时间安排

红色代表梁峥来完成，蓝色代表唐老板完成

数据准备（已完成）：

1. 处理camera和monitor数据集（设计例子：多值、文本数据、标注训练集、两个语义相似的源）（陈老师完成）
2. 查找新的数据集——找数据集合并、restaurant、weather

方法：

1. 数据增强
2. 代码确认（包括：概率融合替换、主动学习方法）——完成
3. 布置主动学习方法修改任务——完成
4. 布置概率融合和同义替换法任务——完成
5. 知识表示
6. 与CTD连接、超参搜索readme撰写（1.4完成）
7. 构建任务特征-embedding参数的经验库，设计初值策略(分割多个数据集、设计多个数据源embedding的计算策略（可以考虑融合）、搜索最接近数据集用其超参初始化) （1.6讨论）
8. 相似度替换
9. 连接3种算法代码（唐：开放值相似度、源相似度怎么做单向、源可信度的接口）（1.4完成）
10. 抽样策略设计与伪代码撰写（待完成）
11. E.A.T
12. A.T到E.T：方案初步设计文档撰写（待完成）
13. E.T：TextCNN尝试（需要数据：要高维（1.4完成），思考）
14. 完成实现，初步测试（待完成）

论文：

1. 对比实验（优化CTD,1.5完成）、消融实验等

唐老板code review：

（1）数据集说明、用了哪些否定约束？

（2）换数据集，用restaurant和weather（陈老师给）

（3）实验流程

将CTD-semi、CTD-GA评价函数改为结合主动学习的抽样RMSE（抽样策略要与主动学习结合SLIMFAST）+无监督参数、值相对变化（需要读论文决定IATD、DART、CTD）

最小化目标还加一个所有数据RMSE（估计）+标注数量（已知）

估计方式：所有数据RMSE/所有数据信息量=部分数据RMSE/部分数据信息量

部分数据信息量= +kmeans数据量

所有数据信息量= +kmeans数据量

唐：先用kmeans抽样标注、运行CTD-semi、运行CTD-GA

陈：用slimfast预测、判断置信度低的样本、作为下一轮的标注

（4）TextCNN如何获得数据？可以用数据增强解决这一问题

任务：

（1）按照（3）做实验改函数评价指标（a+b，归一化后，其中主动学习和陈老师讨论），CTD（唐）、IATD（陈）、DART（陈）修改只修改收敛函数，（1.12初步完成）

（2）源数据（1.12讨论）

（3）换数据集，与陈老师沟通接口（1.12完成）

（4）冷启动策略（1.12思考）