深度模型在推荐系统中的实践

召回阶段采用双塔结构，用户侧与物品侧分别编码，线上通过近似最近邻检索完成候选生成。

特征工程方面，序列特征按时间窗分桶，减少了长尾稀疏带来的过拟合风险。

粗排阶段引入轻量级交叉特征，在算力预算内将排序质量提升到可用水平。

训练稳定性上，采用梯度裁剪与学习率预热的组合策略，收敛过程更加平滑。

精排阶段使用多任务学习框架，同时优化点击率与转化率两个目标。

线上推理通过批量合并与量化压缩，将平均延迟控制在可接受范围内。