

我这样一个尝试跟真正的推荐系统还有那些距离要走

1. 我没有用到隐式数据，只有显式，稀疏度大约 1%-2%，而隐式数据的使用是更加普遍的
2. 还不曾立足于业务，比如在现实中优质和劣质答案的分辨，抖机灵多赞和认真少赞的，所以也还不能自主确定训练数据的成分，目标与衡量指标
3. 现实中有一些反馈，比如 AB 测试，这部分的数据价值如何利用并验证推荐，寻找推荐模型的问题和增长点，我也不清楚
4. Netflix 的冠军模型，那可是 100 多种算法集成在一起的，现实中的推荐，也是应该有复杂的集成，我的集成做法还十分简单
5. 对数据结构对性能的影响了解的还不够深