**基准模型对比方案**

**实验目标**

在相同数据集Criteo (Criteo Uplift Modeling Dataset)、相同预处理和评估流程下，对比八种模型的点击率预测性能。

评估指标包括：Area Under Curve、LogLoss、PR-AUC、训练时间、推理时间。分析各模型在不同预处理、超参数设定下的敏感性。

数据集由25M 行组成，每行代表一个用户，具有 11 个特征、一个处理指标和 2 个标签（访问和转化）。

**数据预处理**

连续特征缺失：统一填充为 0，并增加二元指示特征。

类别特征缺失：填充为特殊字符串 "unknown"。

特征编码：连续特征Z-score标准化；类别特征采用维度8的 embedding；对超长尾类别，出现次数 < 100 的类别合并为 "rare"。

数据划分：随机打乱后按 70/15/15 划分为训练集/验证集/测试集。

**对比模型列表**

| **序号** | **模型名称** |
| --- | --- |
| 1 | FTRL (Follow-The-Regularized-Leader) |
| 2 | FFM (Field-aware Factorization Machine) |
| 3 | Wide & Deep |
| 4 | DeepFM |
| 5 | Deep & Cross Network (DCN) |
| 6 | DMR (Deep Match to Rank) |
| 7 | DIN (Deep Interest Network) |
| 8 | CTNet (Continual Transfer Network) |

**实验实现**

Python + PyTorch + [DeepCTR](https://github.com/shenweichen/DeepCTR)**。**

batch\_size = 1024; embedding\_dim = 8; MLP\_layer = 3; hidden\_unit = [256, 128, 64]; ReLU。

学习率 lr ∈ {1e-3, 5e-4, 1e-4}；

L2 正则化系数 λ ∈ {1e-3, 1e-4, 1e-5}；

Dropout p ∈ {0.0, 0.2, 0.5}；

超参数调优 通过验证集上使用网格搜索或贝叶斯调参。

**评估指标**

主要指标：AUC（Area Under ROC Curve）；LogLoss（对数损失）；PR-AUC（Precision-Recall AUC，用于不平衡场景）。

次要指标：Calibration - Brier score；训练/推理时间。

**结果呈现**

表格汇总：测试集的 AUC、LogLoss、PR-AUC 列表；超参数设置与模型复杂度一览表；训练/推理时间对比。

可视化：ROC 曲线 & P-R 曲线；Calibration 曲线。