

# 知乎搜索排序模型的探索实践

王瑞欣

知乎搜索算法团队

2020年11月

知乎

# 大纲

- 多目标排序
- Unbias LTR
- Context Aware
- End2End
- Session Aware

知乎

# 知乎搜索

- 问答搜索
- 图文搜索

← 面试自我介绍

综合 实时 用户 话题 视频 盐选专栏

Live 更多 >

面试中的「自我介绍」怎么做？

知

小红拖拉机

🌟

人力资源（HR）等 3 个话题优秀回答者

与机会擦肩而过。那么，在这些问题背后，**面试**官究竟在问什么？在这 7 期 L...

Live · 2.1 万热度

面试自我介绍5句话公式

苏三：No2不练够十小时不要上台！ 请： 大声读**自我介绍** 记住**自我介绍**的重点和顺序。 背诵**自我介绍**，直到听起来...

1.4 万赞同 · 211 评论 · 2016-12-03

🌸 圆桌精选

面试时怎样自我介绍比较好？

夏姑娘：前面说到，**自我介绍**不是**自我介绍**本身，是**面试**的开场。从这个不...

2.6 万赞同 · 184 评论 · 04-29

👤 你关注的「小红拖拉机」赞同了该回答

面试应答有哪些话术和技巧？

知乎求职：灯下说书人如何在**面试**的无领导小组讨论环节中脱颖而出？ **面试**...

40 赞同 · 11-05

← 隐秘而伟大

为什么娱乐圈的爆料，渐渐发现基本上都是真的？

叭叭：迪丽热巴后面的红毯风格方向会走那种轻熟性感风 《**隐秘而伟大**》一直不播不是政策原因，而是现在有两家...

4455 赞同 · 1138 评论 · 10-20

作为圈内人，娱乐圈有多脏？

椒下客：9更： 1.李易峰、金晨主演的 《**隐秘而伟大**》豆瓣评分8.4，口碑挺...

1317 赞同 · 81 评论 · 11-09

《**隐秘而伟大**》男二号王泷正为何不红？

娱乐圈有瘾： 这一次，王泷正等来了《**隐秘而伟大**》这部当红明星李易峰主演...

56 赞同 · 20 评论 · 11-12

电视剧《**隐秘而伟大**》有哪些不易发现但很精彩的细节？

枕南乡： 黄玫瑰 据说是代表顾耀东要投身**伟大**的革命事业了 还有小黄狗（和...

29 赞同 · 5 评论 · 7 小时前

《**隐秘而伟大**》“青成”之恋仅留余声，他终究只是抱了抱那个女孩

郭健：王泷正和金晨主演的民国剧 《**隐秘而伟大**》正在热播中，很难想象出...

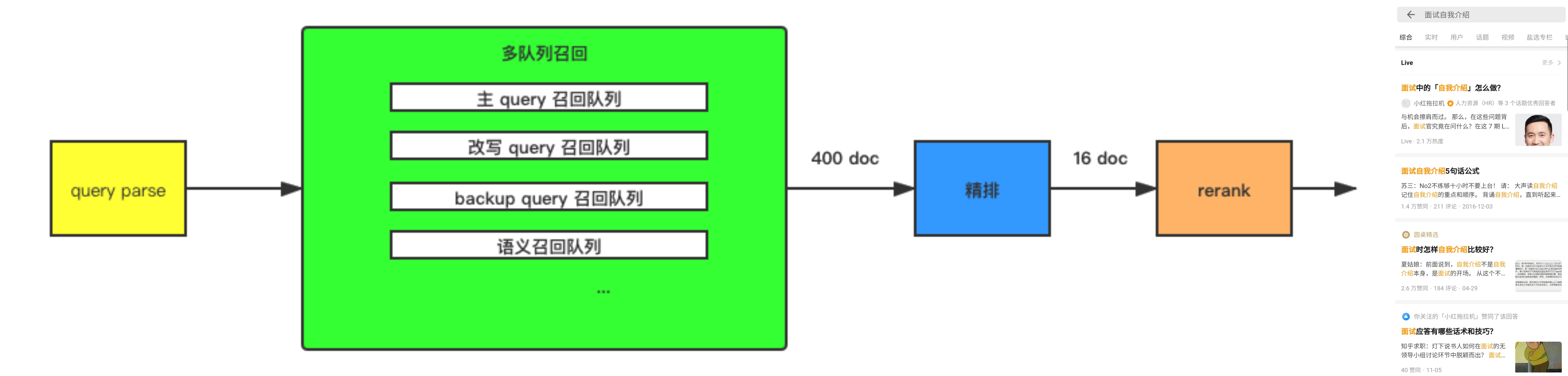
知乎

# 发展历程

2017.12	搜索算法团队成立
2018.08	深度语义相关性
2019.04	BERT & 语义向量索引
2019.08	GBDT -> DNN 升级
2019.10	Unbiased LTR
2019.11	MultiTask LTR
2020.02	SERank
2020.03	回答静态分 DNN 模型
2020.04	LTR End2End
2020.05	LTR MMoE
2020.06	LTR Session Aware

知乎

# 搜索架构



(a) 搜索架构

知乎

# 从 GBDT 到 DNN

为什么要迁移到 DNN?

- 数据量变大可以支持更高的模型复杂度
- 支持最新的研究成果，比如多目标排序

上线收益?

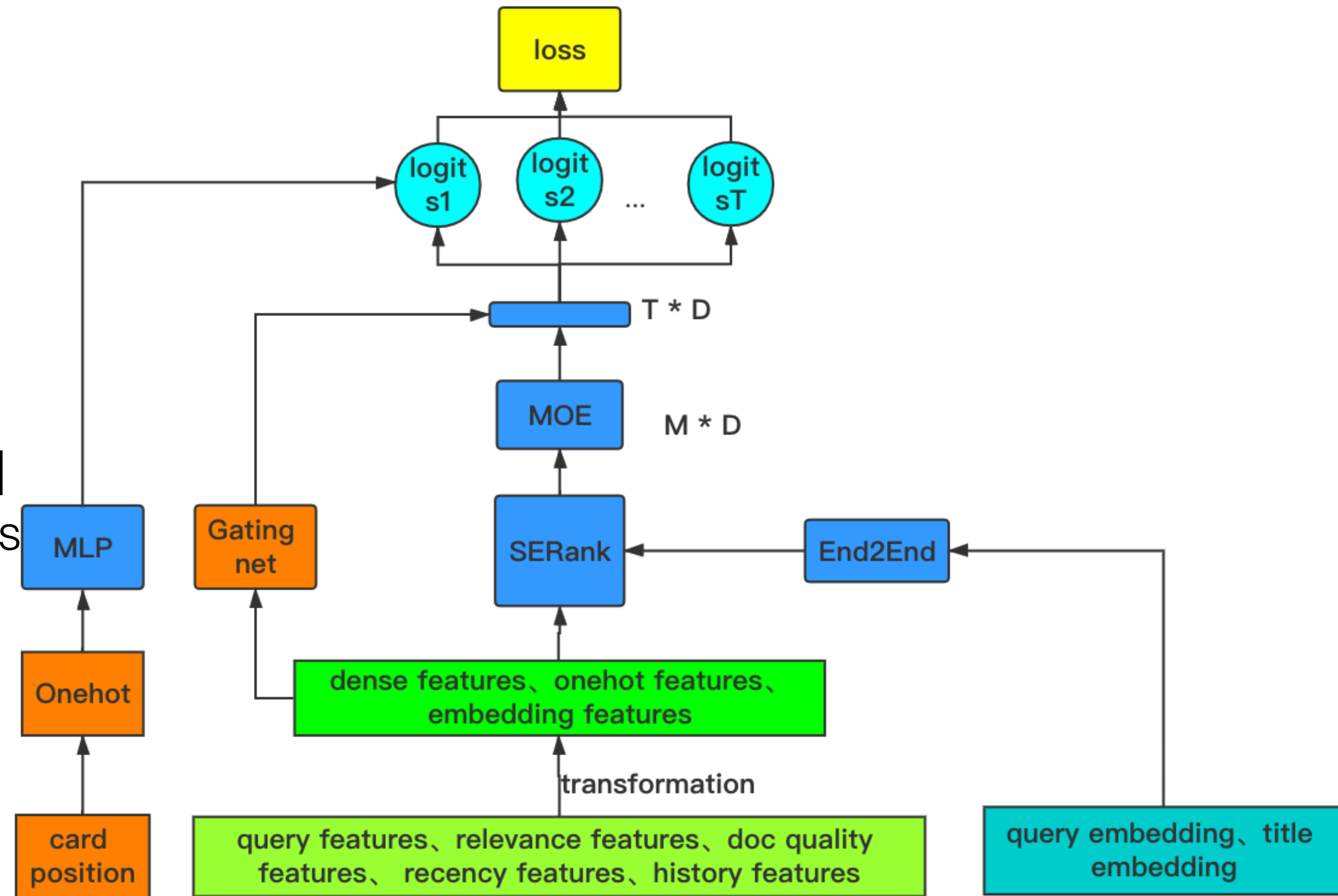
- 点击比 +0.5%
- 点展比 +0.7%
- 人均时长 +0.75%

知乎



# 排序模型

- 基于 TF-Ranking 开发 [1]
  - Ranking loss & metrics
  - Estimator api
  - Multi head



(a) 排序模型结构

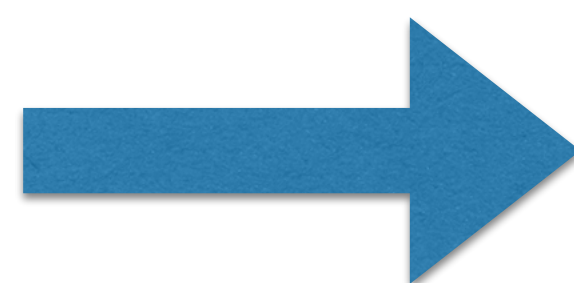
知乎

# 多目标排序

排序任务的目标是什么？

满意度

预估点击率



阅读时长

点赞、收藏、关注、分享、评论

分析与用户留存的关系，确定主要的目标

知乎



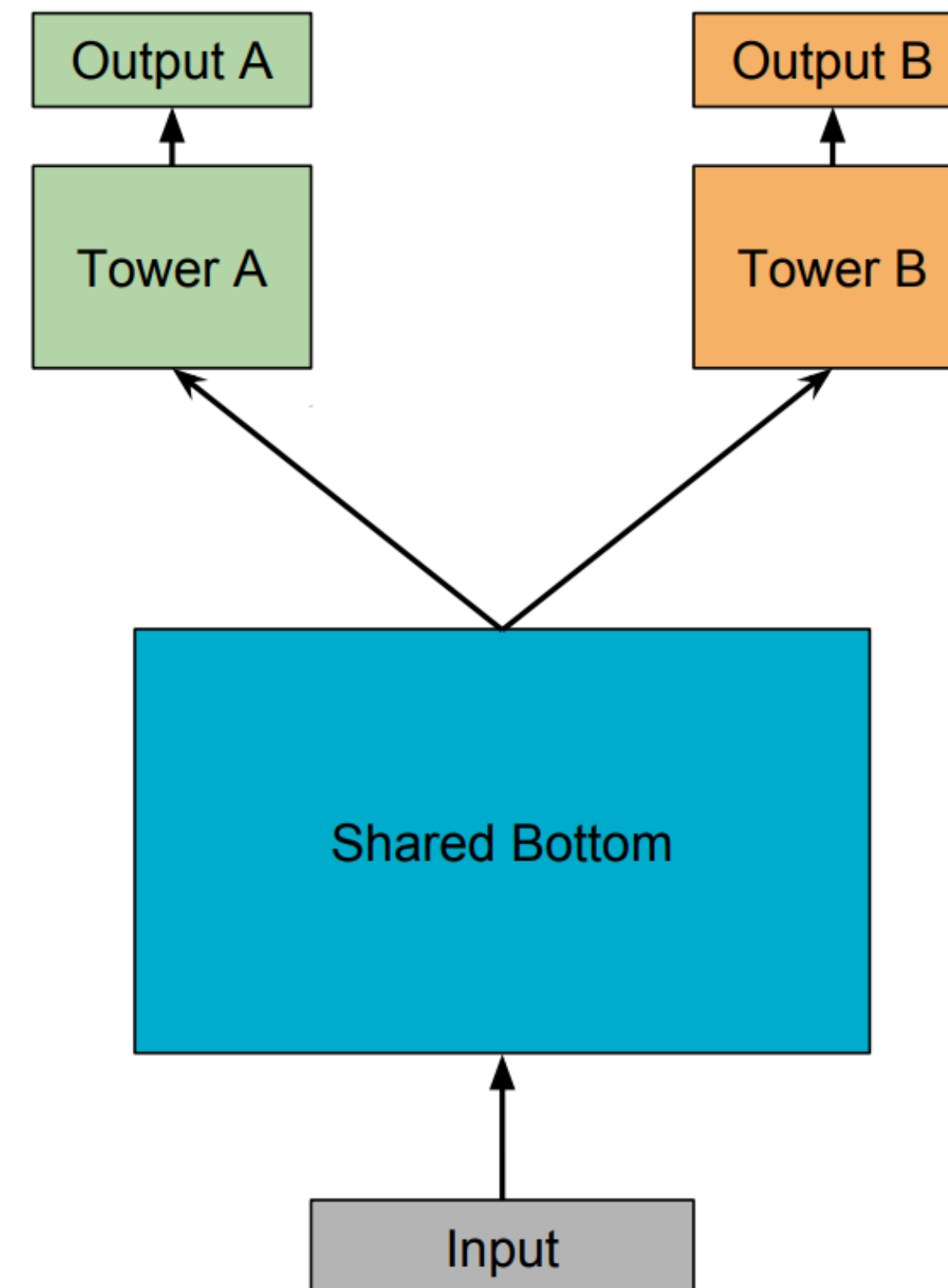
# 多目标排序

## 第一版 hard sharing

- 各任务如何计算 Loss?
- 在线预测时分数怎么计算?

上线收益?

- 人均阅读时长 +1.07%
- 次留 +0.16%



(a) Shared Bottom struct

1. Ma, Jiaqi, et al. "Modeling task relationships in multi-task learning with multi-gate mixture-of-experts." *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2018.

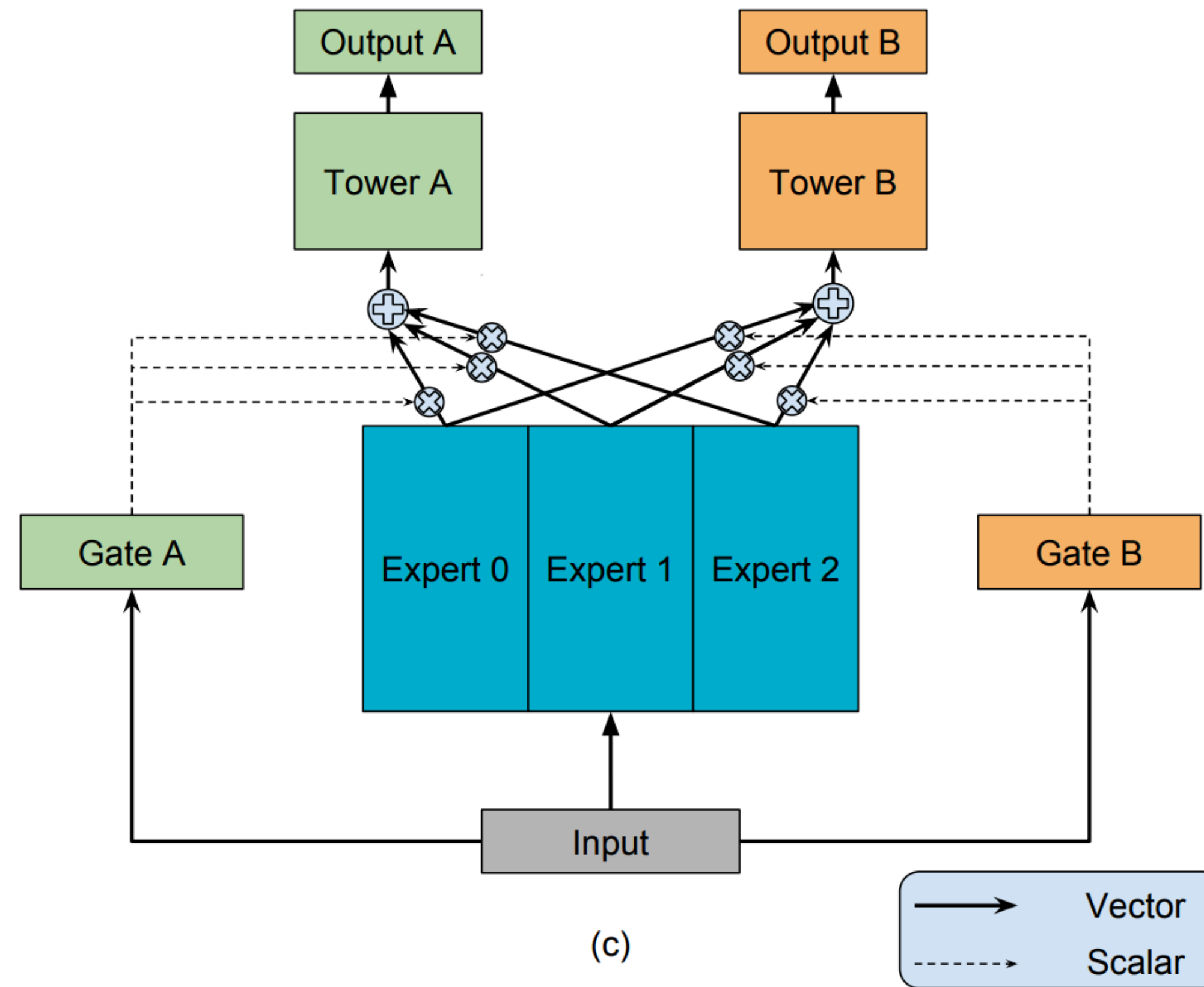
知乎

# 多目标排序

## 第二版 MMOE

上线收益?

- 人均阅读时长 +1.3%
- 次留 +0.25%



(a) MMOE struct

1. Ma, Jiaqi, et al. "Modeling task relationships in multi-task learning with multi-gate mixture-of-experts." *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2018.

知乎

# Unbias LTR

## 训练数据的噪声：Position Bias

两个思路：

- A. 降低 TOP 位置的样本权重
- B. 对 Position Bias 进行建模

1. Zhao, Zhe, et al. "Recommending what video to watch next: a multitask ranking system." *Proceedings of the 13th ACM Conference on Recommender Systems*. 2019.



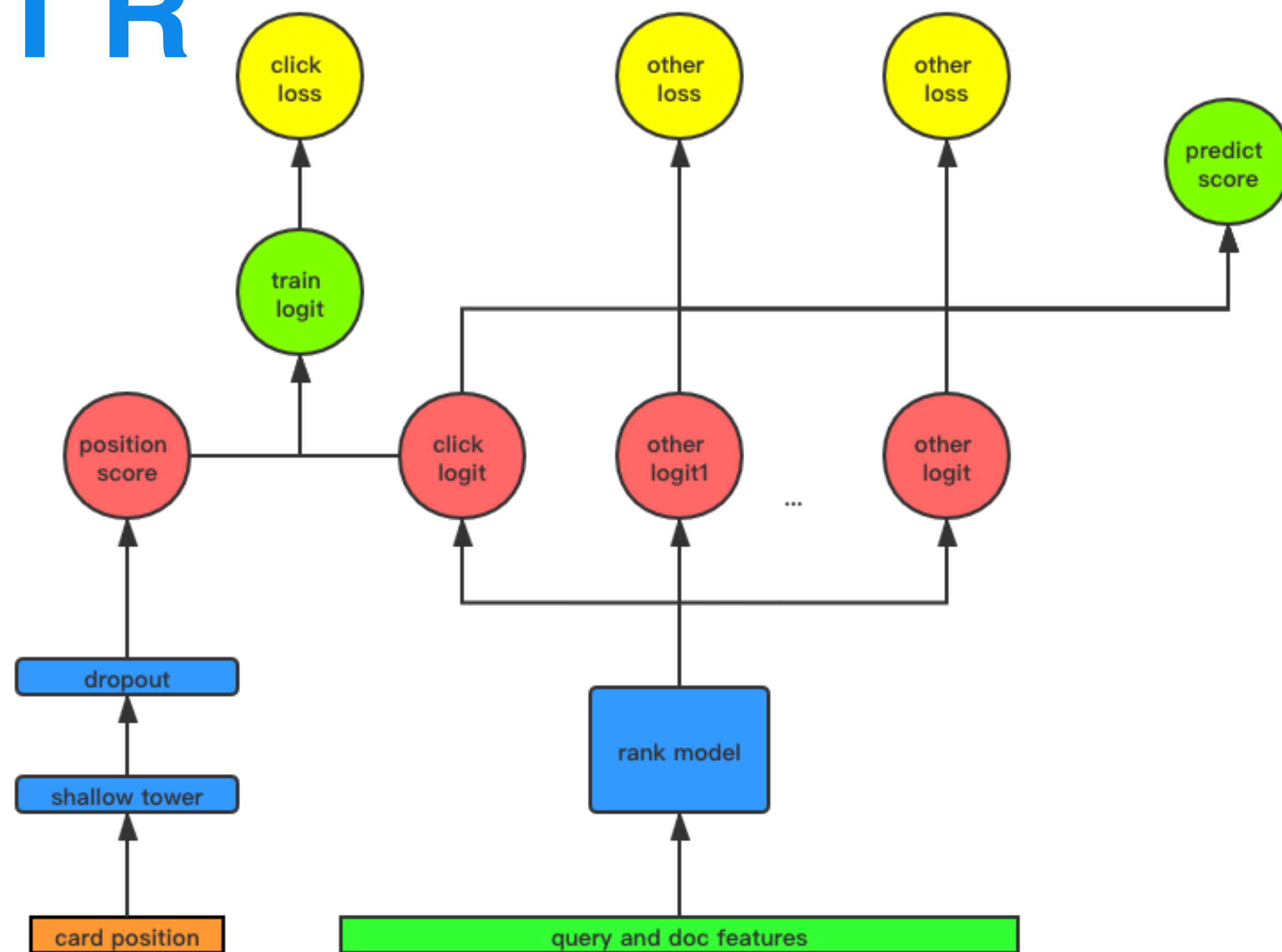
知乎

# Unbias LTR

通过 shallow tower 对  
Position Bias 进行建模

上线收益?

- 点击比 +0.6%
- 点展比 +0.2%



(a) Unbias LTR struct

用户实际点击 = 真实点击率 + Position Bias

1. Zhao, Zhe, et al. "Recommending what video to watch next: a multitask ranking system." *Proceedings of the 13th ACM Conference on Recommender Systems*. 2019.

知乎

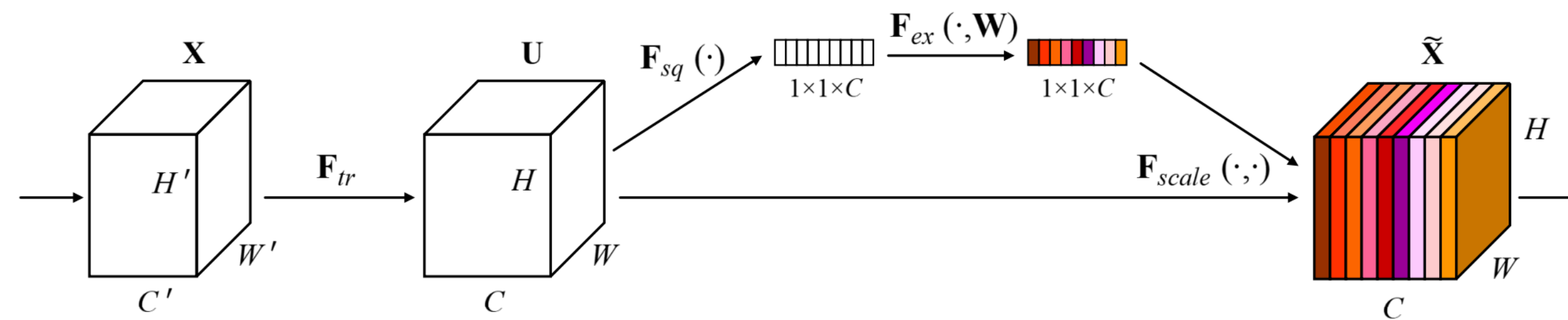
# Context Aware

Point wise or pair wise or list wise ?

- Loss 角度
  - Point wise loss
  - Pair wise loss
  - List wise loss

list wise > pair wise > point wise

- 模型 OR 特征角度?



(a) SE block

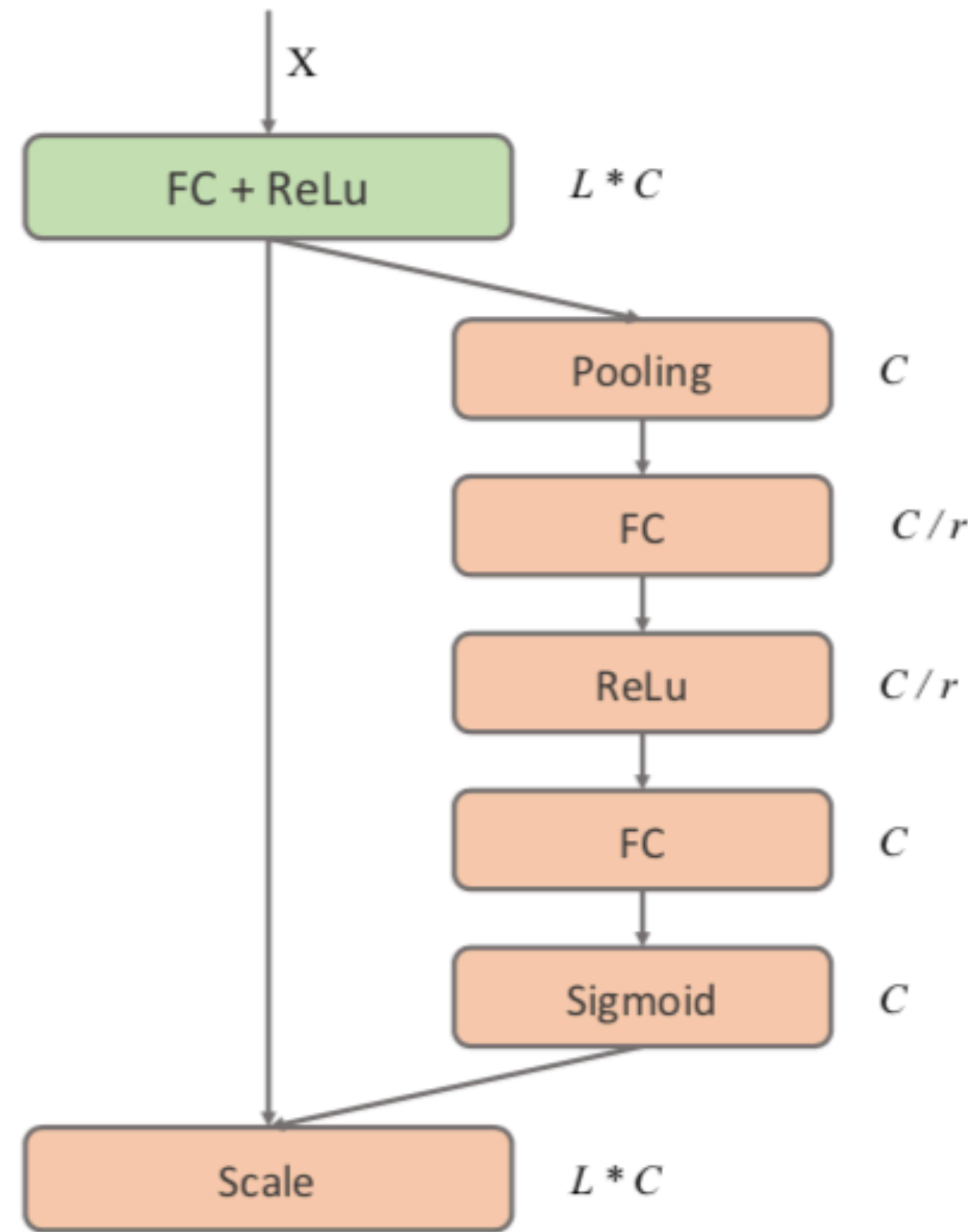
1. Hu, Jie, Li Shen, and Gang Sun. "Squeeze-and-excitation networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

知乎

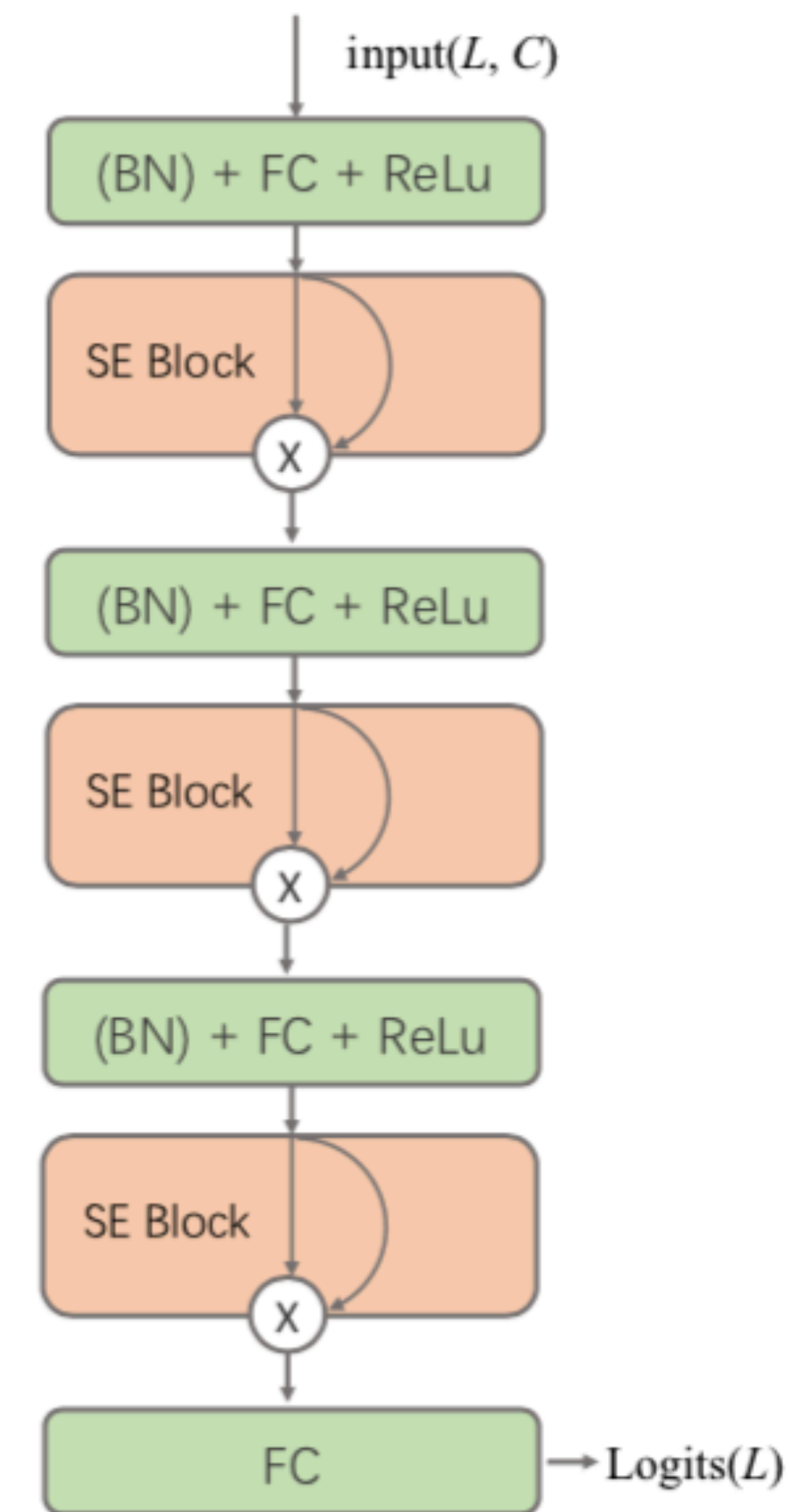


# Context Aware

- 将 List 整体输入到 SE Block
- List size 怎么选?



(a) SE Block struct



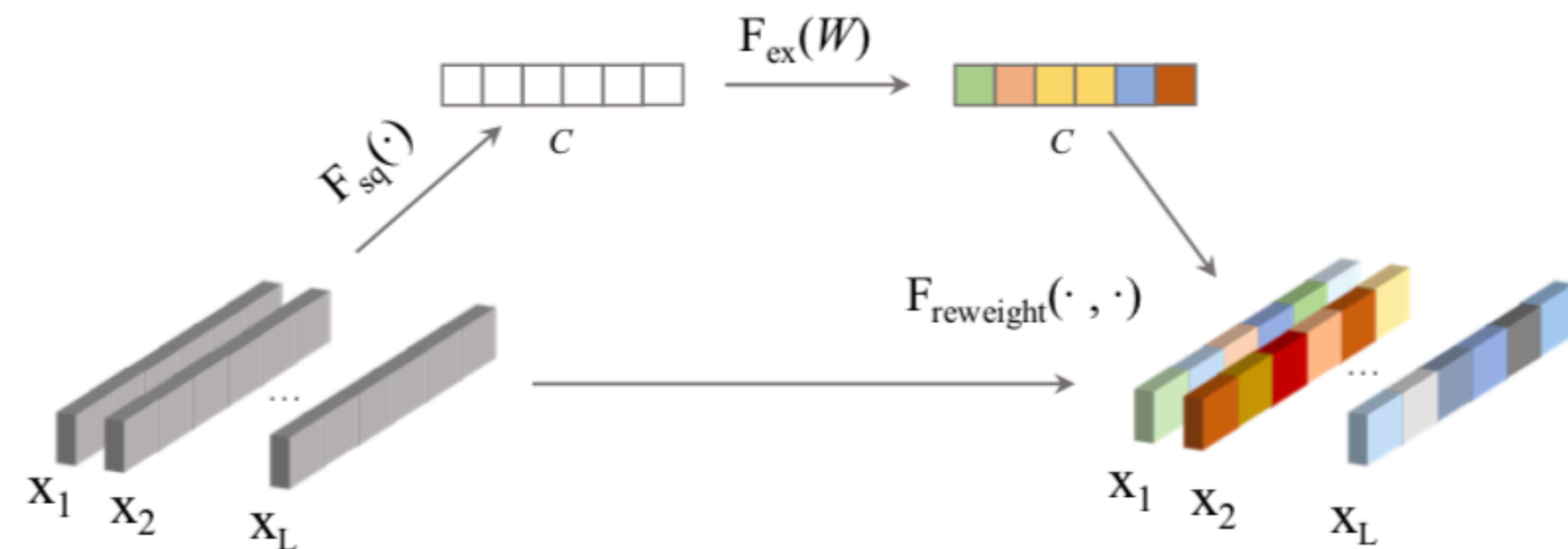
(b) Stack multi SE Block struct

1. Wang, RuiXing, et al. "SERank: Optimize Sequencewise Learning to Rank Using Squeeze-and-Excitation Network." arXiv preprint arXiv:2006.04084 (2020).

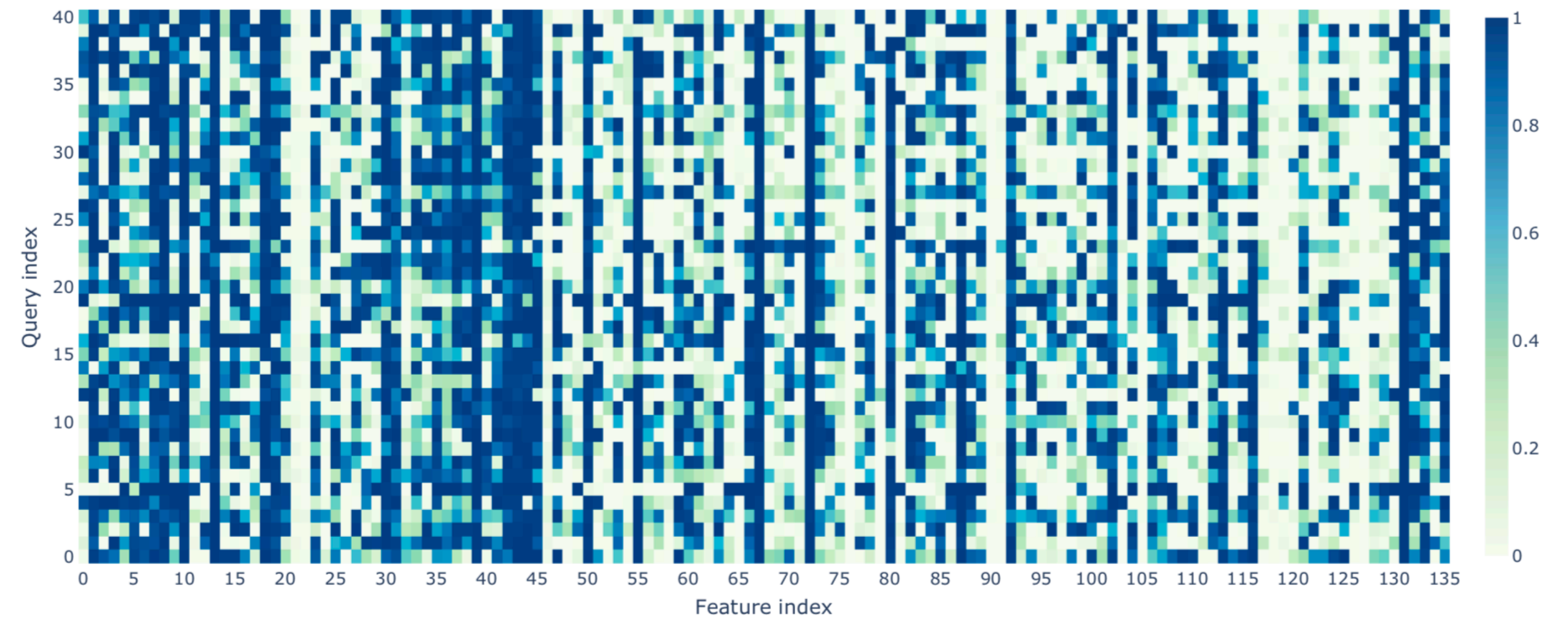
知乎

# Context Aware

- 学习特征在不同 Context 下的特征重要性



(a) SE Block explanation



(b) Feature importance on different context

- Wang, RuiXing, et al. "SERank: Optimize Sequencewise Learning to Rank Using Squeeze-and-Excitation Network." arXiv preprint arXiv:2006.04084 (2020).

知乎



# Context Aware

上线收益?

- TOP1 点击比 +0.9%
- TOP3 点击比 +0.33%
- TOP5 点击比 +0.1%

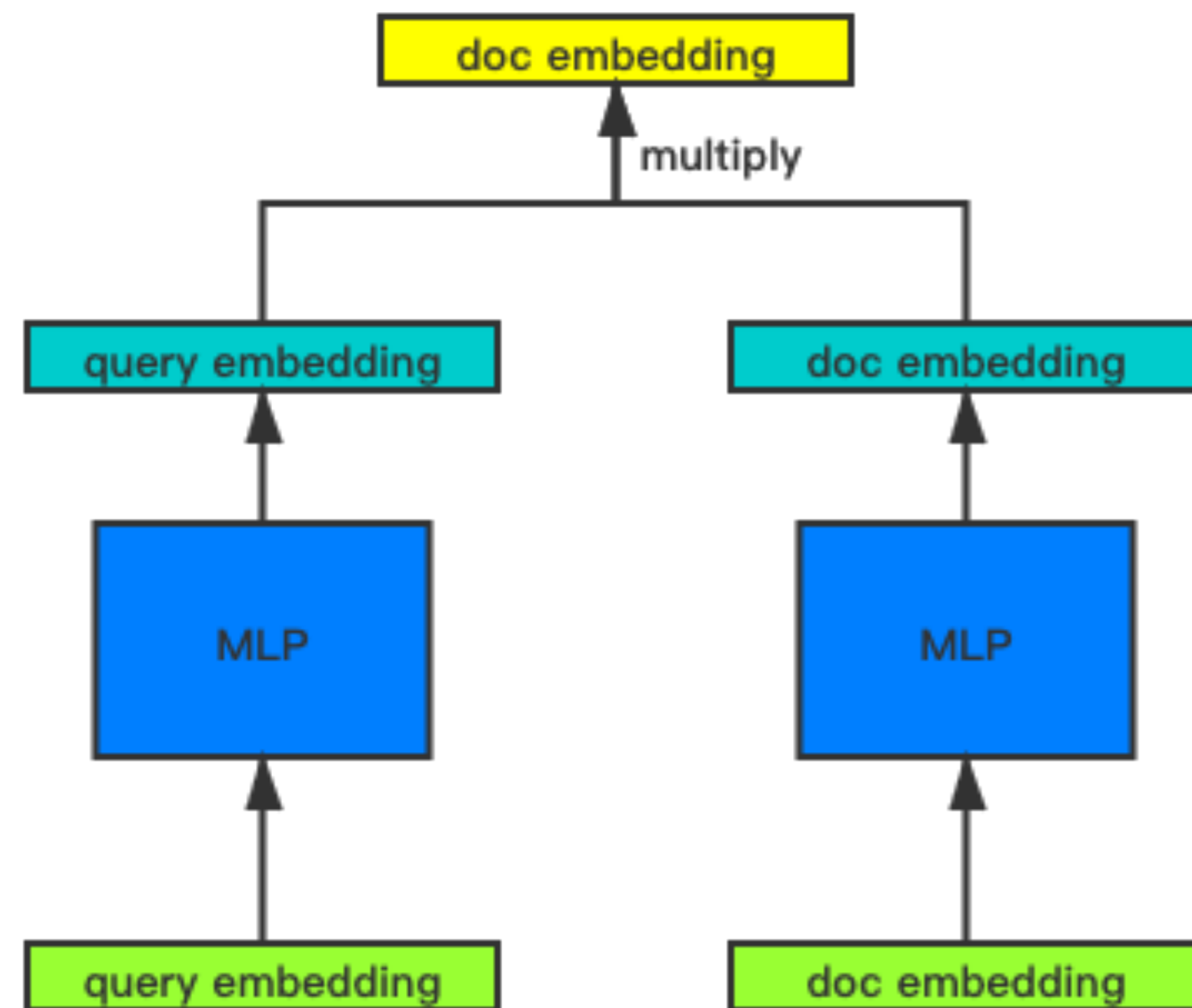
知乎

# End2End

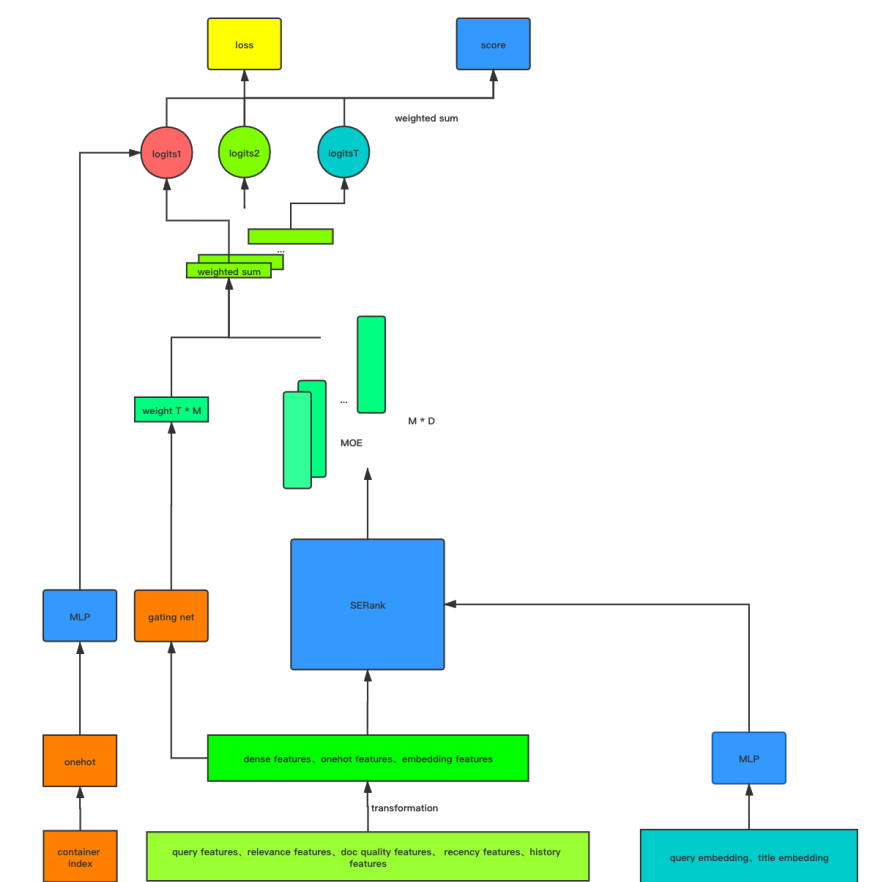
- 将 query、doc 的文本作为特征?
- BERT 能不能 end2end 的训练?

上线收益?

- 点击比 +0.15%
- CTR +0.8%



(a) BERT end2end struct



(b) ranking model

知乎

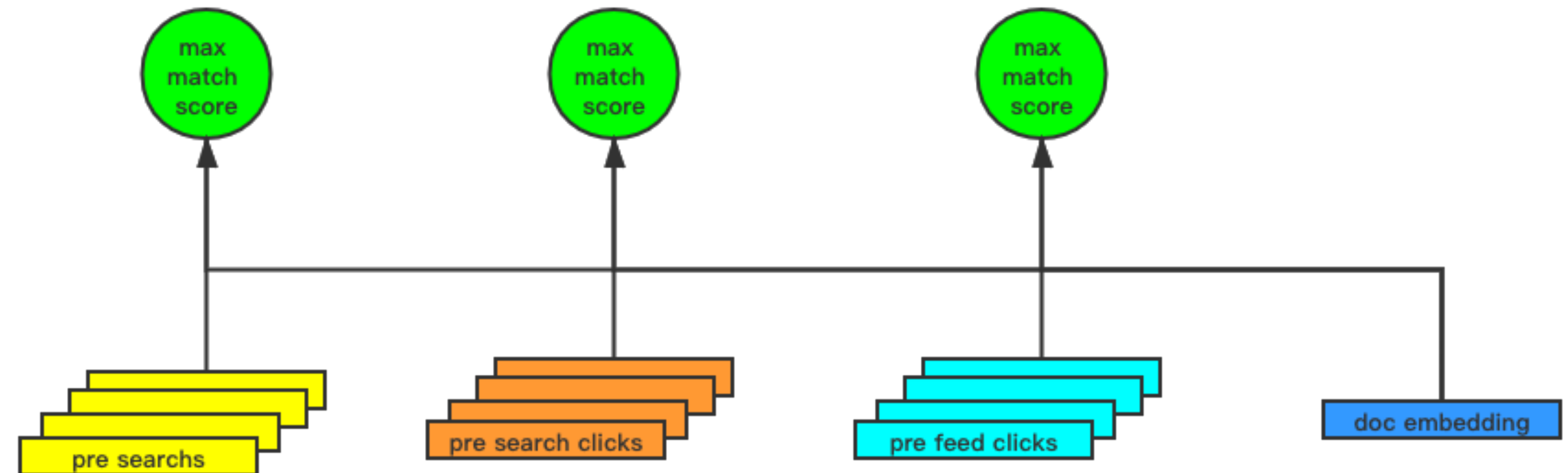
# Session Aware

如何利用用户的近期交互信息？

- 近期搜索 query
- 近期搜索点击 doc
- 近期首页推荐点击 doc

上线收益？

- 点击比 +0.49%
- 点展比 +0.62%

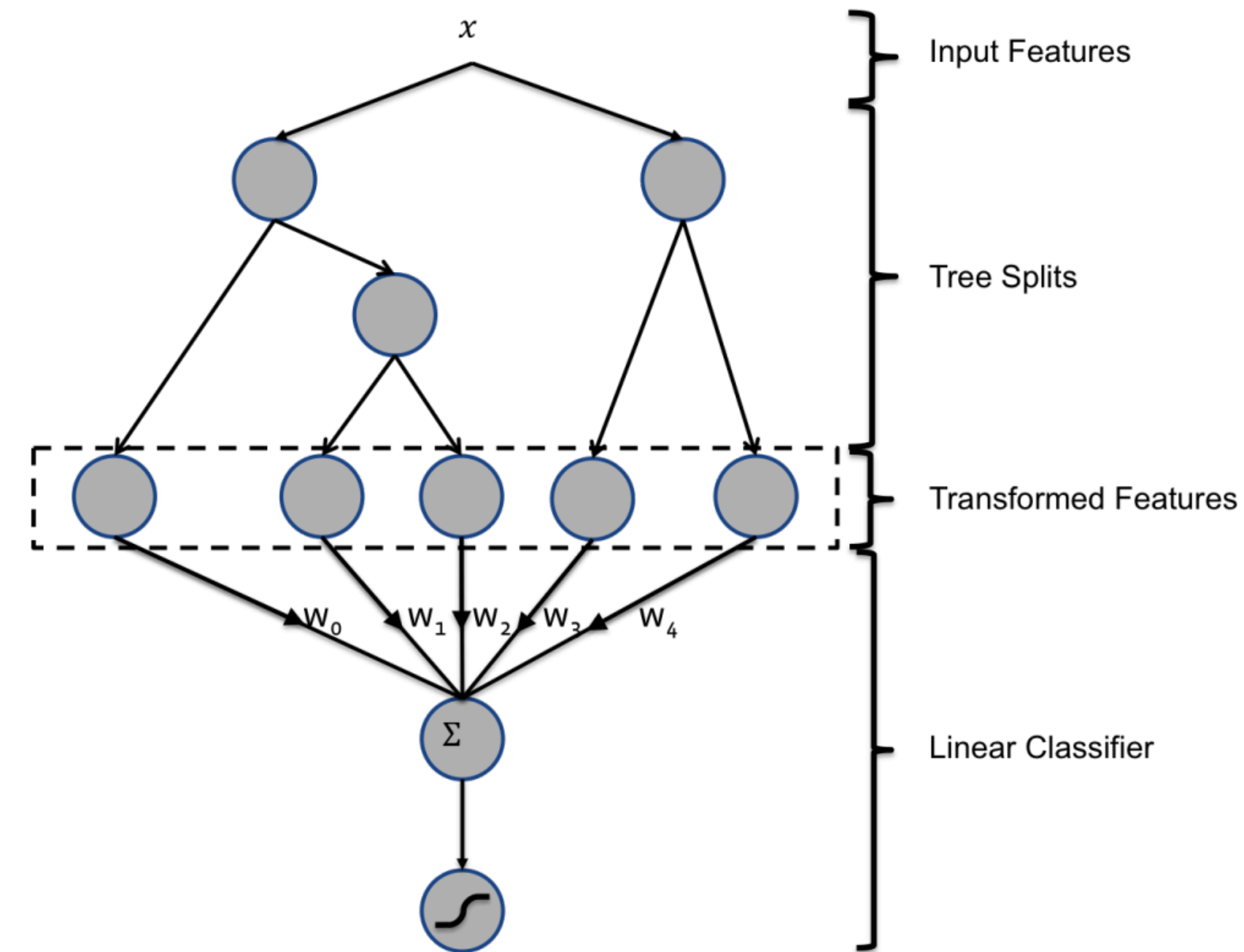


(a) Session aware struct

知乎

# GBDT 特征编码 - 未上线

1. GBDT 模型自动进行特征分桶和特征组合
2. 将 GBDT 输出的叶子节点编号作为 category 特征喂给 NN 模型



(b) GBDT encode

知乎

1. He, Xinran, et al. "Practical lessons from predicting clicks on ads at facebook." Proceedings of the Eighth International Workshop on Data Mining for Online Advertising. 2014.

# IRGAN — 未上线

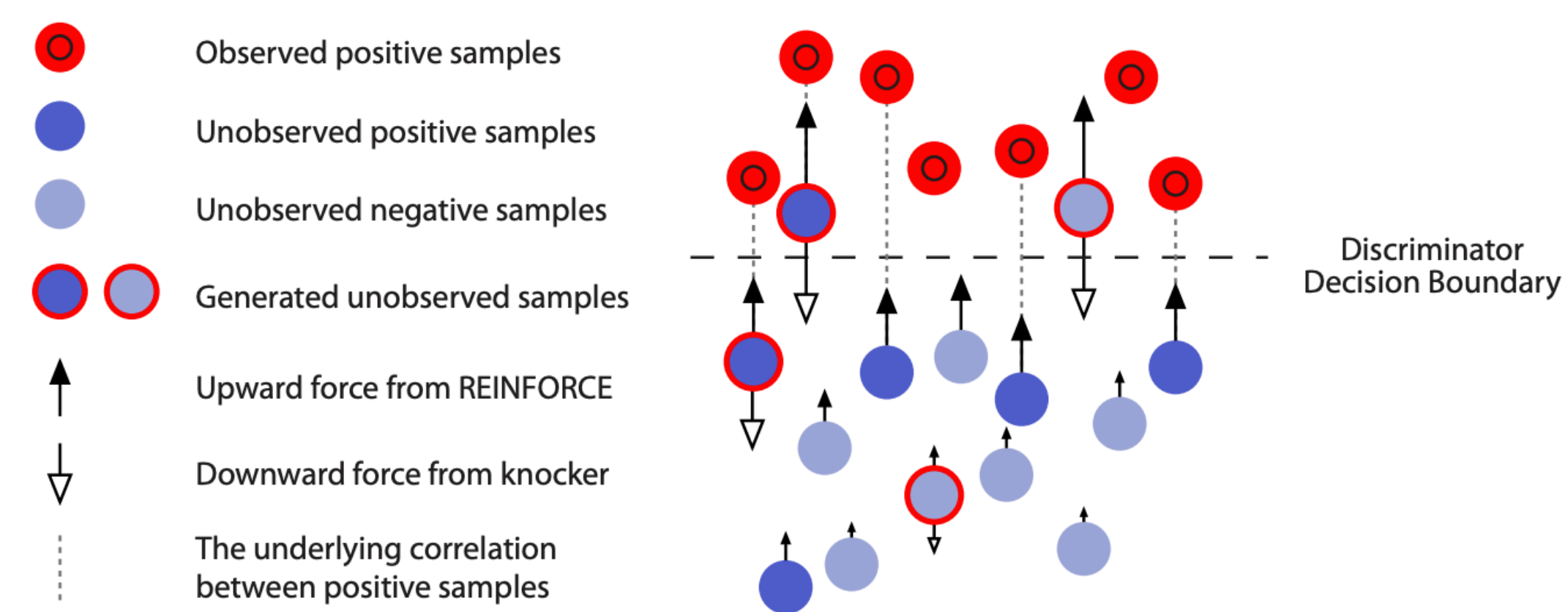
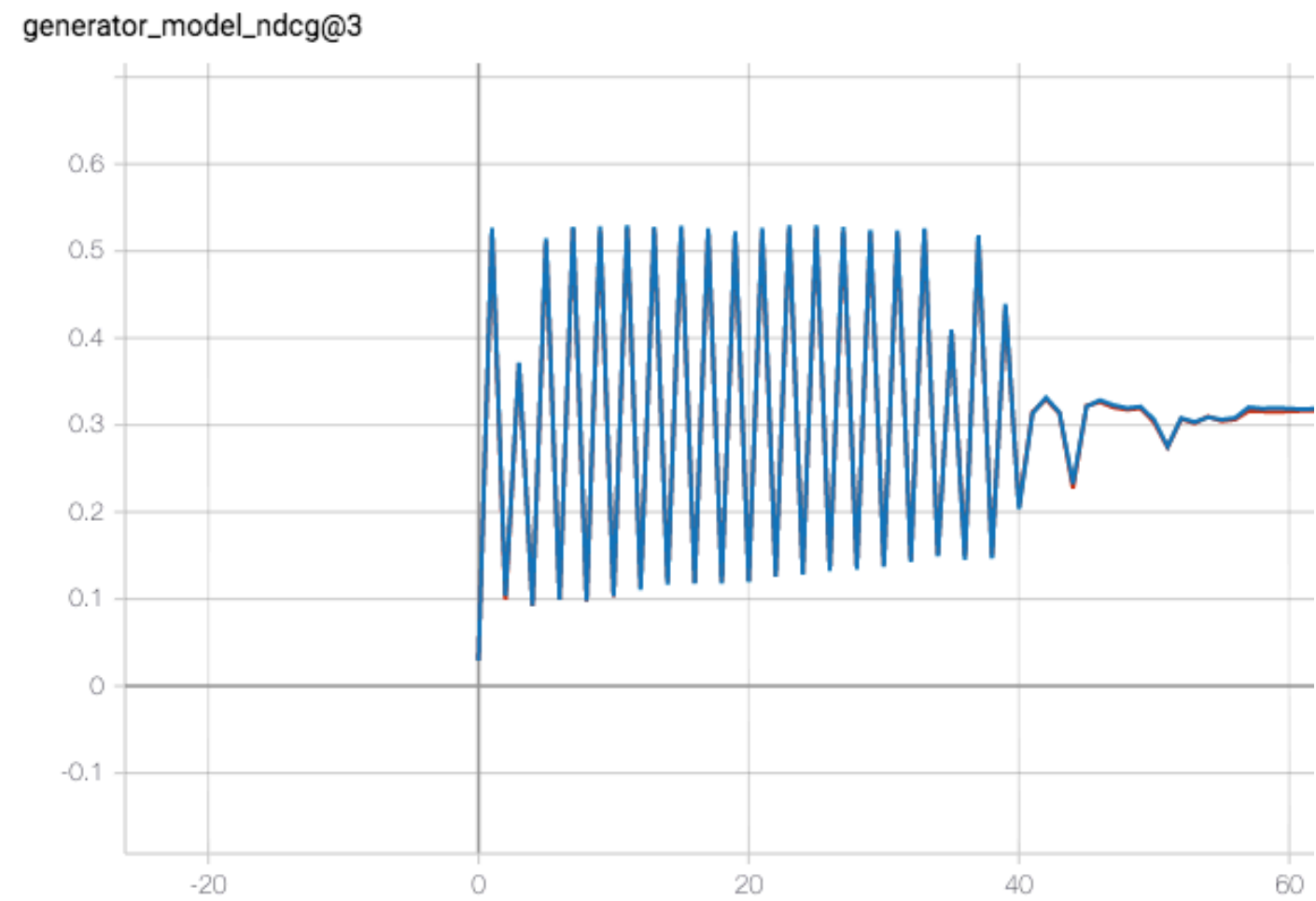
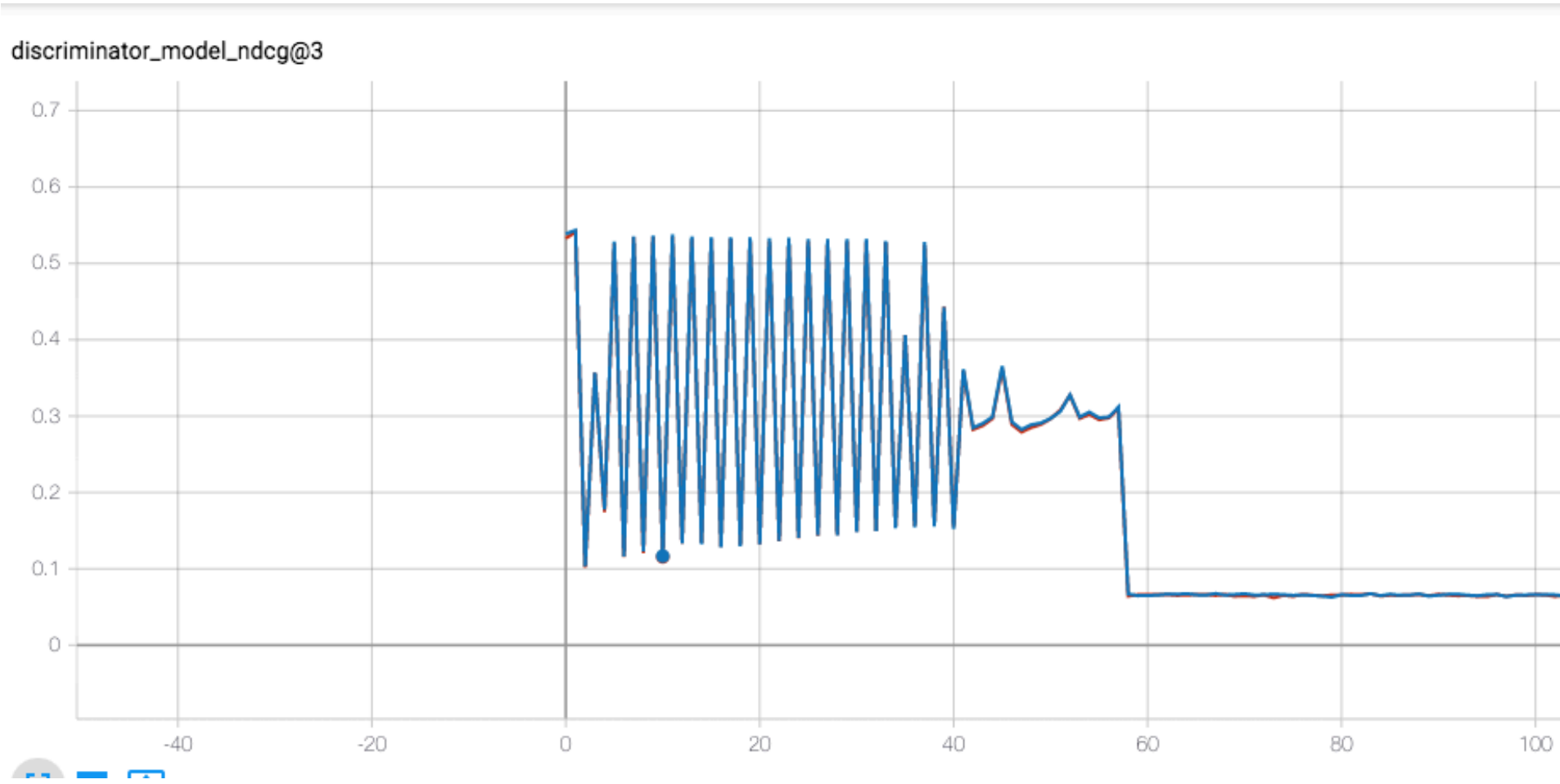


Figure 1: An illustration of IRGAN training.



知乎

1. Wang, Jun, et al. "Irgan: A minimax game for unifying generative and discriminative information retrieval models." *Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval*. 2017.

# 未来方向

- Online learning
- Graph embedding
- 个性化
- 边缘计算

知乎

THANKS



# 知乎

发现更大的世界