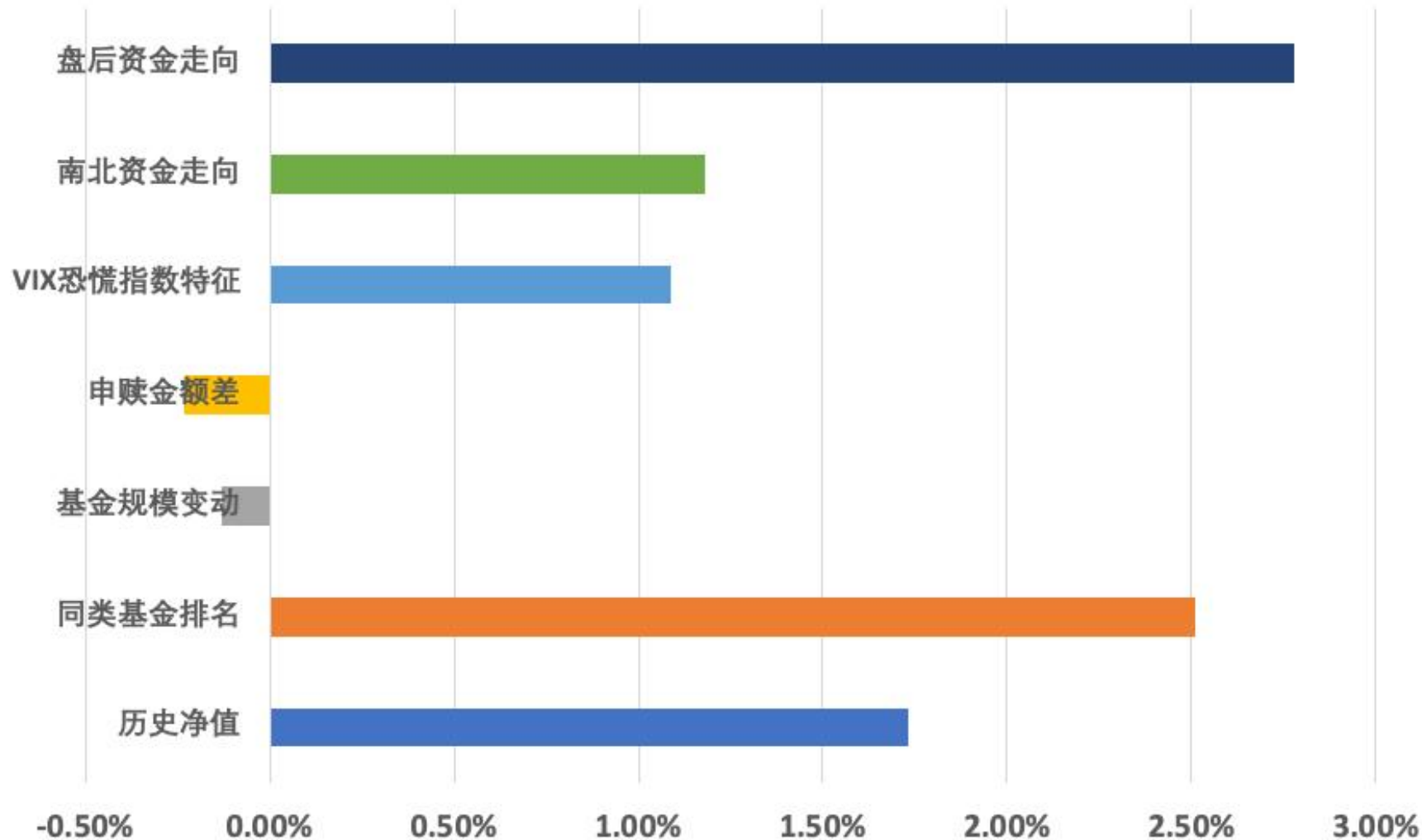


- 赛方提供的曝光 UV 类特征是**核心预测基础**，对申购行为**驱动强**，精度提升较强。
- **时序特征**是影响投资者申购决策的**关键因素**
- 获取与构造的外部特征，发挥了突出的**增益作用**，多类特征**协同**，共同推动模型性能进阶。

其余特征尝试



精度下降百分比



- 图中为淘汰的部分特征
- 虽未覆盖特定基金代码
- 在基金代码迭代、金融场景替换中，存在**适配潜力**



04

训练优化

训练配置以及优化



高效、灵活、稳定的优化方案

开发流程

- 拆分规则：训练集 70%、验证集 30%
- 基金适配：每基金分配 3 验证样本
- 关键参数：滞后期 $\text{lookback_days}=[5-10]$ (网格搜索覆盖)、预测固定 7 天

策略调优

- 调优维度：滞后期 (5-10 天)、注意力头数 (1/2/4/8)、Dropout (0.05-0.25) 等 6 大方向
- 训练控制：Adam 优化器 ($1e-3$ 学习率 + $1e-5$ 权重衰减)、早停机制 (30 轮无提升终止)

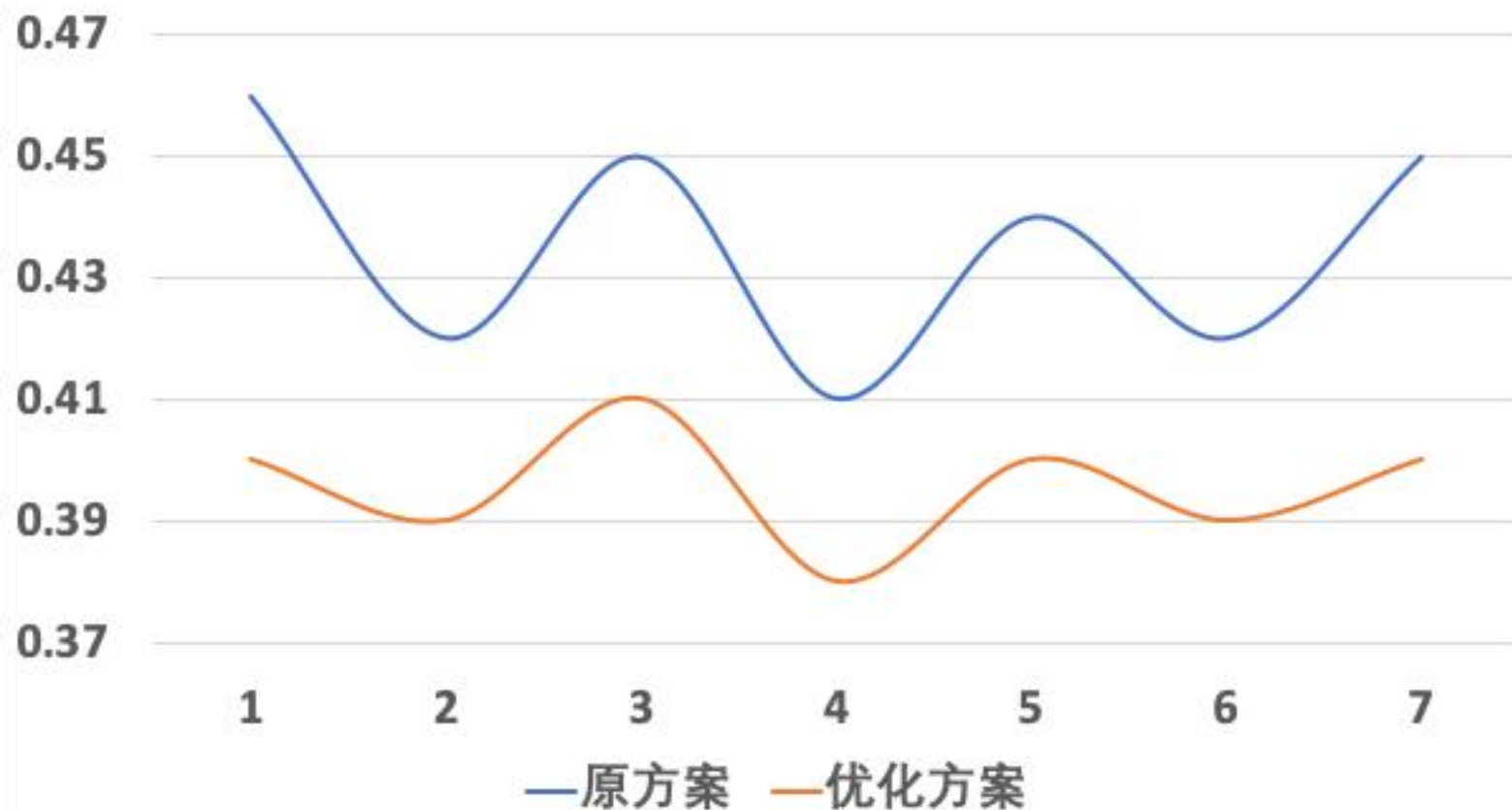
模型集成

- 基础模型：筛选验证集 WMAPE 最低的 3 个单模型
- 融合方式：贝叶斯加权平均 (按精度分配权重，降市场波动影响)
- 效果对比：集成后 WMAPE 再降 5-8%，抗噪音能力显著提升

效果对比



精度效果对比



- 精度与稳定性双提升
- 可复用优化框架，覆盖多金融场景



总结

总结&展望



模型设计

LSTM-Attention

注意力权重

残差连接与层归一化



训练优化

网格搜索调参

贝叶斯加权集成

早停机制



特征工程

大模型特征

热力图

爬虫逆向

未来方向

持续优化

- 新增政策文本等外部特征
- 构建流式计算框架，实现高频特征实时更新
- 新增在线学习模块，支持市场突变动态调参



跨金融场景应用

- 覆盖 ETF 申赎 / 股票量价 / 银行挤兑等场景
- 复用框架，适配场景化特征（折溢价率 / 存款流失率）
- 延伸业务边界，快速复制方案能力



AFAC2025

金融智能创新大赛

挑战组

赛题一：基金产品的长周期申购和赎回预测

感谢垂听，恳请指正

晨曦队