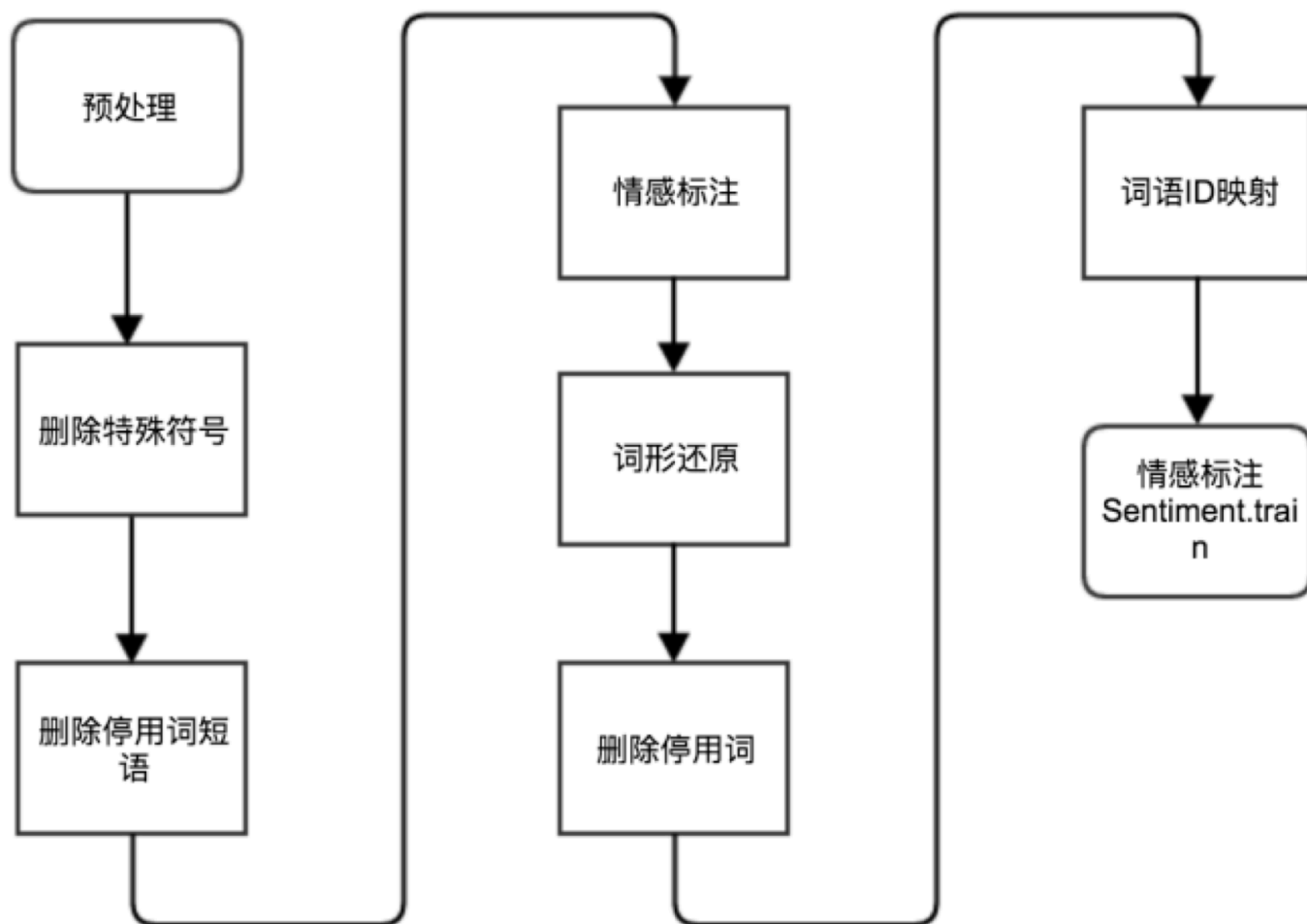


数据预处理

处理流程



文件列表

- 1_Cleaned.py 删除特殊符号
- 2_Sentiment.py 情感标注
- 3_WordRestore.py 词形还原
- 4_StopWords.py 删除停用词
- 5_GeneratedWorDict.py 生成词典
- 6_TransWord2ID.py 将词语映射到ID
- 7_IDF.py 计算IDF
- 8_Weight.py 计算权重
- AddLabel.py 为无标数据添加Label
- autorun.py 自动运行

- review.unlabeled 需要预处理的无标数据
- dic/ 情感标注中用到的字典
 - DegreeLevel.txt 表示程度的词
 - NegativeEmotion.txt 负面情感词
 - NegativeEvaluation.txt 负面评价词
 - OpsitiveEmotion.txt 正面情感词
 - OpsitiveEvaluation.txt 正面评价词

环境配置

1. Python 2.7
2. Pattern.en
3. NLTK

运行方法

1. 将需要预处理的数据放到该文件夹中。
2. 命令行运行：

```
1 | python autorun.py
```