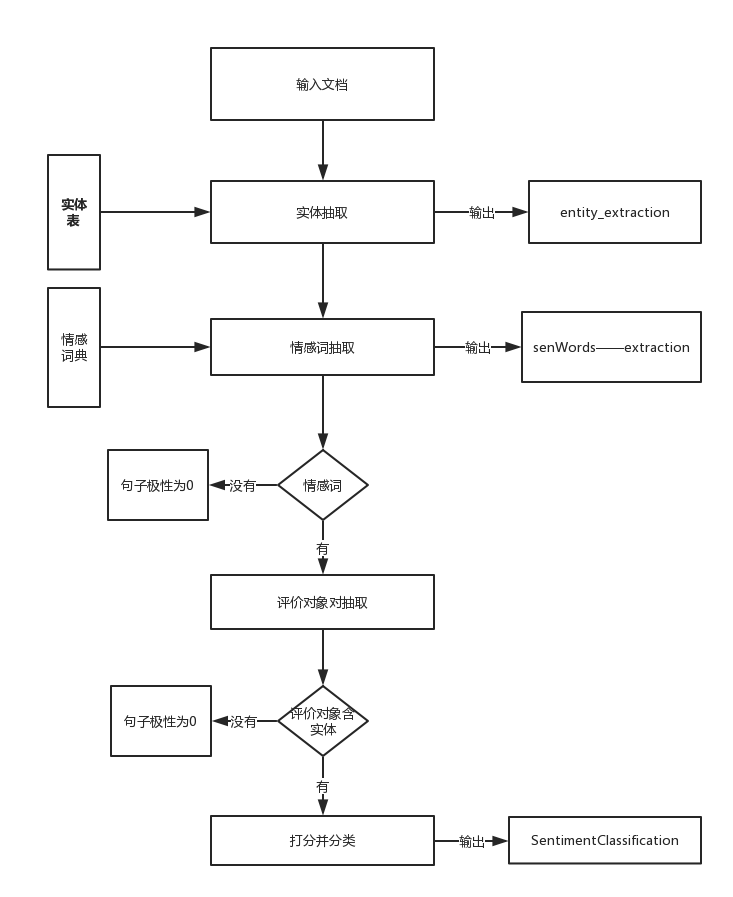
### 2018-06-29

#### 整理代码

将之前的程序整合成了一个demo，同时编写了使用说明书。

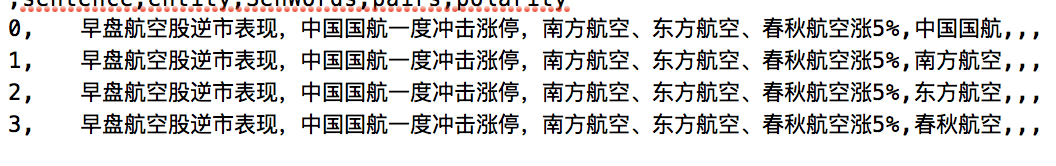
算法框架如下：



#### 优化规则

* 对于公司实体之间存在的并列关系，导致对应的评价对象、情感词对抽取失败的情况。

如下面这种情况：



* 一句话若是单实体且单情感词，则往往该实体和情感词极性一致，增加了相对应的规则。

#### 词典实验

从语料中选取了10w条新闻，按照《Inducing Domain-Specific Sentiment Lexicons from Unlabeled Corpora》的代码，生成SVD embedding向量。

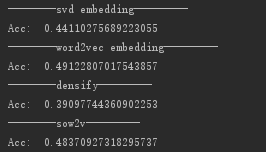
并且标注了400个词作为测试集，分别进行了四次实验。

svd embedding：SVD embedding向量作为输入向量，采用论文sentprop算法（基于图的方法）打分。

word2 embedding：通用的word2vec embedding向量作为输入向量，采用论文的sentporp算法打分。

densify：通用的word2vec embedding向量作为输入向量，采用论文引用并实现比较的另一算法，densify打分。

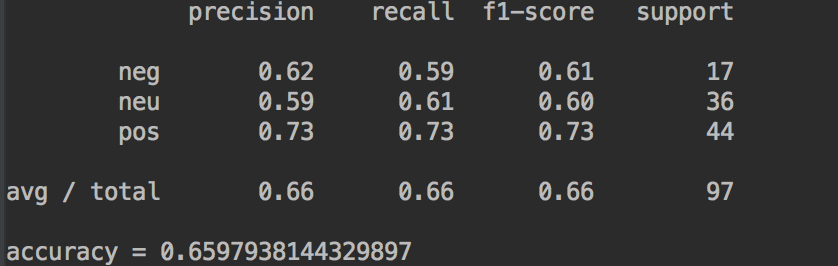
sow2v:：通用的word2vec embedding向量作为输入向量，采用计算词向量之间余弦相似度的算法打分。



#### 实验部分

利用上述的规则优化后，随机抽取了100篇单实体和多实体新闻，小组成员共同标注，计算其准确率。

其中单实体的结果为：



其中多实体的结果为：

