推荐多样性

推荐多样性暂且搁置

AutoFas：

为了提高模型的效果，一些方案会额外使用精排的打分知识进行蒸馏。但仍有两大挑战亟待解决：

1.如果不把时延真正作为一个变量放到模型中进行联合优化，效果必然大打折扣；

2.如果把精排的打分知识蒸馏给一个手工设计的粗排结构，模型的表现也肯定不是最优。

本文使用了神经网络框架搜索 (Neural Architecture Search) 的方法，开创性地提出了AutoFAS (Automatic Feature and Architecture Selection for Pre-Ranking System) 的算法框架，统一解决了以上两个问题：在给定时延限制和精排打分知识指导的条件下，同时选出最优的粗排特征与结构组合方案

目前的推荐链路已经进化为召回-匹配-预排序-排序-重排。粗排和精排用多目标模型为物品做pointwise打分。

经典的推荐系统链路主要包括四个部分：匹配、预排名、排名和再排名。匹配阶段将用户活动历史中的事件以及当前的查询（如果存在的话）作为输入，并从大型语料库（数百万）中检索出一个小的项目子集（数千）。这些候选项目的目的是与用户普遍相关，具有适度的精度。然后，预排序阶段提供广泛的个性化，并过滤出具有高精确度和召回率的前几百个项目。一些公司可能会选择结合匹配和预排名阶段，如Youtube。然后，复杂的排名网络根据所需的目标函数，使用描述项目和用户的丰富特征集，为每个项目分配一个分数。得分最高的项目被呈现给用户，如果没有重新排名，则按其得分排名。一般来说，预排名与排名的功能相似。最大的区别在于问题的规模。在预排名系统中直接应用排名模型将面临计算能力成本的严峻挑战。如何平衡模型性能和计算能力是设计预排位系统的核心部分。