

实训计划

内容：食品安全事件检测与舆情分析

1 / 第一周

- 余连玮：理清《网络评论文本的细粒度情感分析研究》和《文本话题识别算法的研究与实现》论文所用到的技术；
- 王子昂：理清《基于网络数据的食品安全事件检测与分析》和《社交网络推文情感分类系统的设计与实现》所用到的技术，将论文中涉及的技术整合到一个框架中，确定之后的研究和开发方向
- 两人共同完成：讨论并将上述论文的技术整合到一个框架，确定后续深入的方向。

- 输出：得到一个整合上述所有技术的大纲。

2 / 第二周

- 王子昂：配置环境，整理命名实体识别代码（标注、识别）、话题提取部分、事件检测部分的代码，理解代码结构和实现方式。基于源代码的基础上进行代码的重构，添加必要注释，修改其中存在的问题，使整体结构清晰，代码易于重用。结合得到的新的数据源，重新进行数据处理，命名实体识别、构建新的库语料、训练词向量、得到新的 LDA 模型。
- 余连玮：寻找新的适合的视频或音频数据源，进行文本提取，形成可调用的 api，集成到邓娜代码中。帮助王子昂重构代码。
- 输出：得到更加丰富的数据集、并对新的数据集进行数据处理、得到重新训练的可用模型。

3 / 第三周

- 余连玮：复现《文本话题识别算法研究与实现》中 VSM 模型部分，与 LDA 模型相结合计算相似度。

- 王子昂：复现《文本话题识别算法研究与实现》中结合 LDA、VSM 模型得到的相似度，进行文本聚类，实现话题提取。
- 输出：得到一个改进后能提取话题的系统。

4 / 第四周

- 余连玮：复现《社交网络推文情感分类系统的设计与实现》中的图像情感分类模型的搭建，包括 MFES 多级特征提取器模型，以及 MFIC 多级特征交叉分类器。
- 王子昂：复现《社交网络推文情感分类系统的设计与实现》中的图像文本联合分类算法中训练方式的实现，包括两个算法（TIS 算法：自动标注算法 CH 算法：图像去重算法）。
- 输出：实现一个图像情感分类模型。

5 / 第五周

- 共同完成：寻找中文推文情感数据集，搭建 LSTM 神经网络模型提取文本

特征向量，复现《社交网络推文情感分类系统的设计与实现》中的图像文本混合网络 (VTFN)，引入相关性损失函数，用于模型的训练。

- 输出：得到一个可运行的中文社交网络推文（图文）情感分类系统。

5 / 第六周

- 共同完成：将图像文本联合分类算法整合进食品安全事件的舆情分析部分，一方面将只针对微博、知乎的文本评论转换为更加丰富的图像文本联合信息，另一方面得到更细粒度的情感划分。
- 输出：得到一个完整的食品安全事件检测与分析系统：采集文本、视频、音频数据，进行事件检测与提取，最终对相关社交网络的推文进行细粒度的情感分析。