

微醺

Contents

目录

1 团队介绍

2 赛题理解

3 特征工程

4 模型介绍

5 总结与思考



团队介绍



刘冀
南京大学
硕士二年级



周宏
算法从业者
数据竞赛爱好者



王志昊
哈尔滨工业大学
硕士二年级



赛题理解

赛题理解

任务

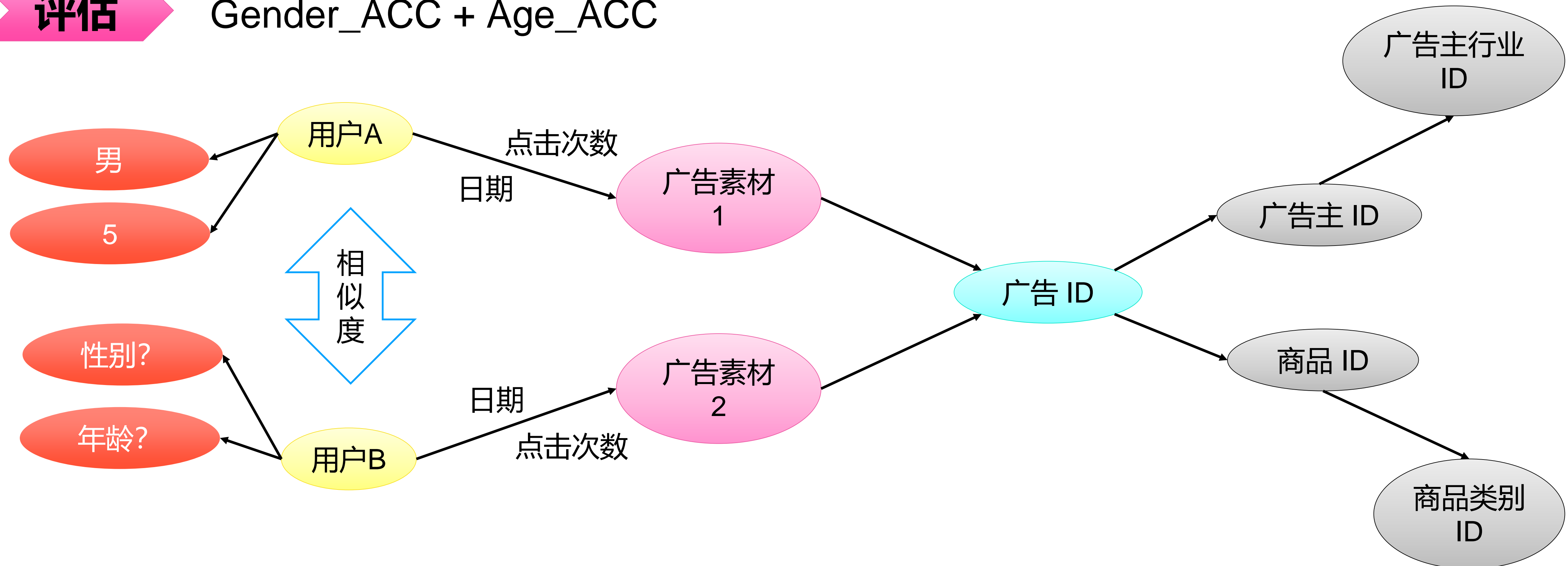
根据用户在广告系统中的 **交互行为** 预测用户的 **年龄 (10)**、**性别 (2)**

数据

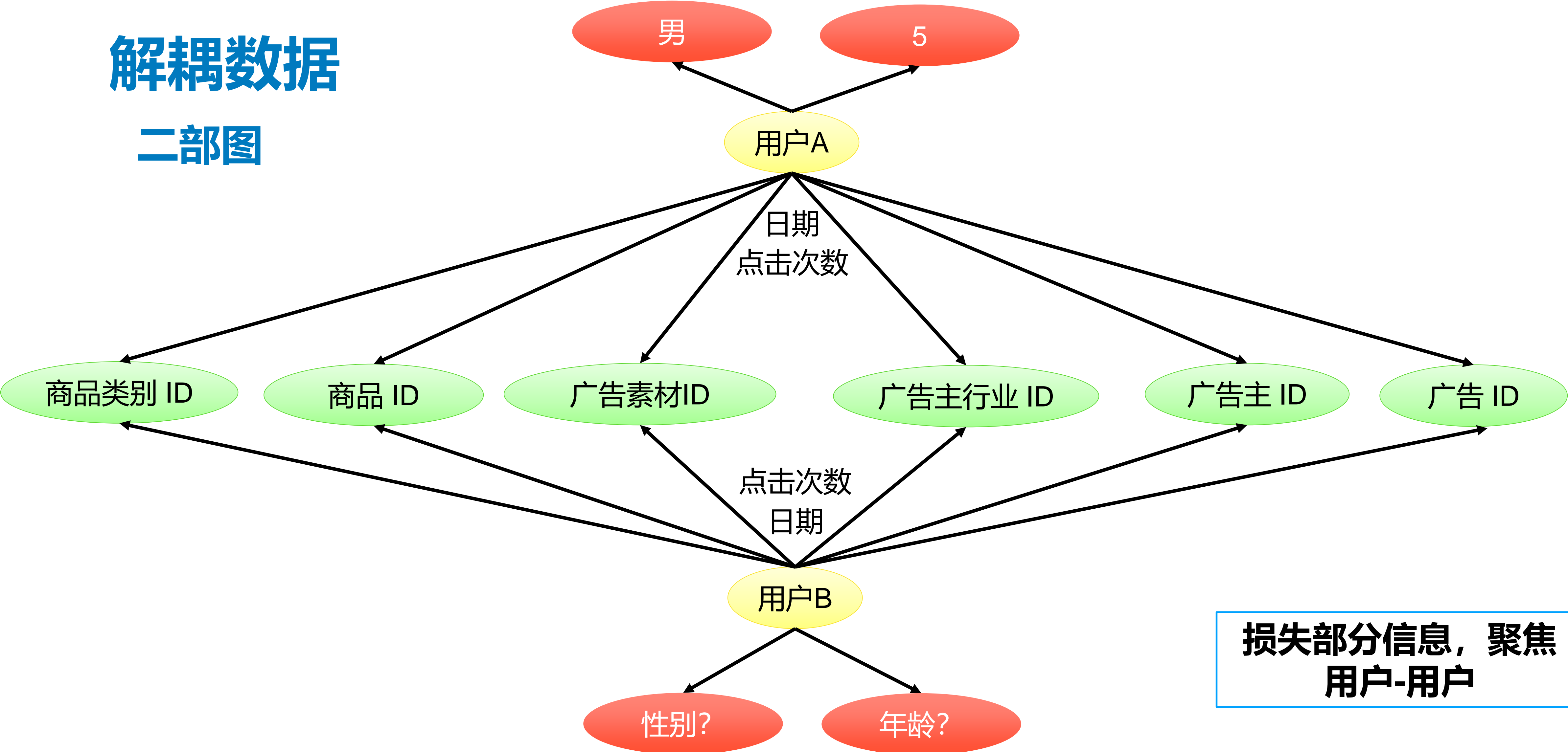
- 用户在91 天时间内的广告点击记录
- 被点击的广告的信息

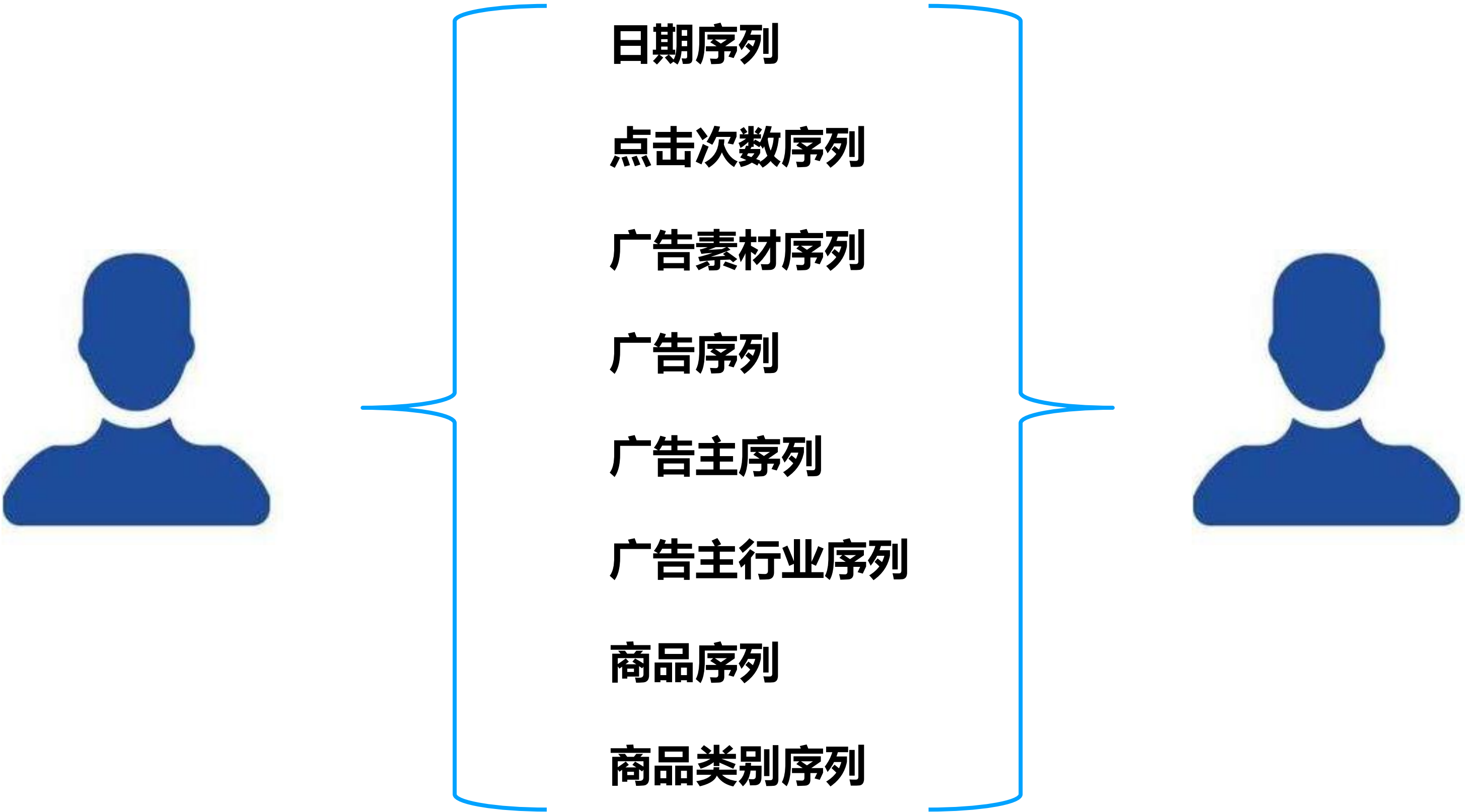
评估

Gender_ACC + Age_ACC



解耦数据
二部图

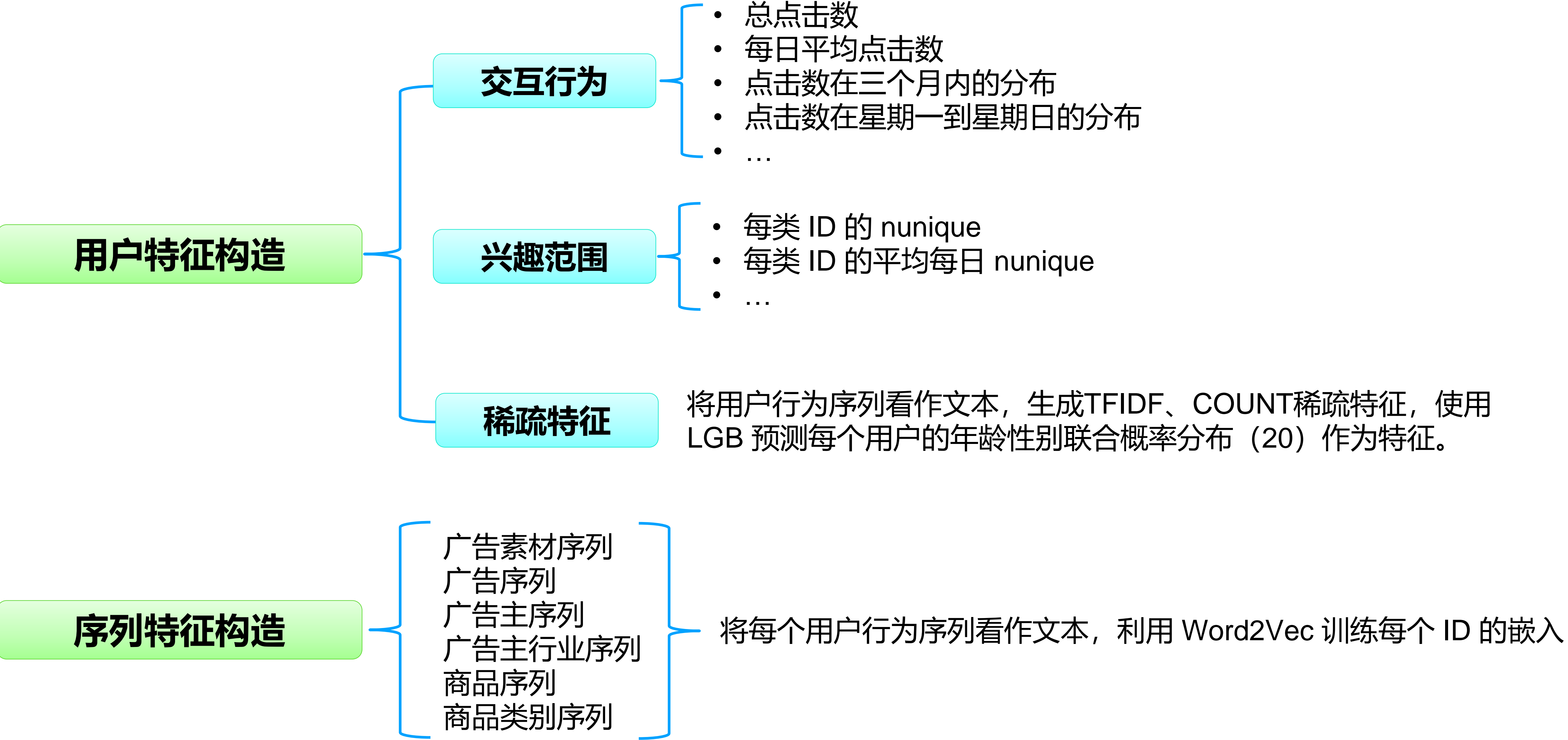




将用户交互日志记录解耦并转换为用户行为序列



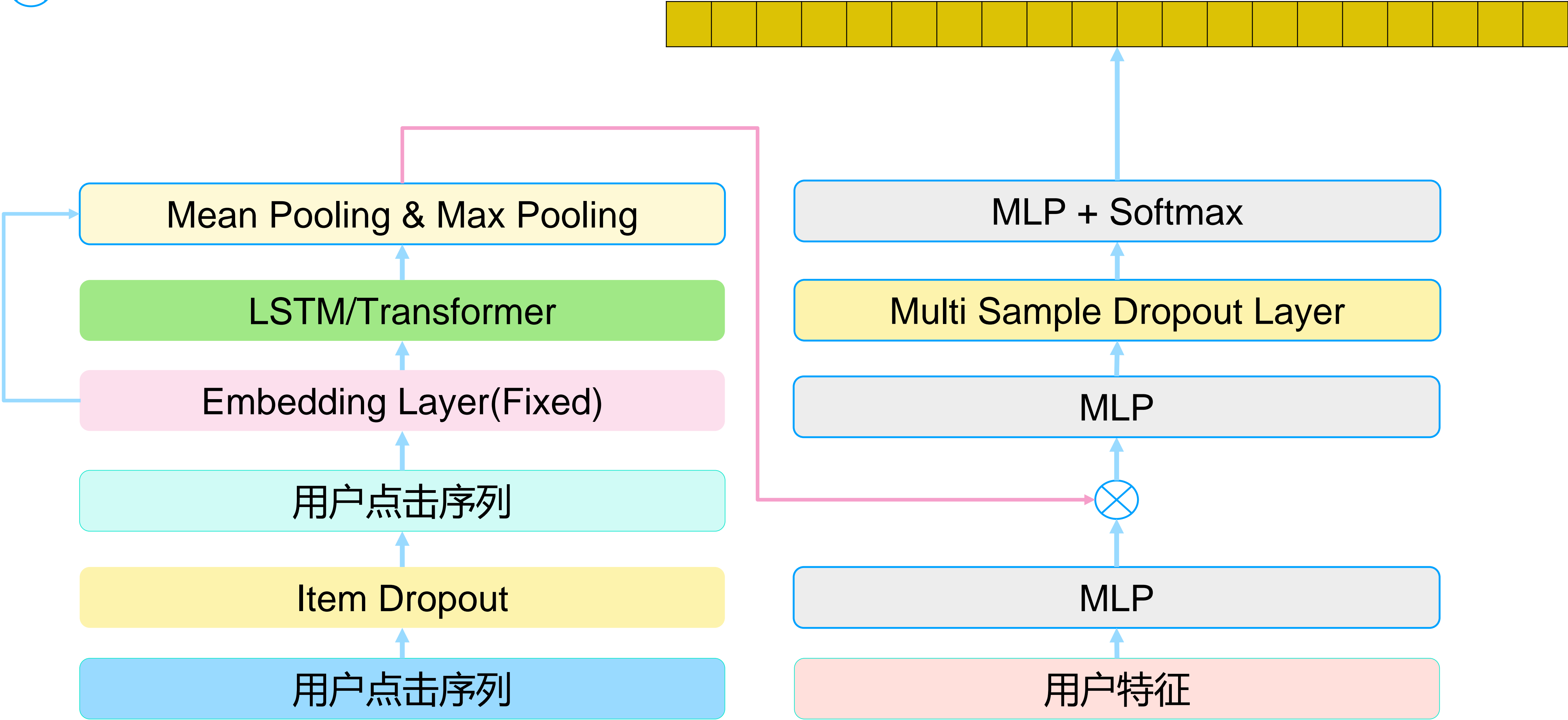
特征工程





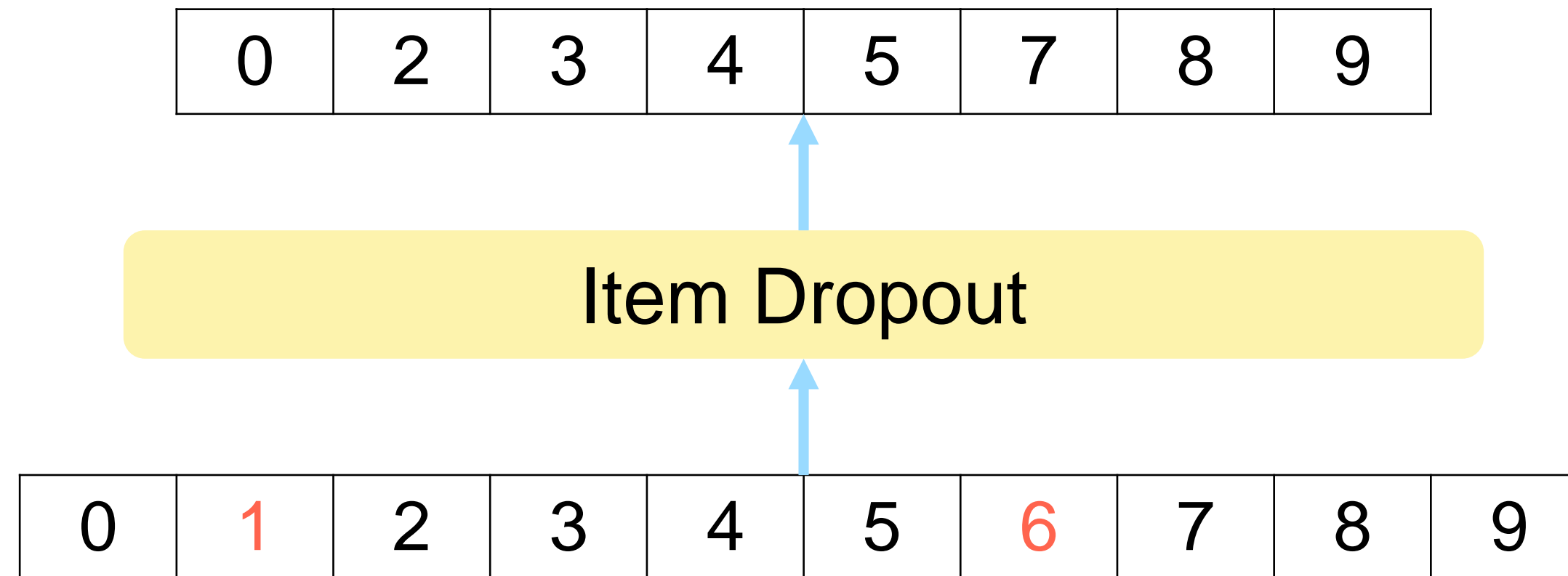
模型介绍

⊗ Concatenate

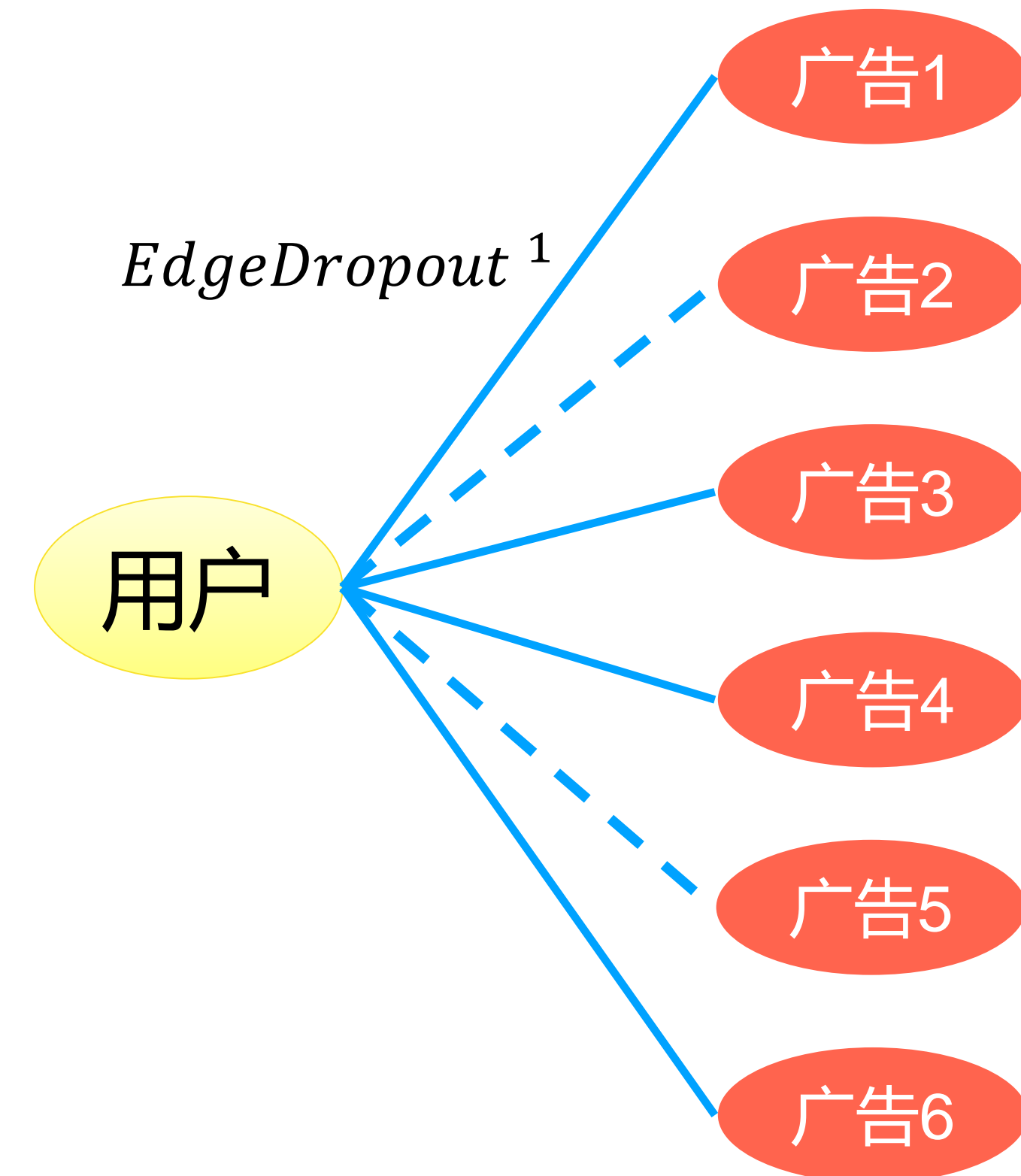


模型介绍

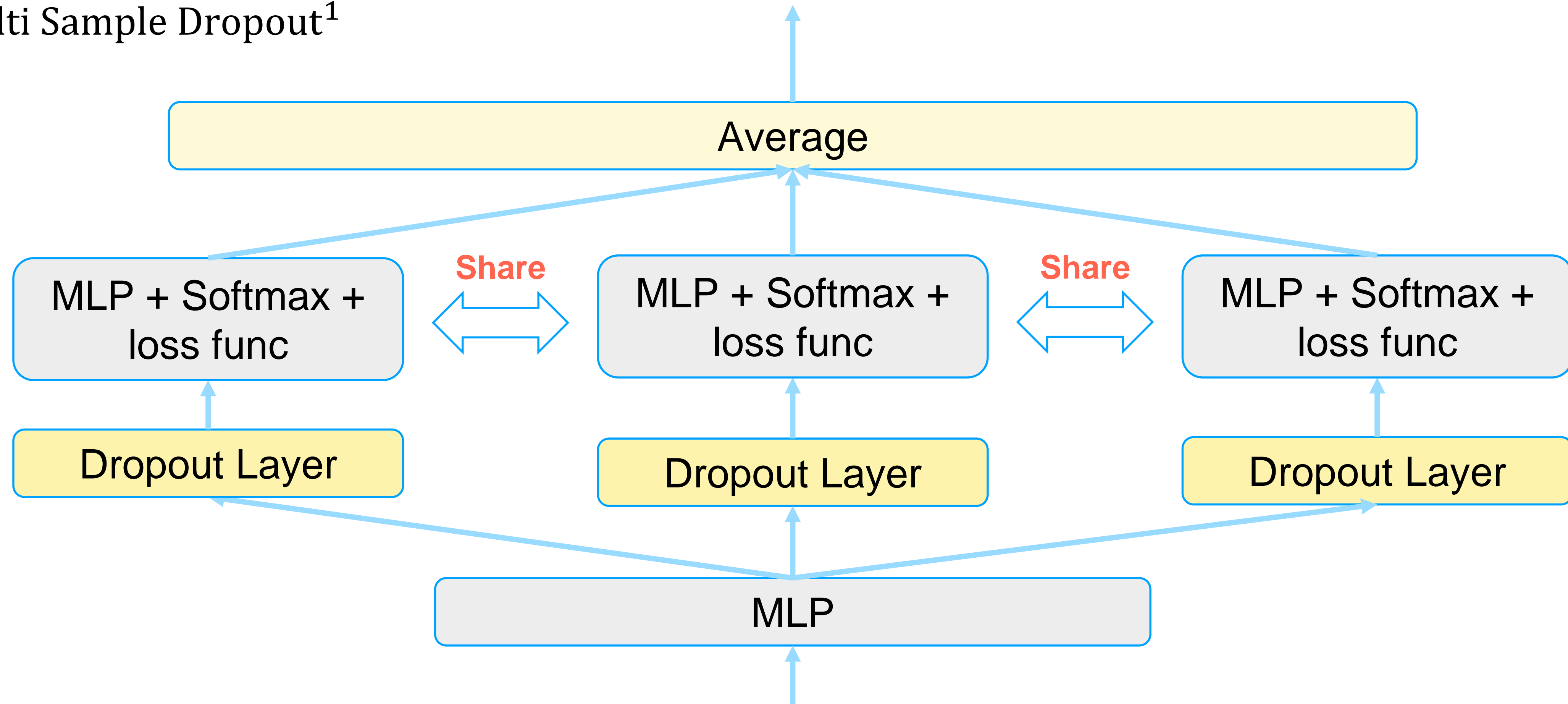
增加样本多样性，减轻过拟合现象
Dropout 的概率值较为敏感



预测阶段，仍然可以使用 Item Dropout 来增加样本多样性，产生多个 logits，以均值作为最终的预测结果



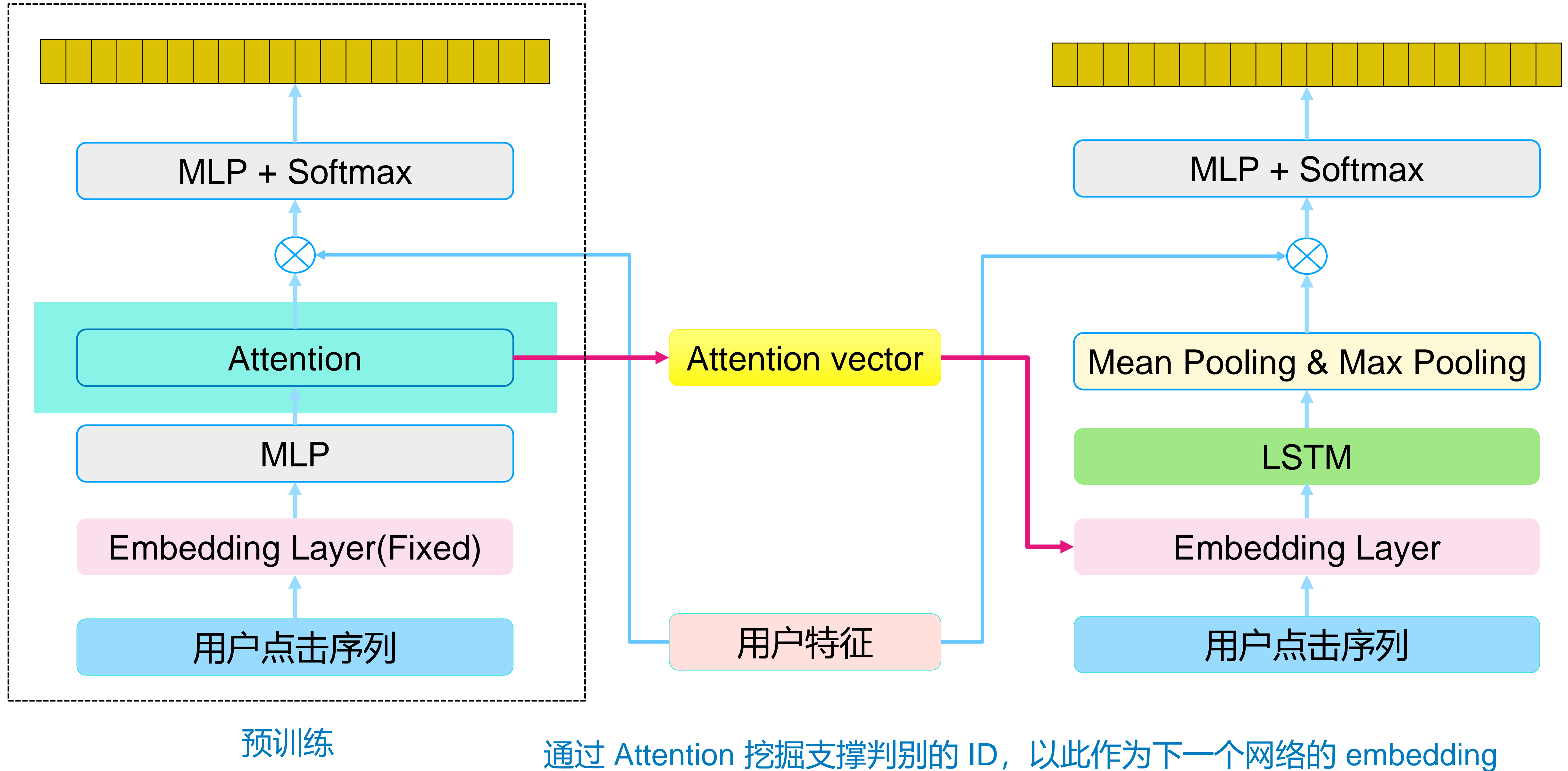
1. Rong Y , Huang W , Xu T , et al. DropEdge: Towards Deep Graph Convolutional Networks on Node Classification[J]. 2019.
2. <https://towardsdatascience.com/transformers-are-graph-neural-networks-bca9f75412aa>

Multi Sample Dropout¹

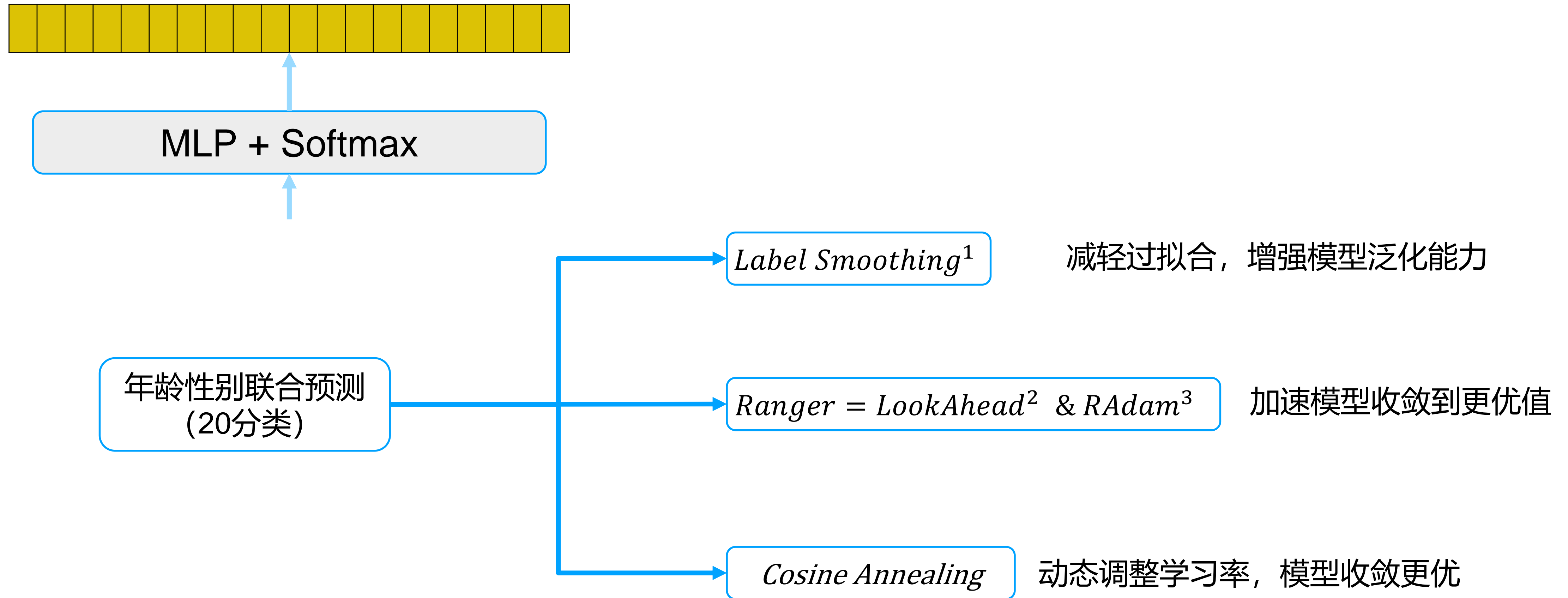
增加 Mini-batch 内多样性，在不显著增加计算成本情况下，加速模型收敛，提高模型泛化能力

1. Inoue H . Multi-Sample Dropout for Accelerated Training and Better Generalization[J]. 2019.

模型介绍-支撑ID预挖掘网络



模型介绍-模型训练

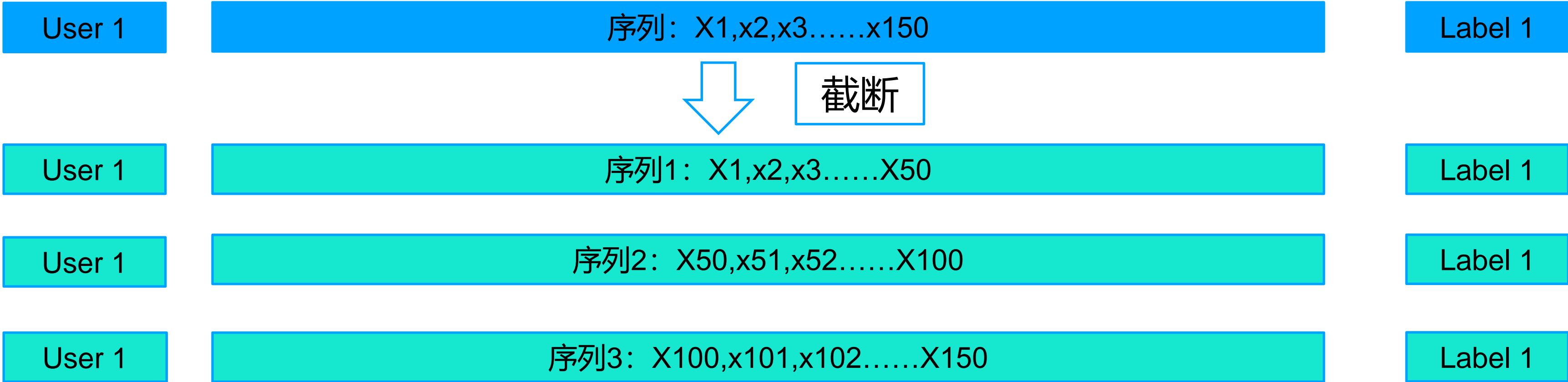


1. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016 pp. 2818-2826.

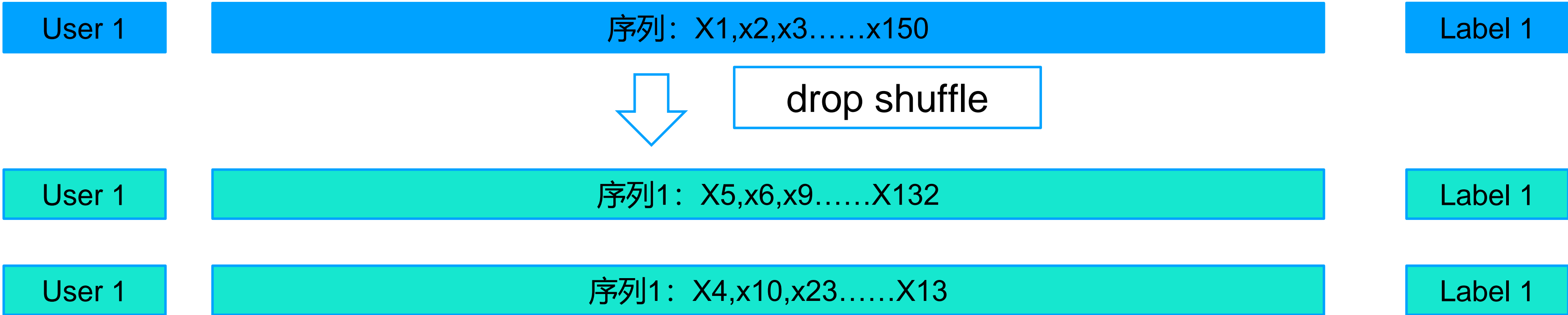
2. <https://arxiv.org/abs/1907.08610>

3. <https://arxiv.org/abs/1908.03265>

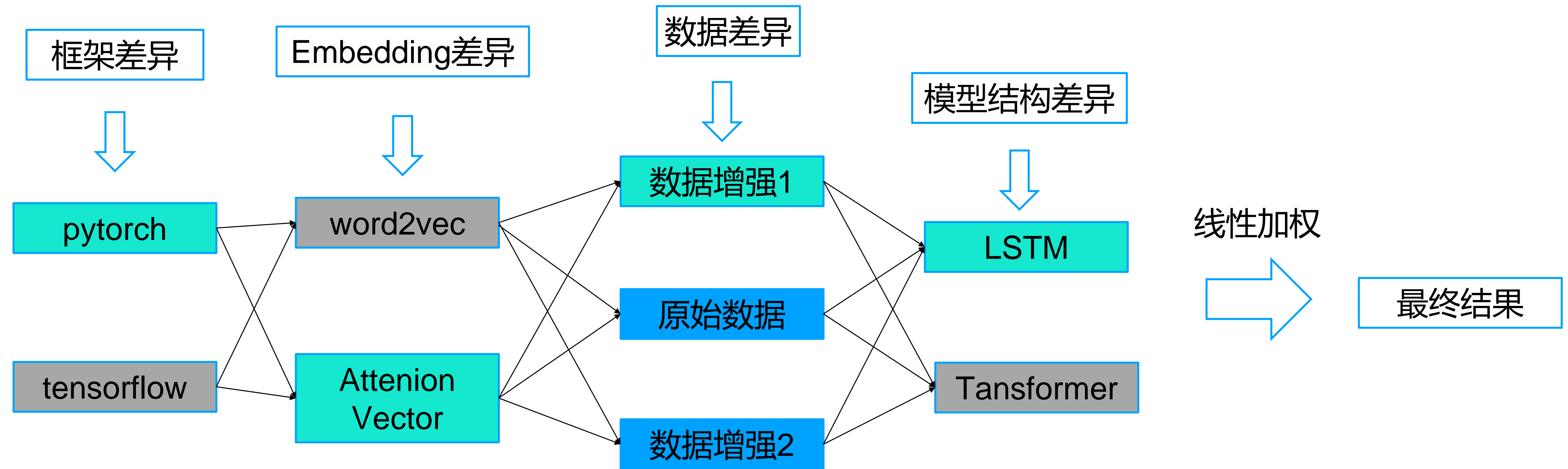
方案一：将一个用户的序列进行截断，分成多序列，增加样本数量，同时预测时候可以对一个用户预测多次，加权平均



方案二：将序列进行乱序，drop，再拼接原来数据



模型介绍-模型融合



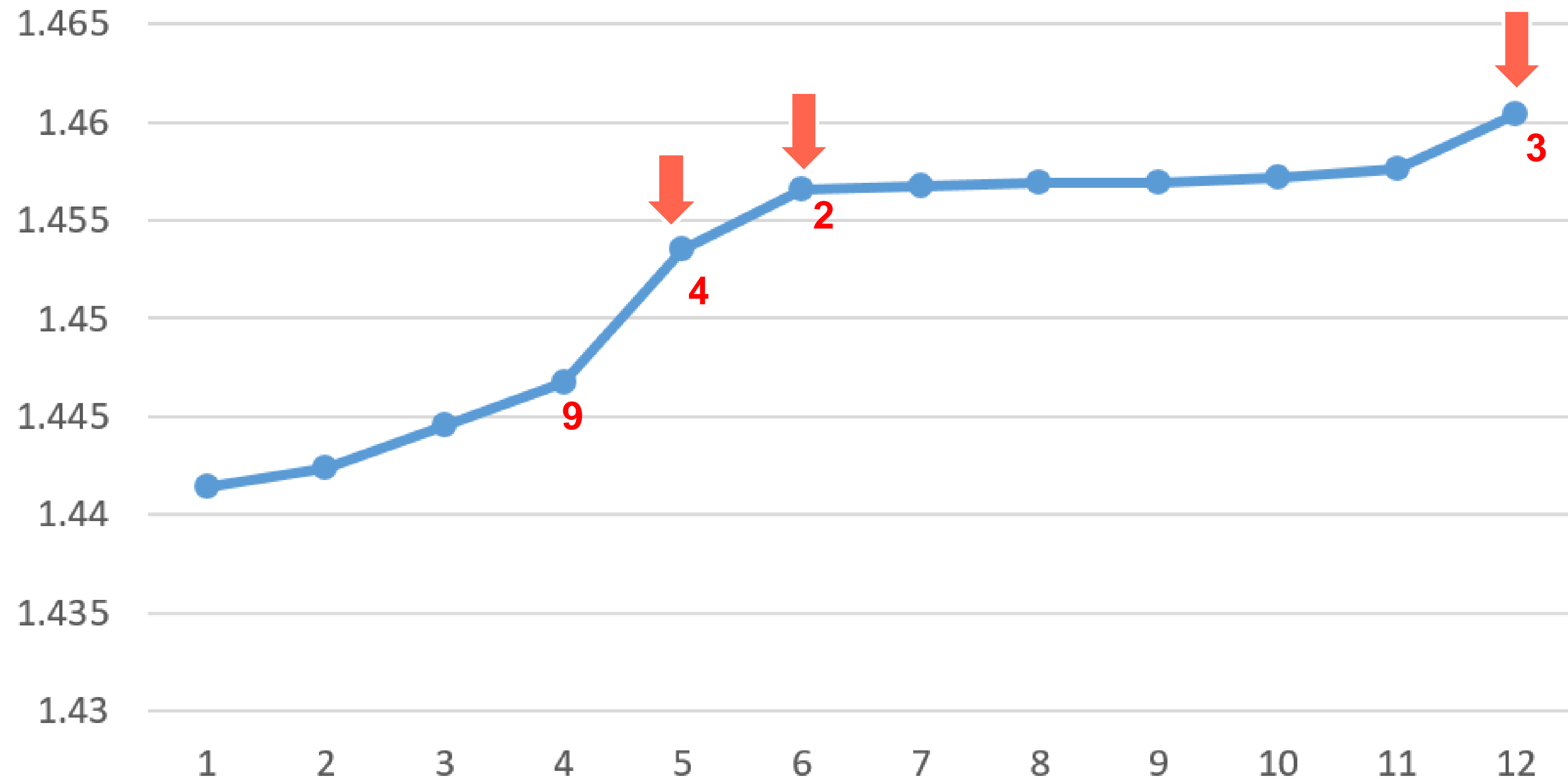
通过将每个模块的差异点，进行交叉组合，训练多个模型，再对结果进行线下加权



总结与思考

DEMO-初赛分数变化

初赛分数

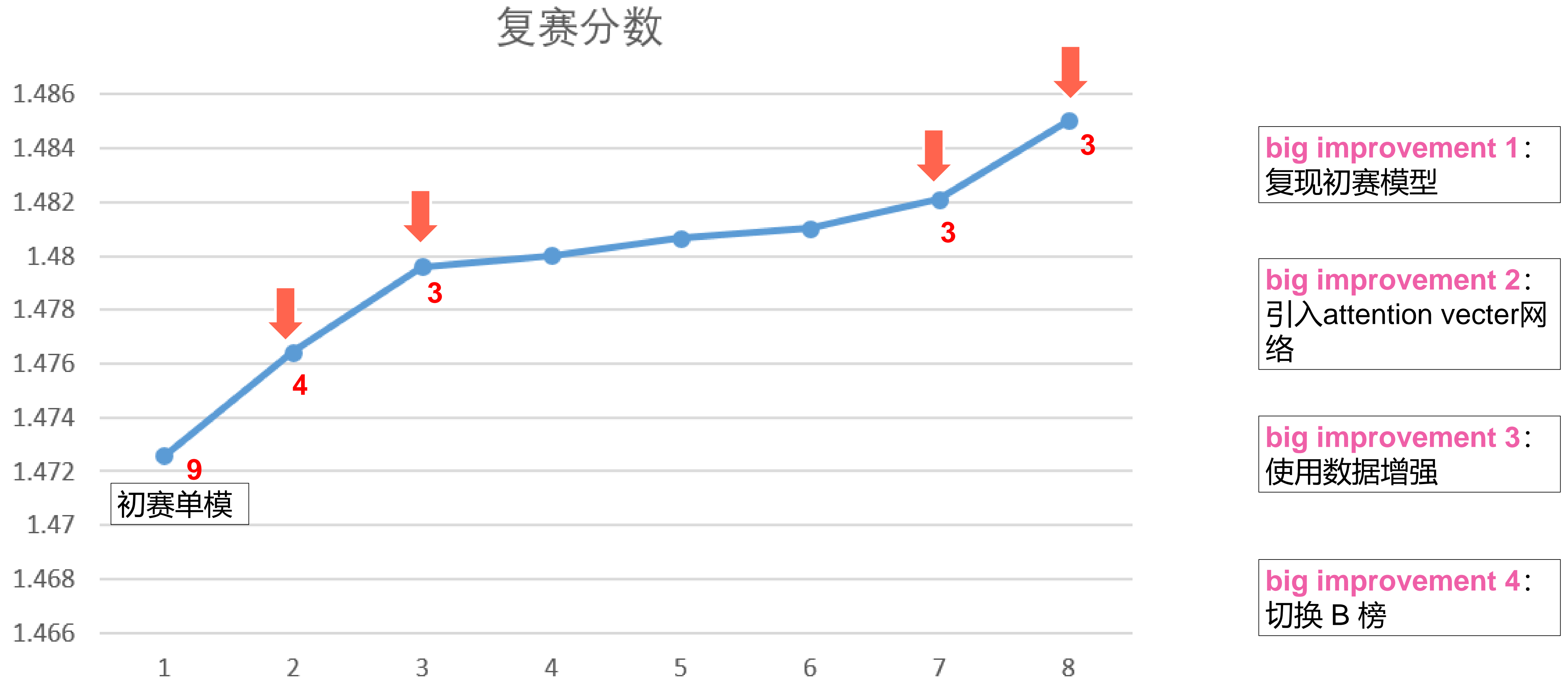


big improvement 1 :
改用多序列输入 + label
smoothing + 余弦退火

big improvement 2 :
组队融合

big improvement 3 :
切换 B 榜

DEMO-复赛分数变化



最终我们的模型在初赛A榜B榜和复赛A榜B榜都获得了前三名的成绩

总结与思考

数据探索

数据探索与处理是模型能获得强大泛化能力的基础。结合对实际业务的理解，去除异常数据，增广数据能有效的提高模型的泛化能力。

模型设计

模型结构的设计是至关重要的，不仅要保证训练过程的高效率、高质量，还要保证构建的模型具有差异性；差异性模型的融合往往会收获意想不到的结果。

训练技巧

模型能否收敛到最优点直接决定了模型的泛化能力，优秀的训练技巧和经验可以使得模型快速地收敛到更优的值。在训练模型的过程中，尝试各种训练技巧是非常有必要的。

思考

我们将人口属性预估任务转化为序列建模建模的同时，丢失了许多拓扑结构信息，如果可以将任务转化为图上的半监督学习任务，使用 GNN 对拓扑结构信息建模，相信模型的性能还会再上一层楼。

THANKS

