



腾讯算法大赛

团队介绍—SYSU_至九

队长：郭达雅 队员：张俊鸿 刘昕

- 中山大学2018届本科毕业生
- 第一届腾讯社交广告高校算法大赛top10



目录

1

赛题分析

2

特征工程

3

模型设计

4

性能优化

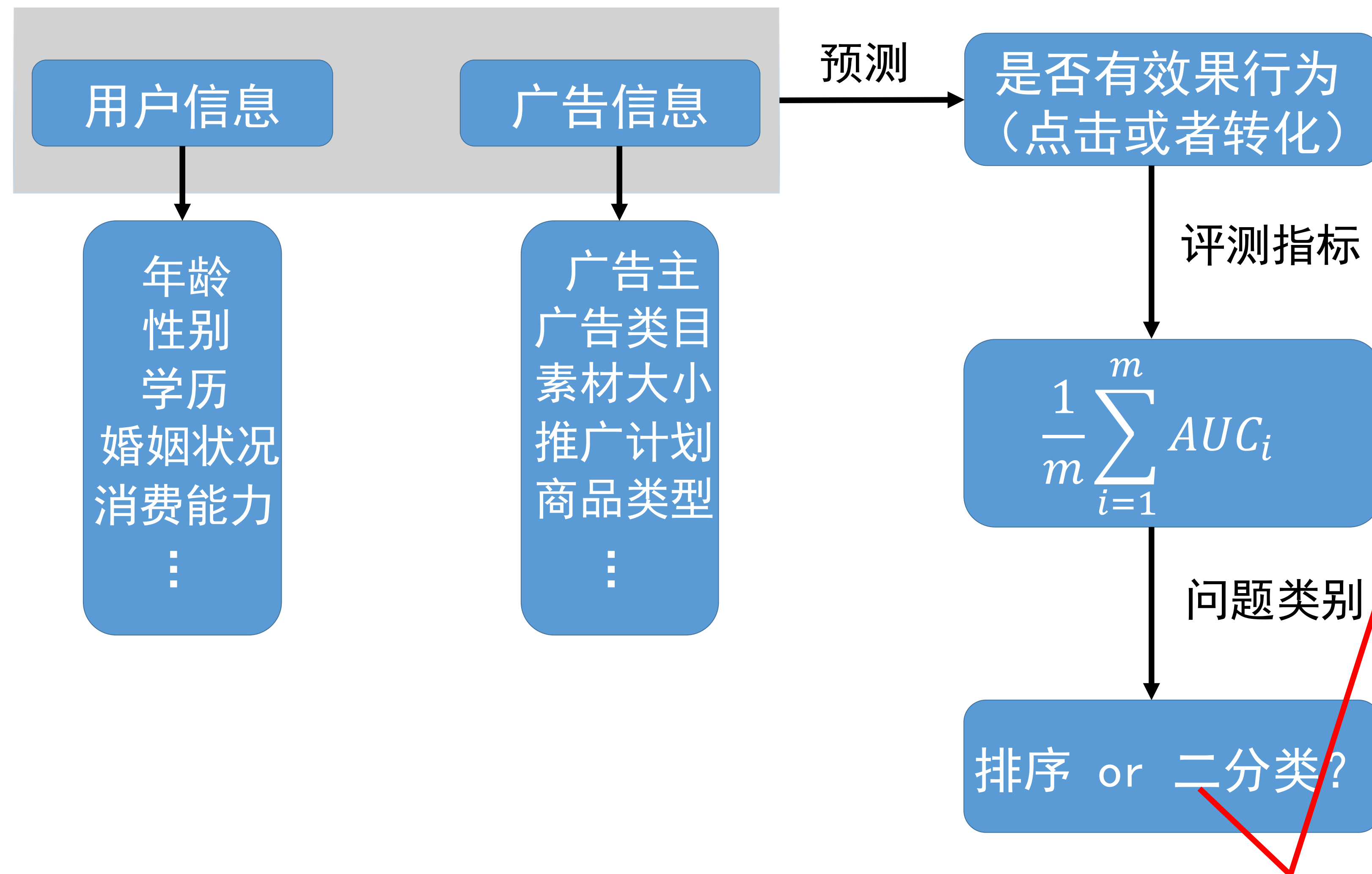


赛题分析

1



赛题分析



2

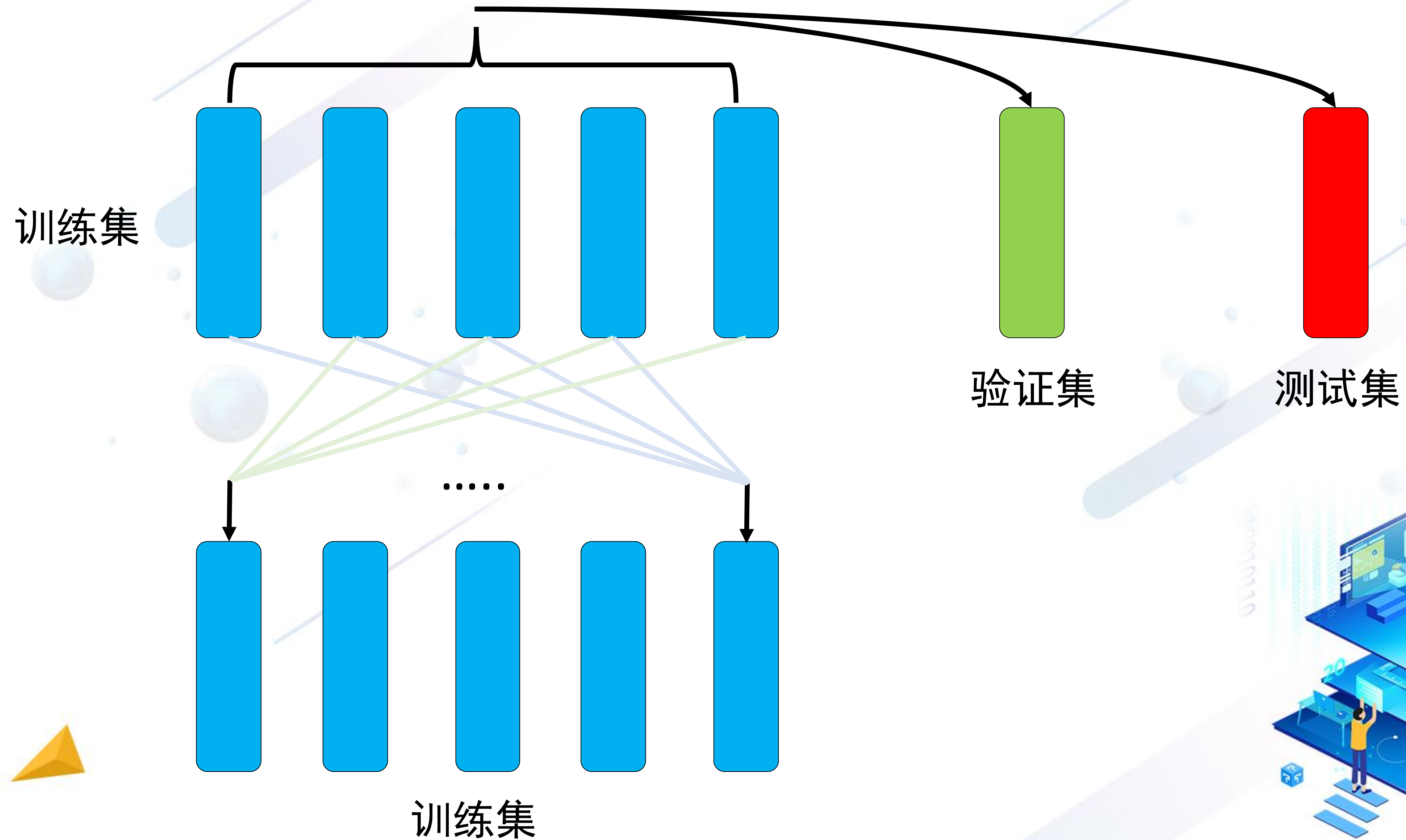


特征工程

- 基本特征：31个ID特征
 - 15个单值特征，16个多值特征
 - 单值特征如用户的年龄性别，广告的素材ID和商品ID等
 - 多值特征如用户感兴趣的关键词，近期安装的app等
- 计数特征：11个特征
 - 如近期安装的app总数，用户感兴趣的关键词总数
- 统计特征：17个特征
 - 利用K-fold统计方法对用户ID构造的特征
 - 用户ID出现的次数和正负样本数（3个）
 - 用户ID与不同广告特性的交互特征共同出现的正负样本数（14个）
- 总特征：31+11+17=59个



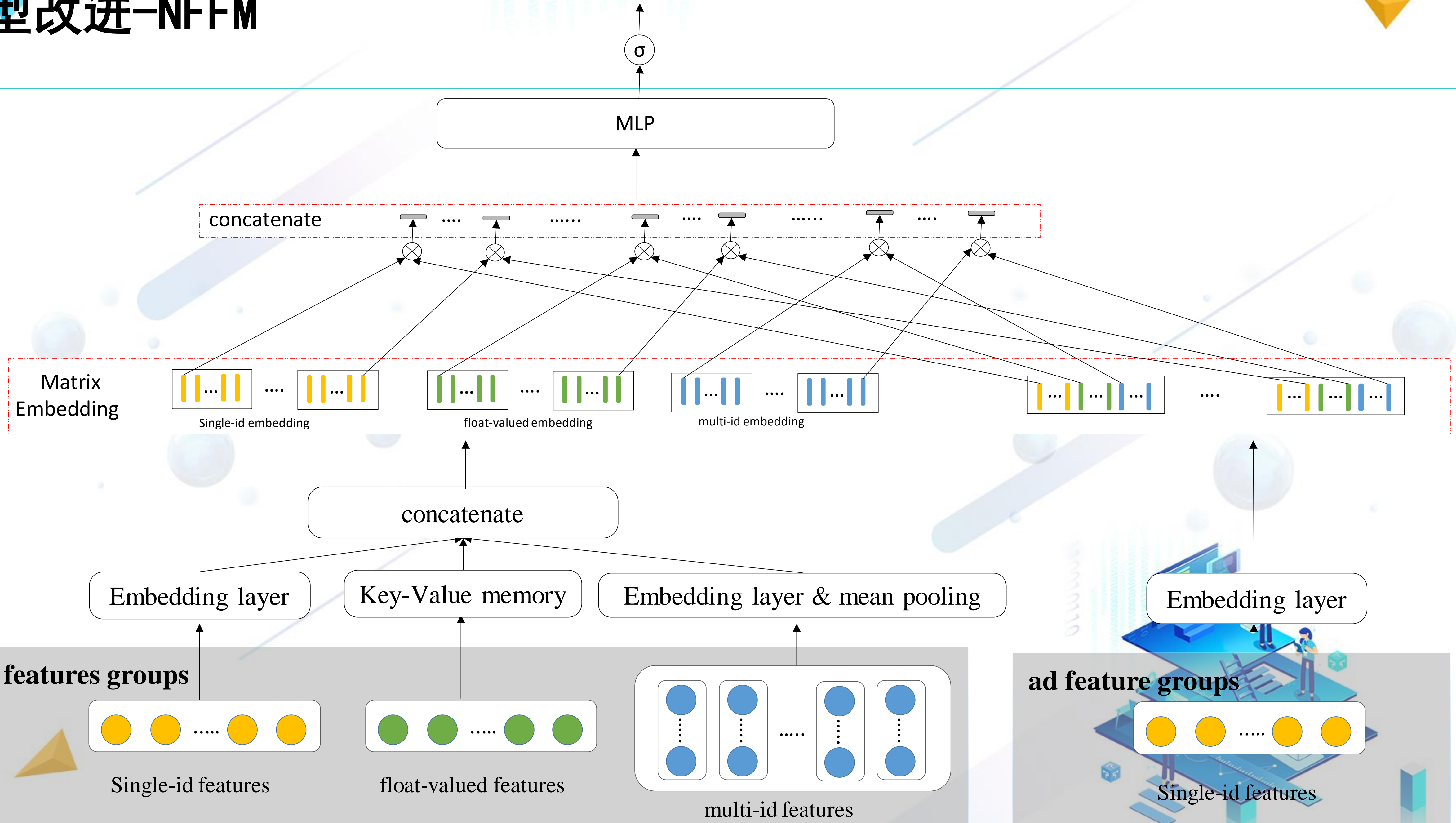
K-fold统计



3

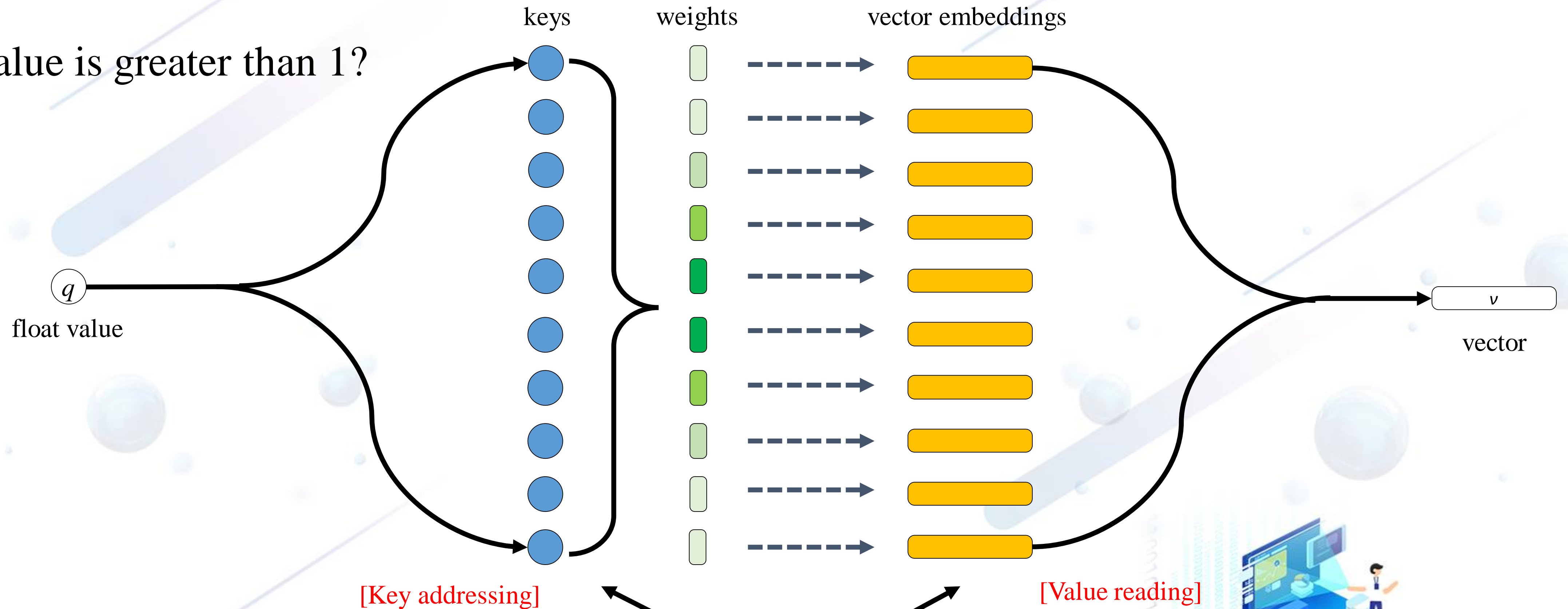


模型改进-NFFM



Key-Value Memory

Q: float value is greater than 1?



Parameter: $N=100$ $k_i = \frac{i}{N}$

Key addressing: $w_i = \text{softmax}(\frac{1}{|q - k_i| + e^{-15}})$

Value reading: $v = \sum_{i=1}^N w_i v_i$

(k_0, v_0) (k_1, v_1) (k_2, v_2) (k_3, v_3) (k_N, v_N)

Key-Value Memories



分析与结果

- 将特征分成用户和广告两个特征组
 - 有效减少空间消耗，提高运算速度
 - 相较于传统的NFFM，性能有所提高
- Key-Value Memory
 - 一种将浮点数转成为空间向量的方法
 - 相较于简单的数值 \times 向量，KVMem具有非线性特点
 - 相较于分桶，相邻的向量具有相关性
- 模型结果

模型	AUC	Rank
NFFM	0.760	38
NFFM + features	0.770	12
NFFM + features + ensemble	0.7734	9
NFFM + Xdeepfm+ features + ensemble	0.7748	7

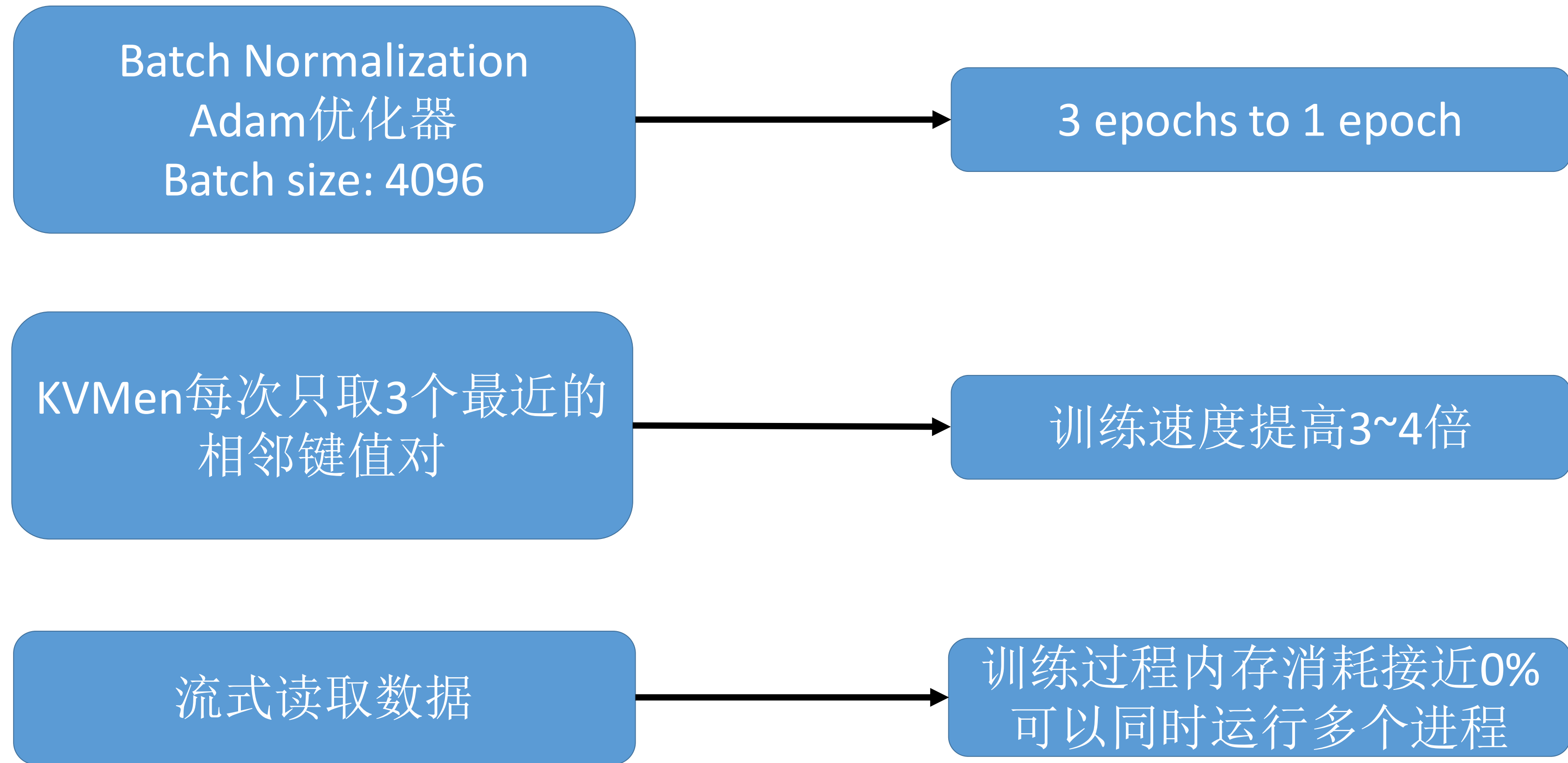


性能优化

4



性能优化



参考文献

- [1] He, Xiangnan, and Tat-Seng Chua. "Neural factorization machines for sparse predictive analytics." *Proceedings of the 40th International ACM SIGIR conference on Research and Development in Information Retrieval*. ACM, 2017.
- [2] Lian, Jianxun, et al. "xDeepFM: Combining Explicit and Implicit Feature Interactions for Recommender Systems." *arXiv preprint arXiv:1803.05170* (2018).
- [3] Miller, Alexander, et al. "Key-value memory networks for directly reading documents." *arXiv preprint arXiv:1606.03126* (2016).



Thanks