**会议主题：**1 任务进度报告：

2 可用的模型测试

3 个人问题解决

**会议时间：**2021.3.3

**会议主持：**王修文

**会议内容：**

王修文：

个人进度：以上一次的基本结果为比较，测试不同的数据处理和 构建word-embediing方式， 数据处理：去掉数字，英文，网址 搜索文件，构建word-embediing方式 字向量 词向量 tf-idf n-gram

一个没有感情的抄材料机器

个人进度：对Age任务 采用词袋+lda，分层抽样

问题：过程中发现有两条数据是纯英文的 一条数据全部为百度网盘链接，另一条数据全为英文单词

刘州

个人进度：对数据进行清洗，代码部分没跑通。分类的交叉验证代码有样例，因为卡在上面，暂时没有测。建模部分看了一些案例和视频，暂时没有思路

问题：替换网址的函数操作有问题，想偷懒直接替换。之前说的语句部分的思路是--需不需要单独分解出来作为单独一列，进行拆分，这样有没有意义，需不需要这么操作？

陆天浩：

个人进度：对数据进行清洗，代码跑出来有点乱码问题，目前在进行交叉验证

问题：k折设置一下就好的，感觉这代码有点无从下手

模型选择：暂时未定，视情况做，NB、SVM、CNN、RNN