1. 任老师：自己本地是否保留数据？尤其是推荐所需的数据

答：推荐所需的数据是推荐时才爬取，所以分析会比较久，其他的数据是保存的。

1. 任老师：为什么考虑推荐数据现爬

答：一方面考虑到保存本地可能导致分析所需数据量和种类较少，不够多样。另一方面考虑到这一部分数据更新可能没法及时跟进。而通过个人数据，得到喜好标签，再去对应网站爬取数据，虽然速度较慢，但是种类和及时性都较好

1. 任老师：所以推荐的算法就是根据用户喜好获得标签，再用标签去寻找是吧？

答：是的，而且如果用户近期没有数据更新，也会智能的推荐上一次的数据

任老师：但是这样的推荐可能意义不大，用户可能永远只看到自己想看的，但是真正的推荐不应该是单向比较，而应该是推荐自己想不到的，横向的比较，以及同类用户的比较才比较合适。

答：其实我们也考虑过做同类比较，但是数据量需求比较大，同时机器学习我们也有用于自己的项目中，比如淘宝的商品分类。

任老师：这个问题总结起来就是，本地应该存储大量数据，并用一些简单的机器学习的方法进行分析推荐，当然还有很多问题比如是否用户愿意把自己的数据用于分析等等。

答：真正的分析推荐大概是不同用户的用户画像之间作比较，然后相似的进行一些算法处理，但我们真正可用的数据量比较小，最后生成出来的推荐可能还是只能随机的。

任老师：但是至少你们可以两类一组试一试，简单的横向比较做一做。

4. 助教：用户注册进来之后服务器可以定期的爬去一些数据，这样用户登录之后就有爬取的数据，不然用户登录之后还要等待获取数据，而且推荐算法看起来，八个数据之后没有一个共同推荐，这样看起来你们做一个和做八个没有太多区别。

答：关于爬取的问题，其实每次爬取之后用户都会将数据存在我们的数据库中，用户除了第一次登陆之外，其实都已经有数据存储，不需等待，只不过需要手动更新，因为需要获取账户的一些登陆信息，只能登陆之后用户手动发送请求。同时推荐算法确实综合的部分比较缺乏，一方面前期爬虫工作量较大，另一方面我们花了比较大的心思放在了单个数据分析上，因此对于综合推荐确实有不足。

1. 任老师：那你们数据是怎么存的呢？

答：用mysql，感觉需要的数据用mysql存比较合适

任老师：那为什么用mysql呢？

答：一个是数据结构化程度比较高，同时存的数据没有超大文本和图片，因此没有考虑用mongo

助教：那你们对于数据的存储花费的时间有做过测试嘛？如果用mongoDB无论读和写都有很快的提升

答：因为我们数据存储是和爬虫一起，边爬边存，因此没有单独测试

1. 任老师：那对于压力测试的数据展示有做过优化吗？

答：有做过一些优化，比如之前测试时有发现微博的端口特别慢，最开始发现是数据库跨了服务器，后来放在一台服务器之后还是有很长的时延，于是后端做了分页，很快速度就提上去了

任老师：那你们有想过更快吗？因为虽然分页，但是有些数据仍然会很大，对于一些没有必要一口气展示的数据，可以先去一部分，做一些缩略。

答：对于知乎那一块，渲染的时候做了一定缩略，先展示总的，如果想看细节，再用id去拿