

一、本周研究内容

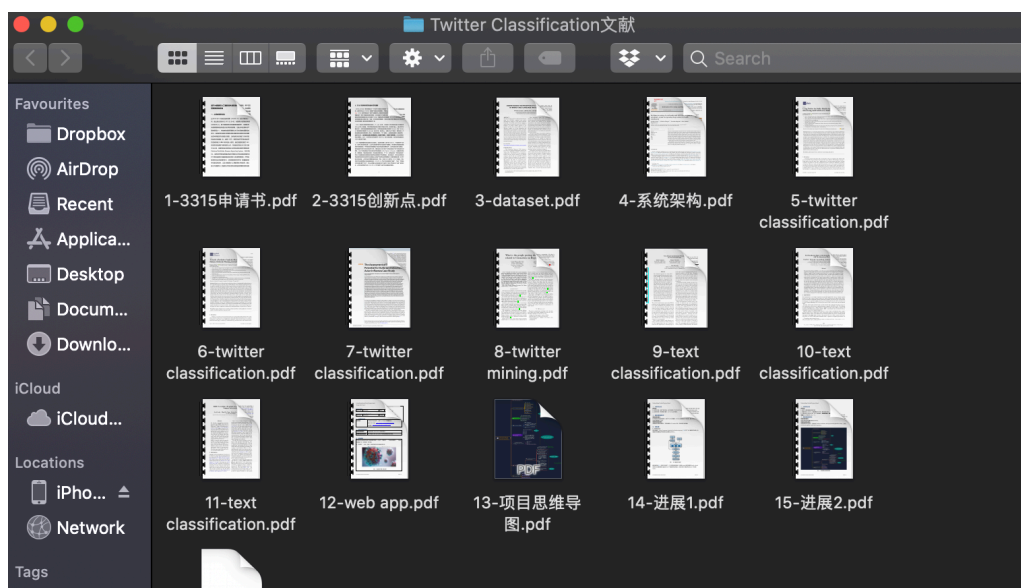
1. 列出了详细的 “COVID-19 Twitter Mining and Text Classification” 项目计划方案，包括我的部分以及 Ke Chen 同学的部分，以及相关的文献资料。
2. 继续对 Text Classification 数据集进行手工标注，本周标注了 2000 条 tweets，标注总进度 3000/7000，预计下周完成全部标注。
3. 实现了基于传统机器学习的算法 KNN 和 SVM 用作该项目的 baseline。

二、项目实施当前状态

已经完成了近一半的数据集手工标注以及建立了 baseline，预计下周完成全部数据集标注。

三、本周成果

1. 项目计划方案：分享给 Ke Chen 同学的资料括 15 个文档以及目前处理后待标注的 twitter 数据集。15 个文档分别涵盖了项目申请书、创新点、预警系统架构的相关文献、twitter classification 的相关文献、twitter 数据集的相关文献、text classification 的相关文献、我在周报中汇报的项目的目前进展以及项目整体思维导图。数据集为文件 ‘coronavirus_merged_label.pkl’。写了项目计划，对我和 ke chen 同学接下来的任务进行了详情规划。



图一：整理给 Ke Chen 同学的文献资料

2. 手工标注数据：本周标注了 2000 条 tweets，标注总进度 3000/7000。

3. Baseline 建立：代码实现了 KNN 和 SVM。

四、下周计划

预计下周完成全部 7000 个数据的手工标注，并与 Ke Chen 同学一起对项目进行探讨。