

AUC 与 GAUC 的核心差异及敏感度分析

算法小小怪下士

2025 年 12 月 19 日

1 AUC (Global AUC)

1.1 定义

AUC (Area Under the Curve) 通常指 ROC 曲线下的面积。在排序任务中，它的物理含义是：随机给定一个正样本和一个负样本，模型将正样本排在负样本前面的概率。

1.2 数学表达

假设数据集为 \mathcal{D} ，其中正样本集合为 \mathcal{D}^+ ，负样本集合为 \mathcal{D}^- 。 $f(x)$ 为模型对样本 x 的预测分数。

$$AUC = \frac{1}{|\mathcal{D}^+| \cdot |\mathcal{D}^-|} \sum_{x^+ \in \mathcal{D}^+} \sum_{x^- \in \mathcal{D}^-} \mathbb{I}(f(x^+) > f(x^-)) \quad (1)$$

其中 $\mathbb{I}(\cdot)$ 是指示函数，条件满足为 1，否则为 0。

1.3 AUC 的局限性

AUC 将所有用户的样本混在一起进行比较。这会引入两个问题：

1. 不同用户间的不可比性：模型给用户 A 的打分偏高（例如 0.9），给用户 B 的打分偏低（例如 0.4）。即使模型正确预测了用户 A 点击了 (0.9) 且用户 B 没点击 (0.4)，但这并不代表模型懂个性化，只代表模型学到了用户 A 比用户 B 更容易点击。
2. 样本对的偏差：全局 AUC 关注的是所有样本对的排序。

2 GAUC (Group AUC)

2.1 定义

GAUC (Group Weighted AUC) 的核心思想是：将样本按用户（User）分组，先计算每个用户内部的 AUC，然后进行加权平均。这消除了用户偏差（User Bias）对模型评估的影响。

2.2 数学表达

假设有 n 个用户，对于第 u 个用户，其内部的 AUC 记为 AUC_u 。该用户的权重记为 w_u （通常是该用户的曝光次数或样本数）。

$$GAUC = \frac{\sum_{u=1}^n w_u \times AUC_u}{\sum_{u=1}^n w_u} \quad (2)$$

3 为什么 GAUC 对高频用户更敏感？

这是 GAUC 设计中最关键的特性之一。我们可以从公式中的权重项 w_u 进行分析。

3.1 权重的物理意义

在工业界实践中，公式 (2) 中的 w_u 通常定义为用户的曝光量（Impressions）或样本总数。

3.2 敏感度分析

让我们考察某个用户 u 的表现变化 ΔAUC_u 对整体 $GAUC$ 的影响：

$$\frac{\partial GAUC}{\partial AUC_u} \propto w_u \quad (3)$$

这意味着：

1. 贡献度差异：一个拥有 1000 次曝光的高频用户的 AUC 下降 0.1，对整体 $GAUC$ 的拉低效果，等同于 100 个拥有 10 次曝光的低频用户 AUC 同时下降 0.1。
2. 流量导向：互联网产品的核心 KPI 往往由头部活跃用户贡献大部分流量。GAUC 通过 w_u 加权，本质上是在评估“在所有展示出的流量中，有多少比例的流量被正确排序了”。
3. 消除噪声：低频用户由于样本极少（例如只有 1 正 1 负），其单体 AUC_u 方差极大（要么 1.0 要么 0.0）。如果对所有用户取简单平均（即 $w_u = 1$ ），指标会被长尾用户的噪声淹没。通过使用样本量 w_u 加权，GAUC 自动降低了这些低置信度用户的影响。

4 总结

综上所述，GAUC 之所以成为推荐系统的主流指标，是因为它不仅剔除了跨用户比较的干扰，它比较对象的从所有的正负样本对（含跨用户），变为仅同一用户的正负样本对。更通过权重分配，确保了模型优化方向与系统的主要流量来源（高频用户）保持一致。