# 10.30-11.6第一周

总结：研读《机器学习》1-10章，理解机器学习原理，初步理解主流机器学习算法。

下周任务：

1）继续学习机器学习算法，掌握每种方法的优缺点

2）熟悉并开始使用Weka工具，使用Weka官网上的数据信息、或Kaggle的数据信息，或其他。

# 11.7-11.13第二周

总结：

1）温习机器学习算法，总结了算法的优缺点

2）使用并学习了Weka工具，下载了一些数据集进行算法简单应用。初步理解weka使用流程和主流机器学习算法的实践、参数调整、结果可视化的应用。

下周任务：

1）学习Weka过滤器中SringToWordVector类的原理作用

2）利用Weka接口编程实现算法

# 11.14-11.20第三周

总结：

1）学习并使用Weka过滤器中SringToWordVector类。

2）利用Weka接口编程实现J48、NavieBayes算法，了解了wekaAPI的基本用法。

下周任务：

1. 将评论数据文件导入数据库中
2. 初步制定毕设问题的解决方案

# 11.21-11.27第四周

总结：

了解了gephi可视化工具、Github等代码版本控制管理工具

1. 将apps.dat和review.dat导入到SQLserver中
2. 将毕设问题初步分解为“文本处理”“文本分类”“文本分析”三部分

下周任务：

1. 将数据信息做基本的统计（如统计英文信息，统计一个app对应的评论个数信息等）
2. 抽取数据集中200-300条信息，划分评论类别
3. 尝试体验StanfordCoreNlp工具
4. 思考文本向量的特征

# 11.28-12.04第五周

总结：

1）统计了数据库中Apps表和Reivews表中的有效数据，筛选出了英文文本、app与评论相互对应的信息

2）抽取了250条评论信息进行人工分类

3）初步翻译了User Feedback in the AppStore: An Empirical Study这篇文献

4）运行了StanfordCoreNlpDemo.java文件，体会了该工具的使用

下周任务：

1. 对app按照对应的评论条数进行分类统计
2. 将每一个评论的类别进行例证，修改类别名称，补充类别定义
3. 尝试使用StanfordCoreNlp工具分析非英文文本，判别是否为无效评论
4. 用StanfordCoreNlp工具分析评论信息，观察分析结果，分析句子结构
5. 主要分析“需求”类的句子结构，提炼共同点
6. 继续思考文本向量的特征

# 12.05-12.11第六周

总结：

1）对app按照对应的评论条数进行分类统计

2）将每一个评论的类别进行例证，修改类别名称，补充类别定义

3）尝试使用StanfordCoreNlp工具分析非英文文本，会识别为名词

下周任务：

1. 继续修正评论类别
2. 再抽取250条评论信息进行人工分类
3. 主要分析“具体评价”和“需求”类的句子结构，提炼共同点
4. 利用聚类的方法自底向上分析评论语法结构，分析文本特征是否为有关特征；思考如何分析语法树的相似度
5. 看论文，提取相关内容

# 12.19-12.25第八周（第七周因考试暂停）

总结：

1. 修正评论类别，去掉“交流”类
2. 随机抽取500条评论信息进行人工分类，可以分为定义好的四类
3. 参考论文，并从“需求”类别的评论中简单抽取了频率高的词汇和句式

下周任务：

利用单词、词组、语法结构、词之间的依赖关系进行特征定义并筛选有关特征。