特征工程 | 特征交叉 (Feature Crosses)

原创 Thinkgamer 搜索与推荐Wiki 2020-08-31

收录于话题

#特征工程 92 #Thinkgamer 11 #精品小系列内容 25



▼ 精彩推荐 ▼

- 1、推荐系统embedding技术实践总结
- 2、从极大似然到对数损失函数和交叉熵损失函数,以及对数损失优化取值范围
- 3、从DSSM语义匹配到Google的双塔深度模型召回和广告场景中的双塔模型思考
- 4、关于推荐算法工程师大家比较关注的几个问题
- 5、CTR预估模型中的正负样本问题

组合特征也叫特征交叉(Feature Crosses),即不同类型或者不同维度特征之间的交叉 组合,其主要目的是为了弥补CTR场景中早期的模型无法有效的进行特征组合而产生的想 法和实践,随着算法模型的进步,虽然一些机器学习排序模型(GBDT+LR、FM等)和深 度学习排序模型能够捕获特征之间的联系,但是前期仍然会生成一些组合特征,比如用户 在某些类别下的行为统计,性别下的数据统计等。

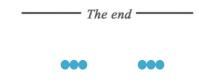
日常工作场景中的特征往往都是以用户和其对应的属性作为出发点,然后和物品对应的属 性特征或者上下文特征进行交叉组合,比如:

- 用户与类别特征组合(eg: 电商平台中, 用户在不同类别下的点击次数, 可以简单衡量用户 的类别偏好)
- 用户不同年龄档与类别特征组合(eg: 新闻平台中,少年用户在娱乐新闻下的点击次数会比 较高,青年用户在社会新闻下的点击次数会比较高)
- 用户身份与时间特征组合(eg: 学生周六周日、节假日,在某娱乐APP内的行为会比较丰 富,上班族早晚高峰,在某娱乐APP内的行为会比较丰富)
- 用户的价格偏好与物品的价格级别特征组合(eg: 用户A偏好的价格级别为中,物品的价格级 别为高,那么组合特征就是0,即两者不匹配)

这里只是举几个具体的例子,在实际的场景中,往往会结合数据分析去挖掘潜在的价值特 征组合。这时候便会借助数据分析、数据挖掘、分析者的敏感直觉和对业务的深入理解去 进行挖掘。笔者之前有幸听到一个分享,大概意思就是说不同人群对于APP推送消息的处 理风格和处理时间很不一样,比如有的人会直接划掉,有的人则会在某个时间段进行消息 查看,这时候便可以考虑对用户Push消息提供个性化时间分发,从而提高用户的打开率。

类似的场景还有很多,比如安装某些APP的用户偏好内容不一致,手机机型不同的用户阅 读的咨询内容风格也会有差别等,所以说做算法、做数据、做分析,脱离了业务理解都是 耍流氓!

在构造好组合特征之后,对于特征的处理方式就和上述介绍的离散特征处理、连续特征处 理方式相同了,不在赘述!



真正的努力,都不喧嚣!



搜索与推荐Wiki All In CTR, DL, ML, RL, NLP

分享,点赞,在看,安排一下?

收录于话题 #特征工程·92个

上一篇

下一篇