



基于注意力机制和列表排序算法的股票预测研究

Stock Prediction Based on Attention Mechanism and Listwise Sorting

答辩人：张雅楠 MF1933122
导师：申富饶





目录
CONTENT

01

研究
背景

02

研究
内容

03

研究
成果

04

总结
展望





01 研究背景



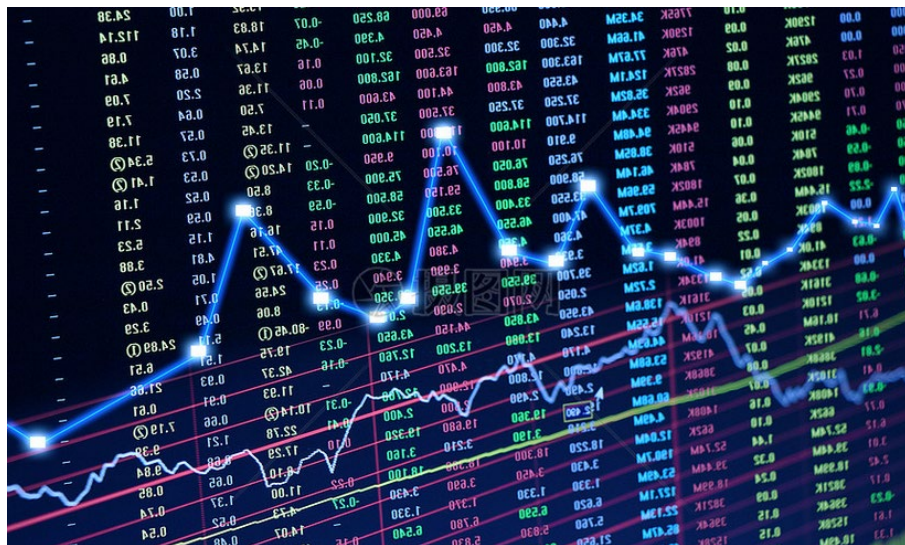
股票预测任务

是什么?

根据股票历史价格波动情况、交易量变化以及市场状况预测未来趋势。

意义:

- 指导投资者们投资，提高收益率；
- 增加市场流通性、活跃度。





相关方法



基于统计学概率的方法

- 确定性时间序列模型
- 不确定性时间序列模型
 - HMM

基于机器学习的方法

- 支持向量机 (SVM)
- 神经网络

基于深度学习的方法

- RNN、LSTM
- CNN
- Attention
- Bagging、Boosting
- 收集相关新闻、讨论数据

...



相关方法



基于统计学概率的方法

- 确定性时间序列模型
- 不确定性时间序列模型
- HMM

基于机器学习的方法

- 支持向量机 (SVM)
- 神经网络

基于深度学习的方法

- RNN | LSTM

针对单只股票分析的分类/回归问题



存在问题

- 单个逐一的预测，没有从**最高利润角度**进行模型的优化与选择；

股票		S1	S2	S3	S4	模型表现		利润 (Top-1)
真实涨跌幅		+50	+30	-10	-35			
预测涨跌幅 (回归问题)	R1	60	10	10	-45	MSE	250	50
	R2	25	35	-10	-25		187.5	30
预测涨跌概率 (分类问题)	C1	0.65	0.85	0.5	0.75	ACC	70%	30
	C2	0.8	0.65	0.45	0.7		50%	50

模型预测表现 \neq 更高的利润

- 单个逐一的预测，忽略了市场内不同股票之间的**相互影响**。
 - 同类型股票
 - 竞争类型股票

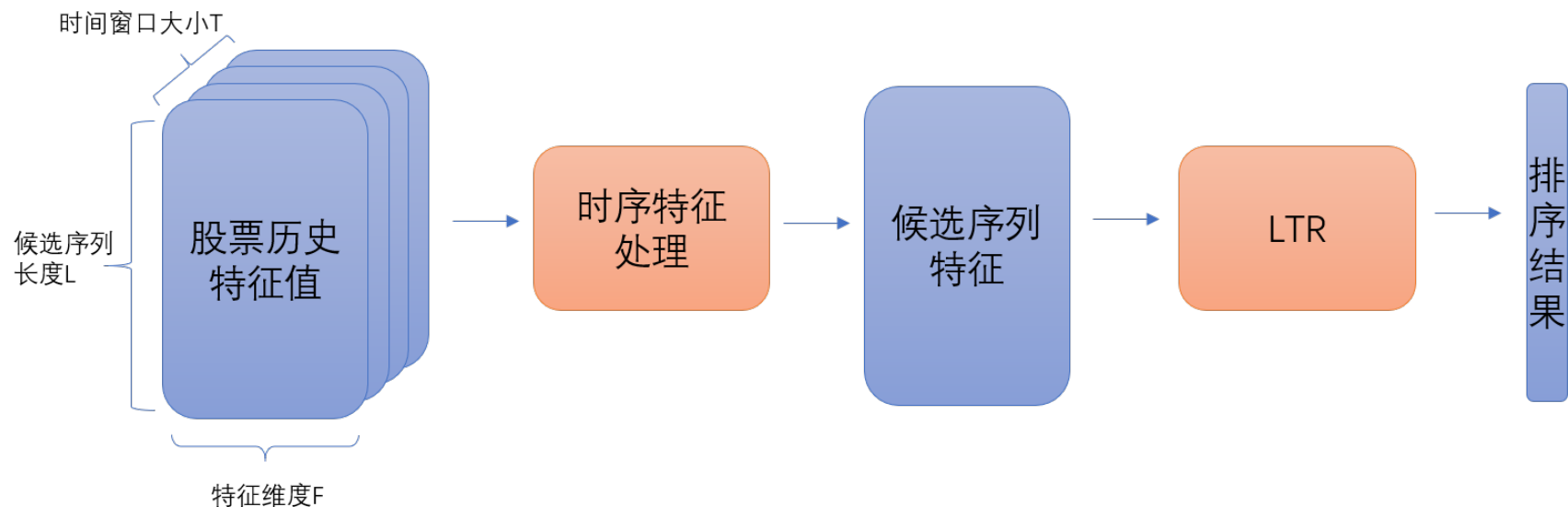
独立假设 **矛盾** 相互影响



问题定义

股票预测问题：分类/回归问题 → 学习排序问题 (LTR)

模型结构：



每日回报率：

$$r_i^t = \frac{c_i^t - c_i^{t-1}}{c_i^{t-1}} \times 100\%$$

最佳排名序列：

$$Y^t = \{y_1^t > y_2^t \dots > y_n^t\}$$

网络结果：（相对回报率大小情况）

$$y^t = f(F_1^{t-T:t}, F_2^{t-T:t}, \dots, F_n^{t-T:t})$$

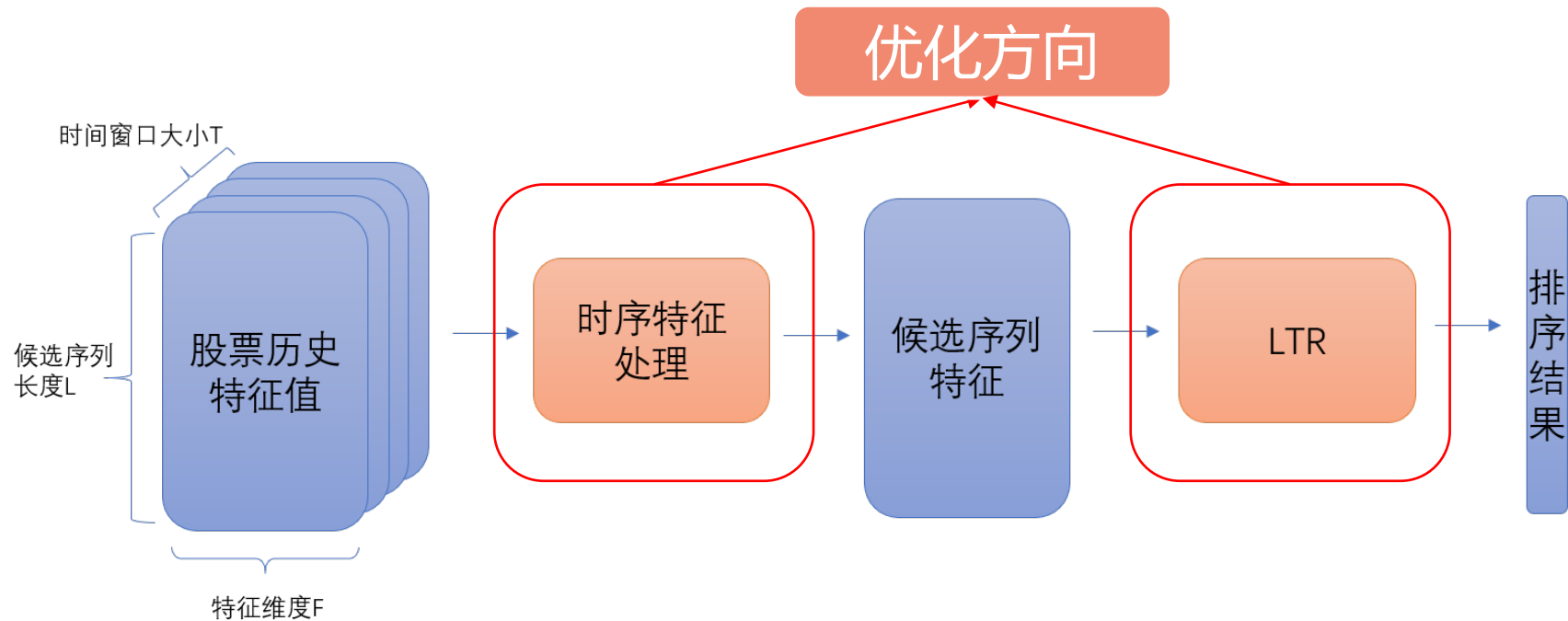




问题定义

股票预测问题：分类/回归问题 → 学习排序问题 (LTR)

模型结构：



每日回报率：

$$r_i^t = \frac{c_i^t - c_i^{t-1}}{c_i^{t-1}} \times 100\%$$

最佳排名序列：

$$Y^t = \{y_1^t > y_2^t \dots > y_n^t\}$$

网络结果：（相对回报率大小情况）

$$y^t = f(F_1^{t-T:t}, F_2^{t-T:t}, \dots, F_n^{t-T:t})$$





02 研究内容

- 基于注意力机制的时序特征处理方法
- 基于注意力解码器的列表排序算法



Part1: 基于注意力机制的时序特征处理方法

研究动机:

- 时序处理网络的长期遗忘问题;

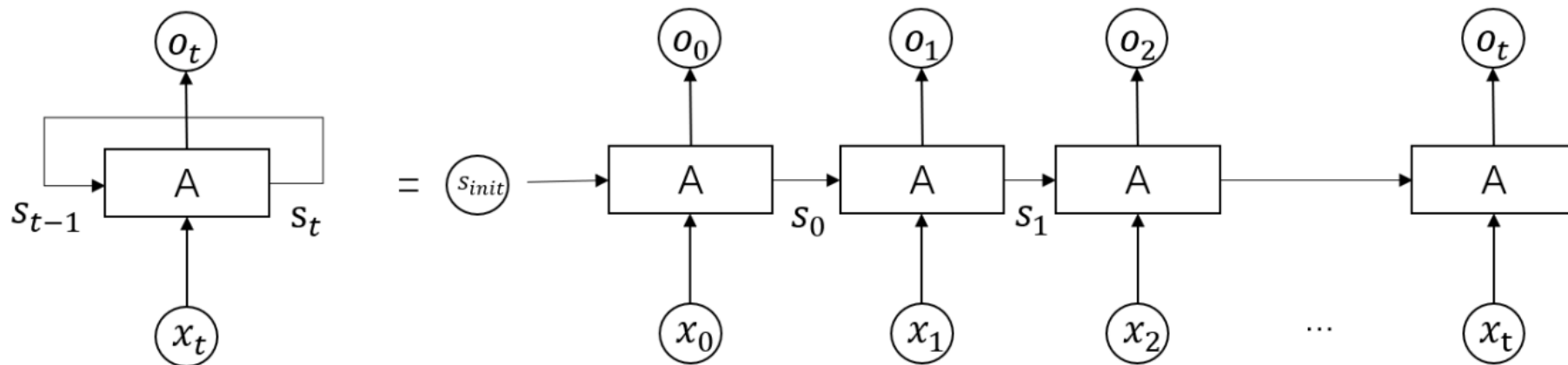


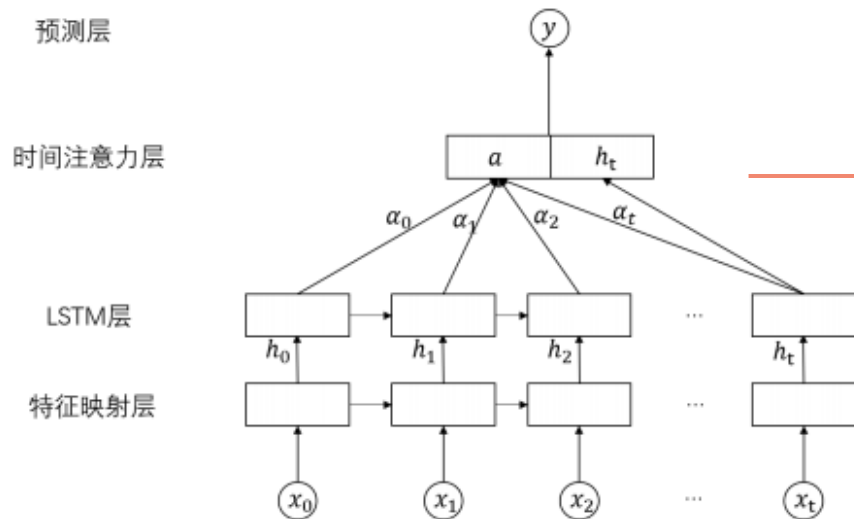
图 3-2: 循环神经网络 RNN 网络结构



Part1: 基于注意力机制的时序特征处理方法

研究动机:

ALSTM:



注意力机制整合历史特征

使用LSTM处理时序特征

$$\tilde{\alpha}_i^s = u_a^T \tanh(W_a h_i^s + b_a)$$

$$\alpha_i^s = \frac{\exp(\tilde{\alpha}_i^s)}{\sum_{i=1}^T \exp(\tilde{\alpha}_i^s)}$$

$$a^s = \sum_{i=1}^T \alpha_i^s h_i^s$$

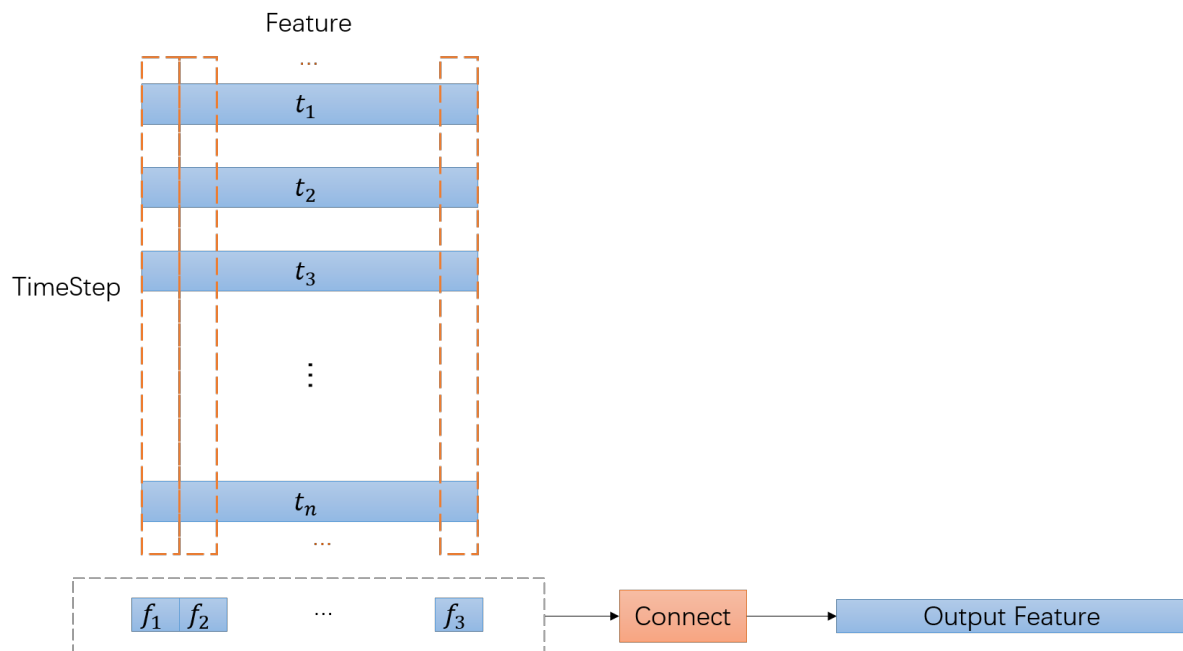
图 3-3: Attentive LSTM 网络结构^[1]

- 实际效率要求。GRU模型相比于LSTM模型具有更少的参数和更快的训练速度;
- 采用学习参数的方法计算注意力值, 增加了网络的不确定度。



Part1: 基于注意力机制的时序特征处理方法

特征卷积结构:



- 卷积核大小: $1 * T$
- 卷积核沿特征维度滑动

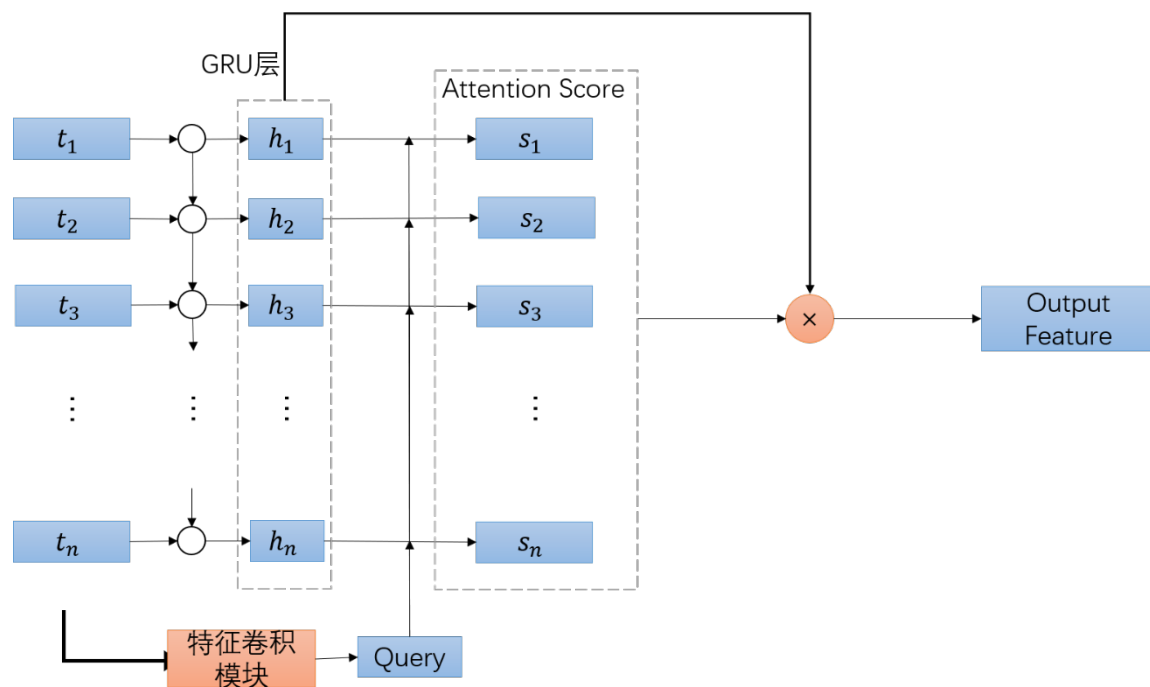
目的:

- 得到不同特征在历史时间步的变化信息



Part1: 基于注意力机制的时序特征处理方法

整体网络结构: AFC_GRU



处理流程:

- 使用特征卷积模块得到Query,
- 按时间顺序将历史特征输入GRU结构, 得到隐层输出,
- Query与隐层输出进行Attention计算。

输出:

- 单只股票的特征表示



Part1: 基于注意力机制的时序特征处理方法

数据集:

- **上证50**: 上海证券交易所内的50只代表性股票。
- **NASDAQ**: 美国纳斯达克股票市场内的股票数据。

特征构造:

- 基础特征 (每只股票每天的开盘价、收盘价、最高价、最低价、成交量、成交额)
- 变化趋势特征

表 3-1: 描述股票每日趋势的特征

特征	计算示例
c_open, c_high, c_low	$c_open = open_t / close_t - 1$
n_open, n_close	$n_close = close_t / close_{t-1} - 1$
$5 - day, 10 - day, 15 - day, 20 - day$	$5 - day = \frac{\sum_{i=0}^4 close_{t-i} / 5}{close_t} - 1$



Part1: 基于注意力机制的时序特征处理方法

实验对比:

表 3-2: 在上证 50 数据集上, 不同时序特征处理方法的结果对比

	MAP@5	MAP@10	P@5	P@10	NDCG@10
RNN	0.11389	0.18315	0.22000	0.34167	0.57823
LSTM	0.11222	0.17655	0.20667	0.33833	0.57487
ALSTM	0.14133	0.20552	0.22667	0.35500	0.59536
GRU	0.14239	0.19242	0.25333	0.35167	0.57928
AGRU	0.14322	0.22097	0.22000	0.37167	0.61303
TCN	0.13261	0.17829	0.22667	0.32000	0.58193
AFC_GRU(ours)	0.15294	0.23474	0.24333	0.38333	0.61598

表 3-3: 在 NASDAQ 数据集上, 不同时序特征处理方法的结果对比

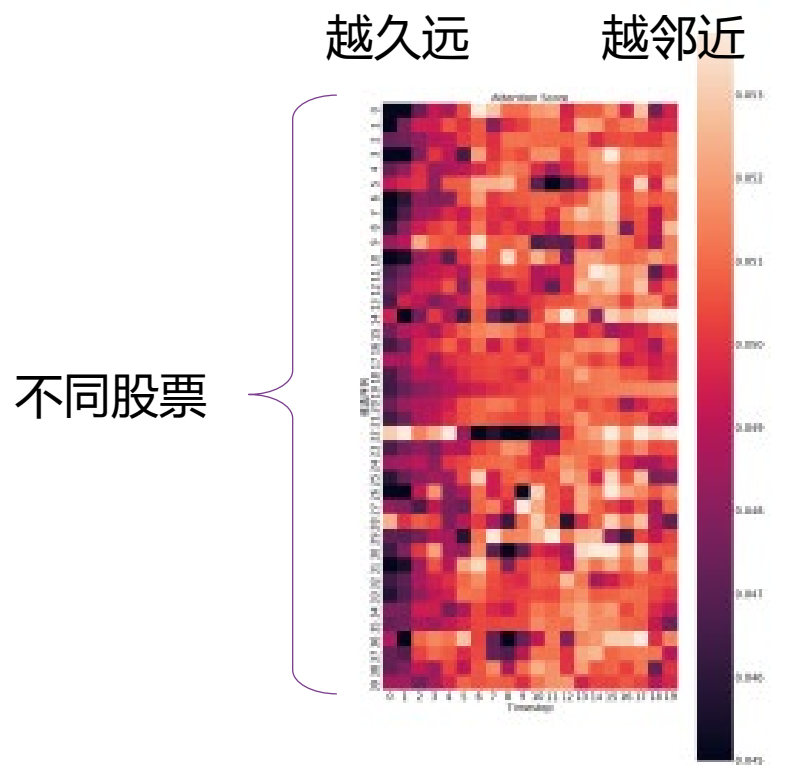
	MAP@5	MAP@10	P@5	P@10	NDCG@10
RNN	0.14592	0.15577	0.24719	0.30000	0.54328
LSTM	0.14022	0.16375	0.24270	0.30562	0.55463
ALSTM	0.12779	0.14743	0.22022	0.28539	0.55077
GRU	0.14408	0.15909	0.22921	0.30000	0.54814
AGRU	0.13273	0.15870	0.22472	0.30787	0.55343
TCN	0.14322	0.15462	0.22697	0.29213	0.54778
AFC_GRU(ours)	0.14801	0.16813	0.23596	0.31685	0.56174

- 与其他时序处理方法比较, 我们的方法在大多指标上结果都最好;
- GRU结构与LSTM方法比较, 排序召回更精准, 说明GRU结构能够更好地表示时序特征;
- 与AGRU方法比较, AFC_GRU排序召回情况更好, 证明了我们特征卷积方法的有效性。



Part1: 基于注意力机制的时序特征处理方法

Attention可视化:



每支股票在历史时间步的注意力分布

- 由于股票数据特性，较近历史的注意力值会较大，
- 但是AFC_GRU算法在一定程度上缓解了长期历史遗忘问题。



Part1: 基于注意力机制的时序特征处理方法

赢利能力回测对比:

使用**每日买入-持有-次日卖出的T+1策略**, 进行模拟投资。

1. 在第t交易日收盘时, 平均分配预算, 购入预期收益TOP的股票;
2. 在第t+1交易日, 卖出交易日t买入的股票;
3. 忽略交易成本。

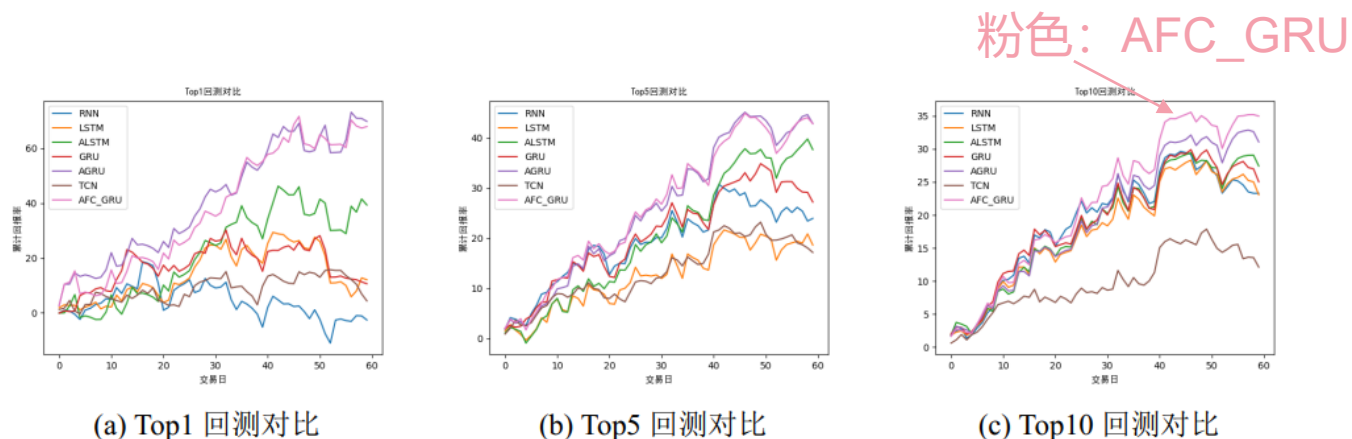


图 3-7: 在上证 50 数据集上, 不同投资模型在各个回测方法下的结果比较

结论:

- 本文提出的AFC_GRU时序特征处理算法能够有效对股票特征进行数据处理, 从而实现提高回报率的目的。



Part1: 基于注意力机制的时序特征处理方法

消融实验:

卷积核大小对模型性能的影响

Kernel Size	MAP@5	MAP@10	P@5	P@10	NDCG@10
1	0.15294	0.23474	0.24333	0.38333	0.61598
2	0.13939	0.22517	0.22667	0.38000	0.61071
3	0.13778	0.19609	0.22000	0.35000	0.58750

特征连接方式对模型性能的影响

Connection Method	MAP@5	MAP@10	P@5	P@10	NDCG@10
replace	0.15294	0.23474	0.24333	0.38333	0.61598
connect	0.14944	0.21526	0.26333	0.37667	0.60403
reduction after connect	0.14628	0.19635	0.23667	0.35167	0.59358

时间步大小选择对模型结果的影响

Timestep	MAP@5	MAP@10	P@5	P@10	NDCG@10
5	0.13178	0.17895	0.23333	0.31833	0.57534
10	0.13561	0.19536	0.21667	0.33500	0.58708
15	0.11533	0.18743	0.20000	0.34333	0.58109
20	0.15294	0.23474	0.24333	0.38333	0.61598
30	0.16983	0.22769	0.27000	0.38167	0.61313
40	0.17861	0.23349	0.26667	0.37500	0.61012

结论:

- 卷积核大小为**1**，能够更好地捕获特征的变化规律；
- 时间步长为**20**，更好地捕获历史变化趋势；
- **直接取代的特征表示**，能够得到更好的时序特征。



02 研究内容

- 基于注意力机制的时序特征处理方法
- 基于注意力解码器的列表排序算法



Part2: 基于注意力解码器的列表排序算法

研究动机:

- 传统排序的打分阶段没有综合考虑候选项之间的影响;

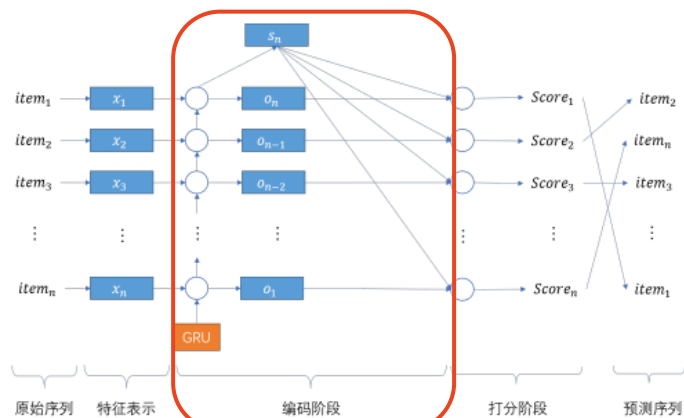


图 4-1: DLCM 模型结构图^[2]

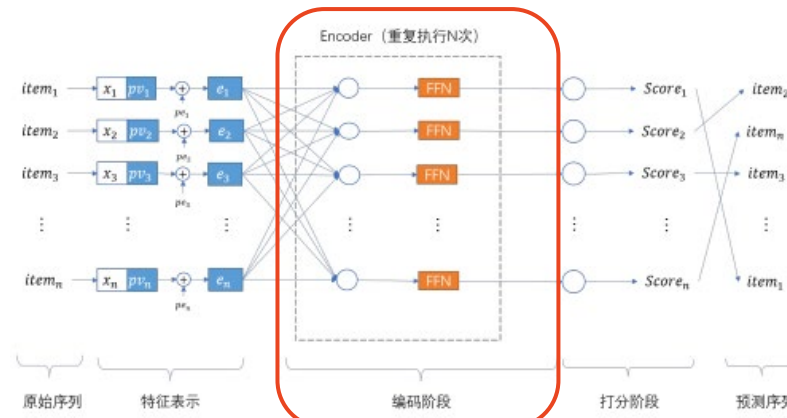


图 4-2: PRM 模型结构图^[3]

- loss函数对整体序列进行约束，而股票预测中更关注前端序列情况。

$$\ell(R_q^n) = - \sum_{d_i \in R_q^n} (a_i^y \log(a_i^S) + (1 - a_i^y) \log(1 - a_i^S))$$



Part2: 基于注意力解码器的列表排序算法

研究动机:

- 传统排序的打分阶段没有综合考虑候选项之间的影响;

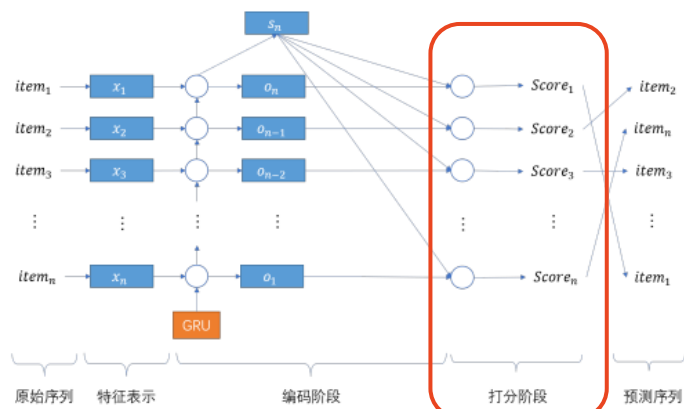


图 4-1: DLCM 模型结构图^[2]

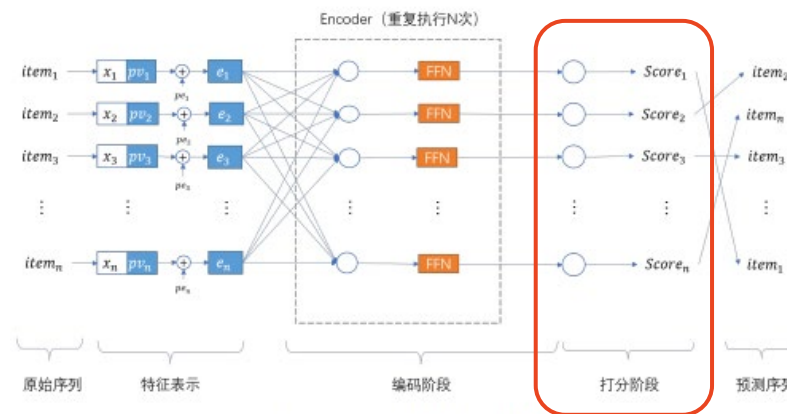


图 4-2: PRM 模型结构图^[3]

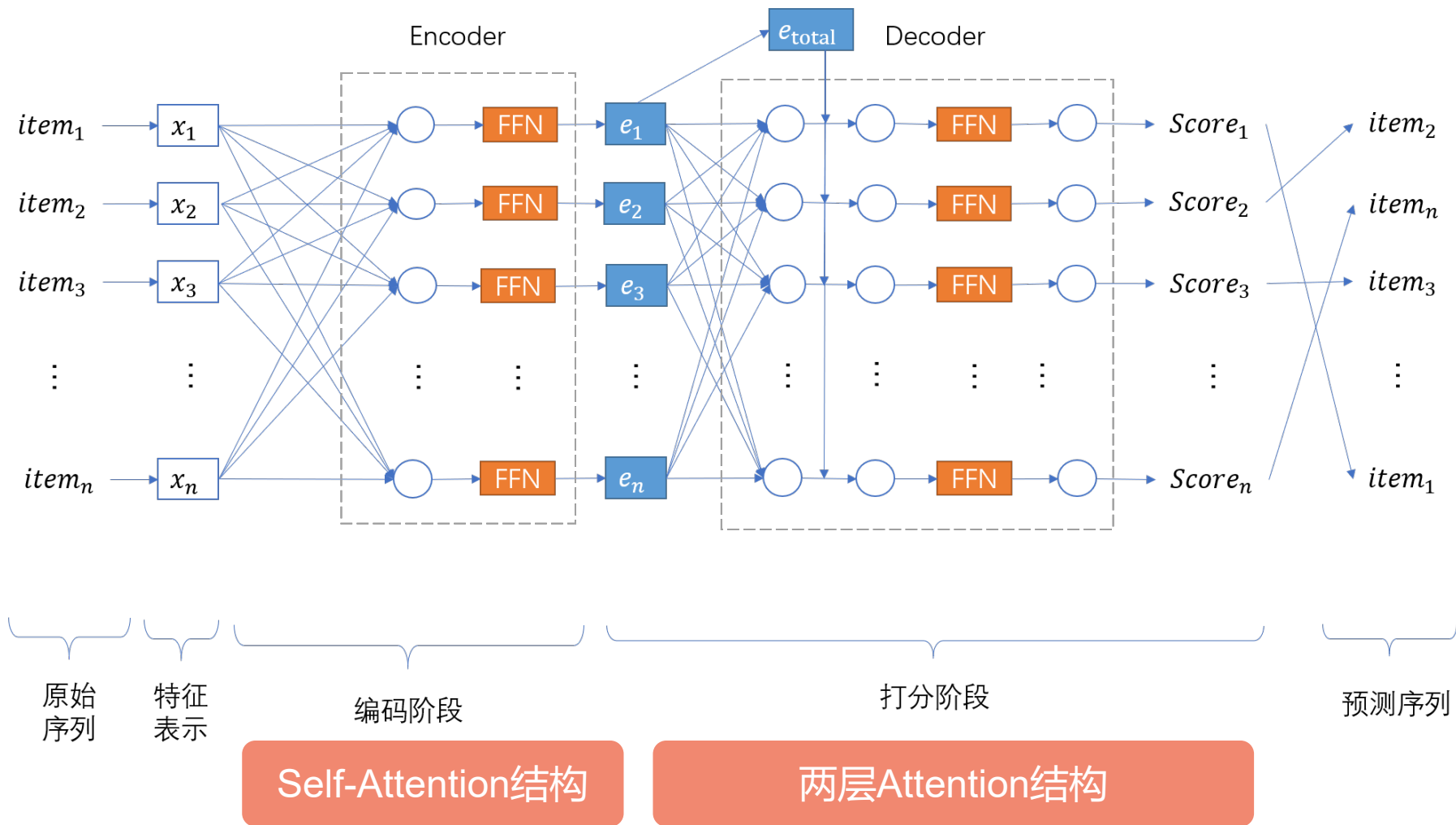
- loss函数对整体序列进行约束, 而股票预测中更关注前端序列情况。

$$\ell(R_q^n) = - \sum_{d_i \in R_q^n} (a_i^y \log(a_i^S) + (1 - a_i^y) \log(1 - a_i^S))$$



Part2: 基于注意力解码器的列表排序算法

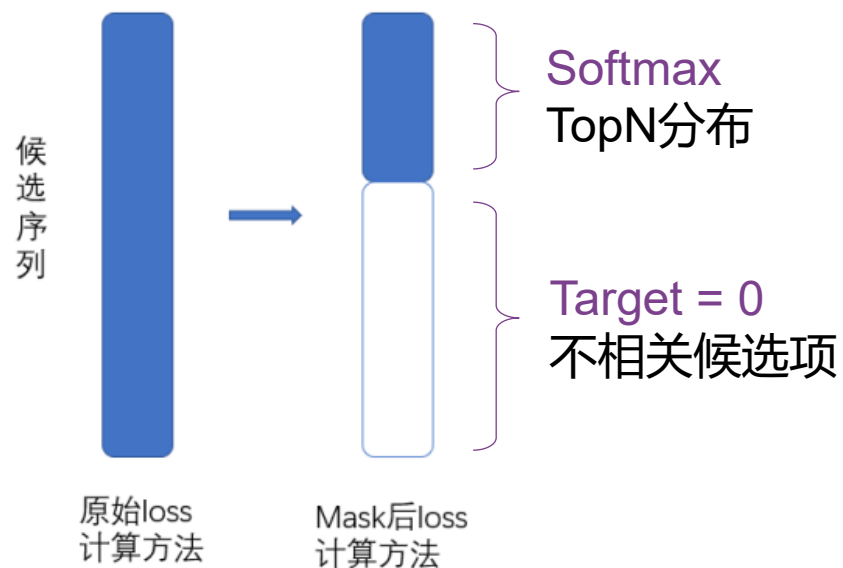
模型结构: AEDRM, 基于Listwise排序方法





MASK Attention Rank Loss:

- 更加关注前端的排序情况;
- 尾部候选项则只关注前后序列的区分。



$$y^{mask} = y \times mask$$

$$a_i^{y^{mask}} = \frac{\psi(y_{(q,d_i)}^{mask})}{\sum_{d_k \in R_q^n} \psi(y_{(q,d_k)}^{mask})}$$

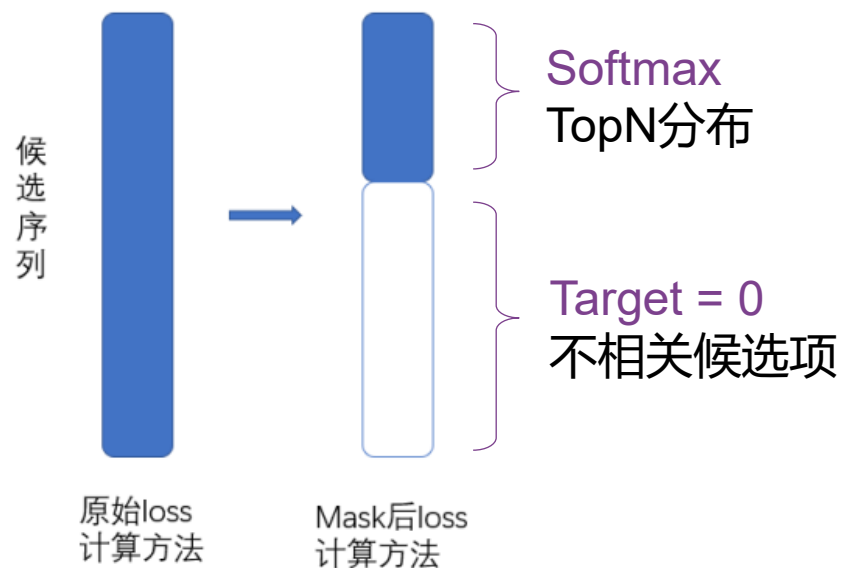
$$\ell(R_q^n) = - \sum_{d_i \in R_q^n} \left(a_i^{y^{mask}} \log(a_i^S) + (1 - mask) (1 - a_i^{y^{mask}}) \log(1 - a_i^S) \right)$$

图 4-6: mask 方法



MASK Attention Rank Loss:

- 更加关注前端的排序情况;
- 尾部候选项则只关注前后序列的区分。



$$y^{mask} = y \times mask$$

$$a_i^{y^{mask}} = \frac{\psi(y_{(q,d_i)}^{mask})}{\sum_{d_k \in R_q^n} \psi(y_{(q,d_k)}^{mask})}$$

$$\ell(R_q^n) = - \sum_{d_i \in R_q^n} \left(a_i^{y^{mask}} \log(a_i^S) + (1 - mask)(1 - a_i^{y^{mask}}) \log(1 - a_i^S) \right)$$

对Topn的注意力分布情况约束

对序列尾部约束

图 4-6: mask 方法



实验对比:

表 4-1: 在上证 50 数据集上, 不同排序方法的结果对比

	MAP@5	MAP@10	P@5	P@10	NDCG@5	NDCG@10	NDCG
PointNet	0.04689	0.08460	0.09667	0.20167	0.51545	0.55002	0.82415
DLCM	0.11133	0.17544	0.19333	0.31333	0.56719	0.58600	0.83883
PRM	0.12744	0.18760	0.18667	0.30833	0.58768	0.59288	0.84481
AEDRM(ours)	0.15894	0.20874	0.25000	0.35500	0.59313	0.60420	0.84570

表 4-2: 在 NASDAQ 数据集上, 不同排序方法的结果对比

	MAP@5	MAP@10	P@5	P@10	NDCG@5	NDCG@10	NDCG
PointNet	0.11000	0.12963	0.20449	0.26180	0.51405	0.54507	0.82107
DLCM	0.10880	0.14021	0.19326	0.28764	0.52891	0.55831	0.82937
PRM	0.12375	0.14638	0.21798	0.29326	0.51545	0.54845	0.82406
AEDRM(ours)	0.14801	0.15515	0.23146	0.31685	0.53533	0.56884	0.83303

- 在MAP、P指标上, AEDRM 优于其他模型, 说明该排序模型能够更好召回前端序列, 并进行排序;
- 整体NDCG只略高于其他模型, NDCG@5与NDCG@10提升较明显, 说明模型加强了对前端序列的关注。



Part2: 基于注意力解码器的列表排序算法

赢利能力回测对比: 使用T+1投资策略进行模拟投资分析

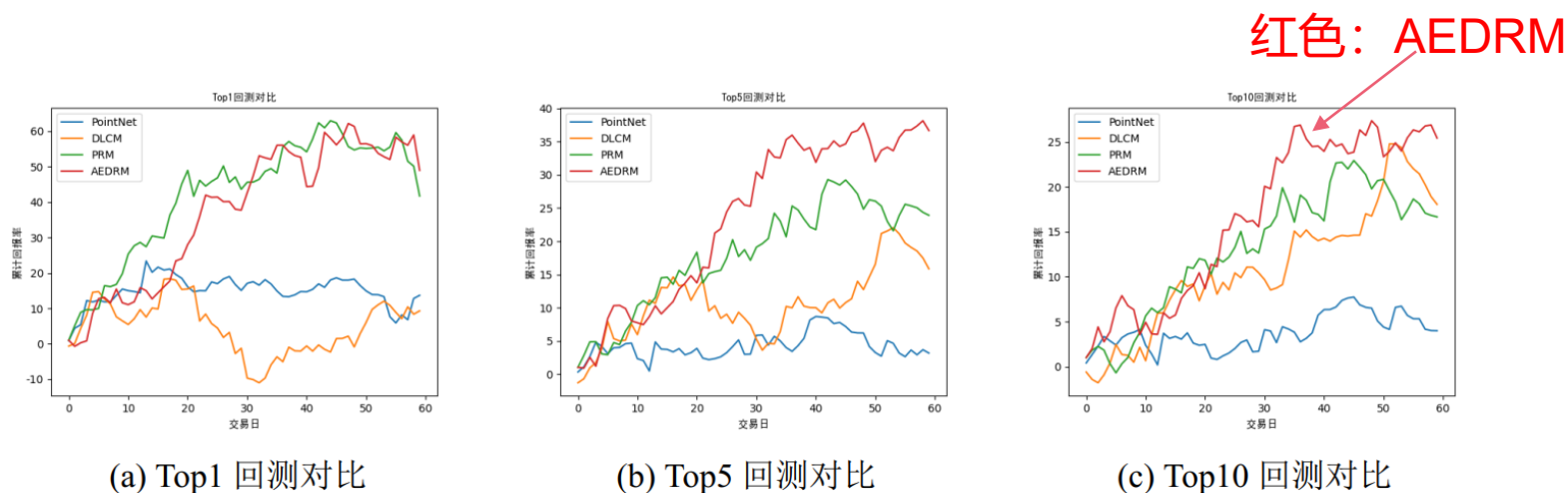


图 4-8: 在上证 50 数据集上, 不同投资模型在各个回测方法下的结果比较

结论:

- **Top1**回测, AEDRM与PRM结果不相上下, **Top5**与**Top10**回测, AEDRM优于其他模型;
- AEDRM 排序模型能够很好地对**回报率更大的股票**进行召回排序, 从而达到提高收益的结果。



消融实验:

编码器个数对模型结果的影响

Block	MAP@5	MAP@10	P@5	P@10	NDCG@5	NDCG@10	NDCG
1	0.12744	0.18760	0.18667	0.30833	0.58768	0.59288	0.84481
2	0.12694	0.18955	0.20667	0.32000	0.58333	0.59212	0.84274
3	0.12617	0.18543	0.21000	0.31000	0.58217	0.59178	0.84191

Loss设计对模型结果的影响

	MAP@5	MAP@10	P@5	P@10	NDCG@5	NDCG@10	NDCG
loss_point	0.06039	0.09549	0.11000	0.22333	0.50888	0.54391	0.82165
loss_attn	0.14822	0.22886	0.25333	0.37667	0.59875	0.60941	0.84717
loss_01	0.15800	0.21839	0.25667	0.36000	0.59026	0.60167	0.84415
loss_mask_attn	0.15894	0.23009	0.25333	0.37833	0.60003	0.60915	0.84778

验证集选择与使用方法对模型结果的影响

	MAP@5	MAP@10	P@5	P@10	NDCG@5	NDCG@10	NDCG
OriDev	0.15178	0.20588	0.23000	0.35000	0.56657	0.59186	0.84104
NoDev	0.15317	0.22764	0.25000	0.37167	0.59361	0.60651	0.84529
RdmDev	0.15894	0.23009	0.25333	0.37833	0.60003	0.60915	0.84778
Fine-tune	0.14489	0.21669	0.23333	0.36000	0.58308	0.59524	0.84257

- 增加编码器个数没有对整体排序结果带来明显提升。Block=1排序结果与效率更加平衡。

- MAR loss计算方法能够更好地对TopN进行召回排序，整体性能也更加稳定，符合设计预期。

- 随机抽取验证集能够有效缓解由于时间导致的数据分布不一致问题，有效利用数据集学习。



03 研究成果



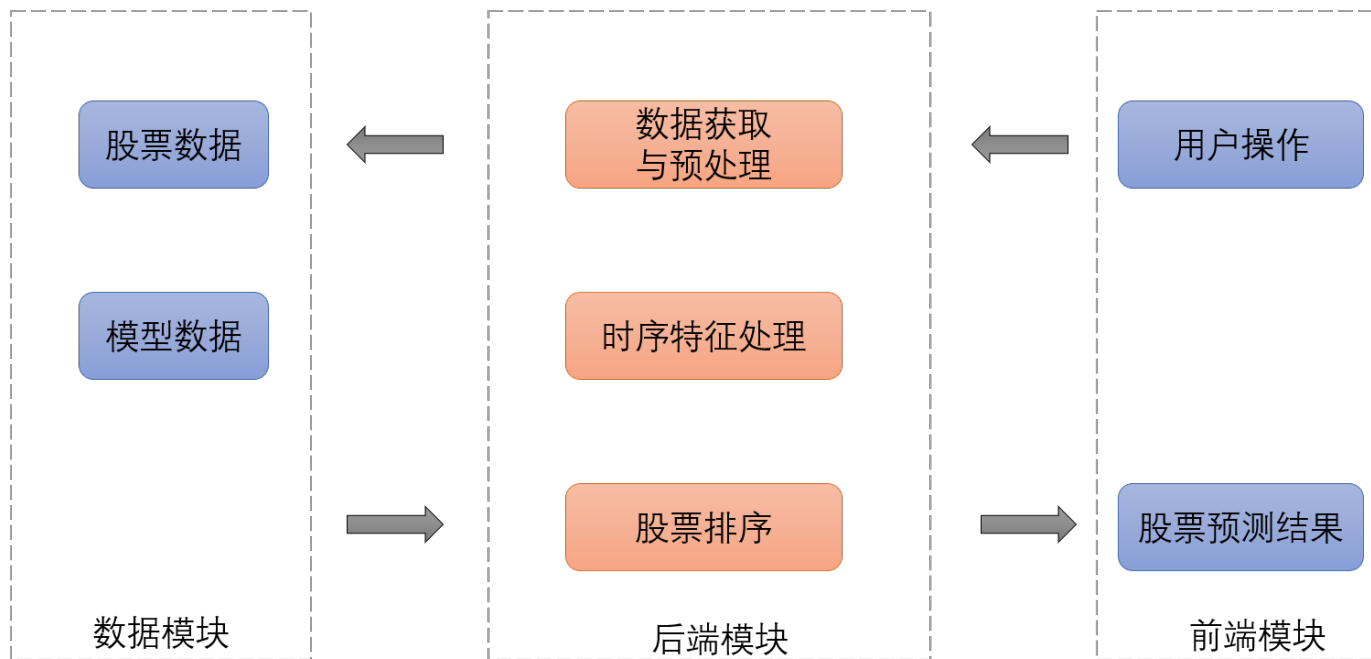
系统设计

股票分析预测系统

功能需求:

- 获取数据
- 预测排序
- 回测分析

系统架构



实现方法





数据查询：

股票数据查询

股票ID:

时间选择:

~

交易时间	开盘价	最高价	最高价	收盘价
2017-01-24	18.59	18.88	18.5	18.87
2017-01-25	18.81	18.95	18.68	18.88
2017-01-26	18.9	19.25	18.82	19.0
2017-02-03	19.08	19.08	18.66	18.7
2017-02-06	18.88	18.9	18.54	18.62
2017-02-07	18.63	18.68	18.55	18.57
2017-02-08	18.55	18.6	18.48	18.6
2017-02-09	18.59	18.93	18.59	18.73
2017-02-10	18.86	18.97	18.73	18.9
2017-02-13	18.9	19.09	18.89	18.97
2017-02-14	19.0	19.03	18.81	18.87
2017-02-15	18.97	19.23	18.9	19.13
2017-02-16	19.19	19.34	19.04	19.14



预测排序:



预测排序

模型选择:

训练新模型

时间选择:

开始预测

序号	股票ID
0	sh.600588
1	sh.600309
2	sh.600050
3	sh.600585
4	sh.603986
5	sh.600690
6	sh.601688
7	sh.601857
8	sh.600031
9	sh.601166
10	sh.601601
11	sh.601398
12	sh.600436
13	sh.601288



预测排序:



预测排序

模型选择: AERM
训练新模型
时间选择: 2022-03-09
开始预测

序号	股票代码
0	sh.600588
1	sh.600309
2	sh.600050
3	sh.600585
4	sh.603986
5	sh.600690
6	sh.601688
7	sh.601857
8	sh.600031
9	sh.601166
10	sh.601601
11	sh.601398
12	sh.600436
13	sh.601288

训练新模型:

训练新模型

模型名称: model1
股票代码: sh.600031 三一重工, sh.600...
时间选择: 2020-03-01
~
2022-03-01

关闭 开始训练

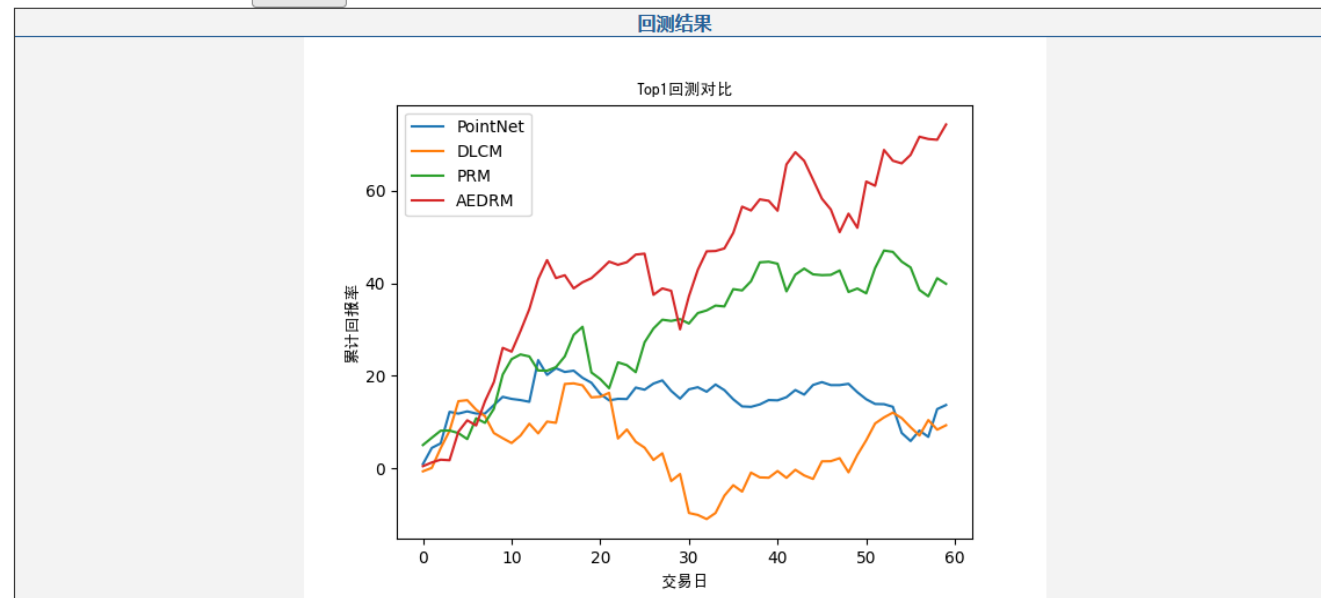


回测分析:

回测分析

模型选择:

时间选择:





专利:

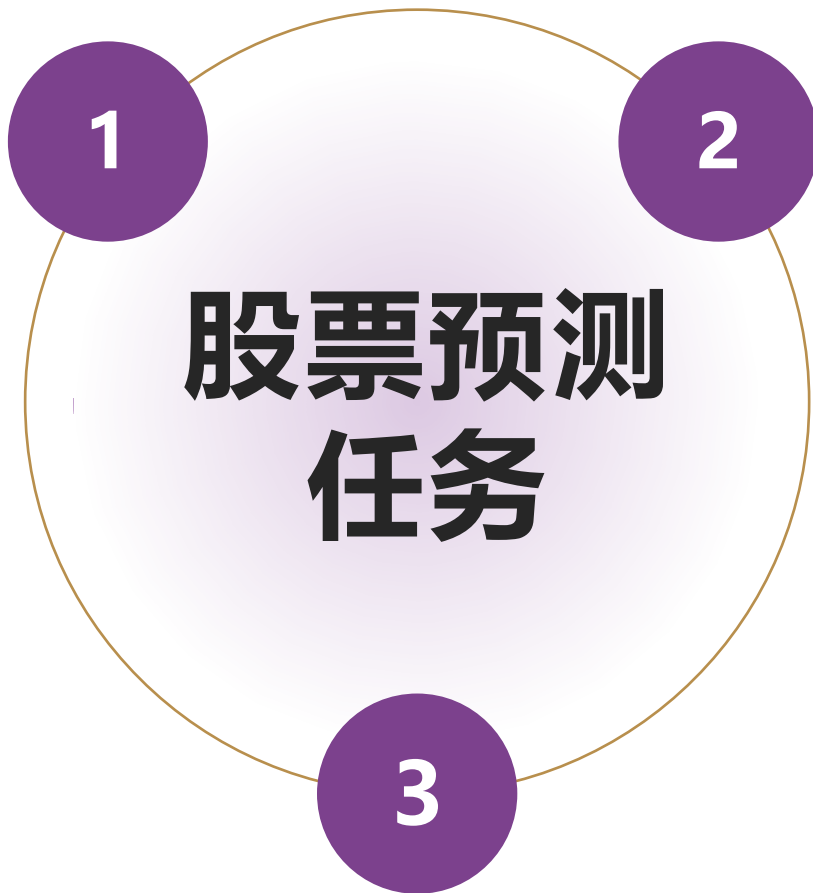
- 吴宜真, **张雅楠**, 严骅, 刘晓涛, 申富饶。 “基于 LDA 和 Bert 的特定类别文本标题二分类方法” (202011177997.X)

项目:

- 国家自然科学基金面上项目。基于深度感知增量式联想记忆神经网络的信息融合系统研究, Information fusion system based on deep perception and incremental associative memory neural networks” (课题年限 2019.01 ~ 2022.12)
-



04 总结



时序特征处理算法AFC_GRU

利用特征卷积结构与GRU进行注意力计算，更新股票特征表示

股票序列排序算法AEDRM

基于注意力解码器的Listwise排序算法，并使用MAR loss增加头部关注

股票分析预测系统

查看历史、预测涨跌、分析对比预测模型



谢谢



南 京 大 学