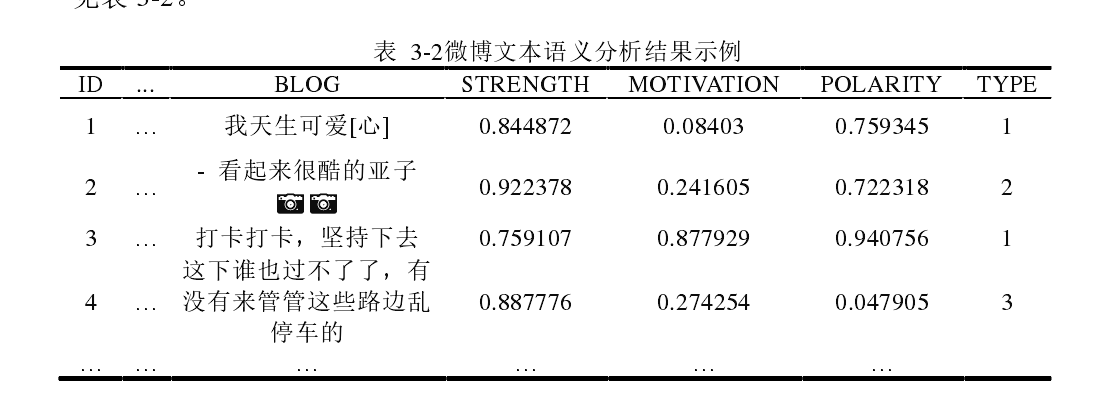
1. **数据获取及预处理**

3.3 情绪数据预处理

根据前文中针对文本语义分析方法的总结，以及微博文本篇幅短，新兴词汇量大，难以用现有的情绪词典简单判断的特点，本研究采用有监督的机器学习的方法来对微博数据进行情绪分类。在人工标注部分微博后，再采用机器学习的算法来训练分类模型，经测试集验证准确率后再应用于全体数据

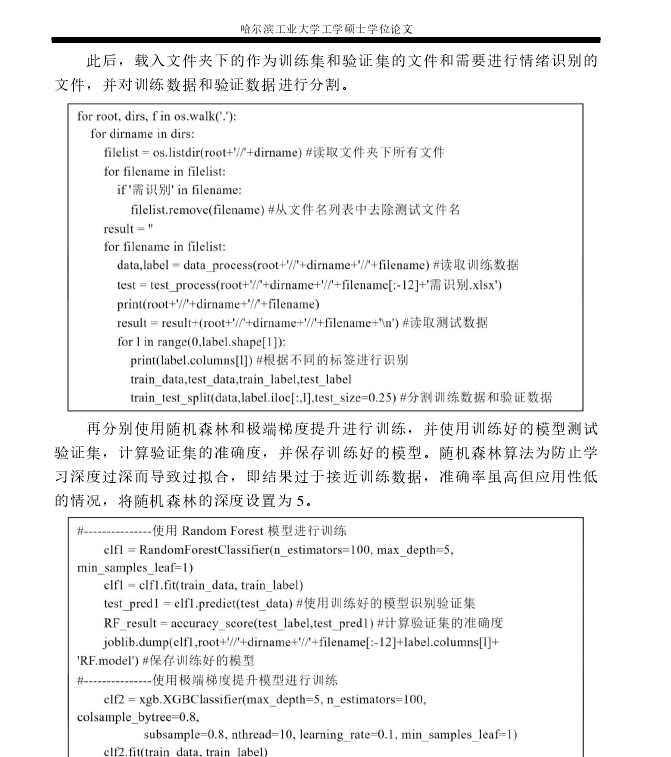
1. 人工标注部分语料
2. 微博语料处理
3. 情绪特征提取

这部分需要我先手动标注情绪强度（0-1）、情绪极性（-1到1）、情绪类型（乐怒哀...）

需要您这边以我标注的结果进行模型训练然后输出上面这样的表格↑ 我的数据是带有地理坐标的，接下来就可以用gis出地理可视化和分析的图了

**第四章 历史城区情绪地图分析**

4.1 情绪识别模型构建

这里需要一些代码和过程文件如下图

4.2 整章都基于上一章节的结果进行分析

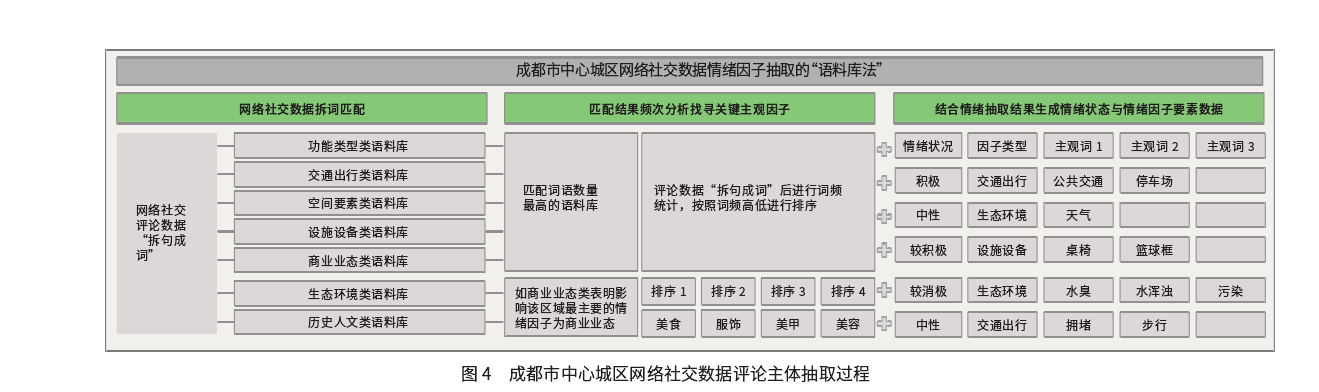
4.2.1 情绪强度分布

4.2.2 情绪极性分布

4.2.3 情绪类型分布

...

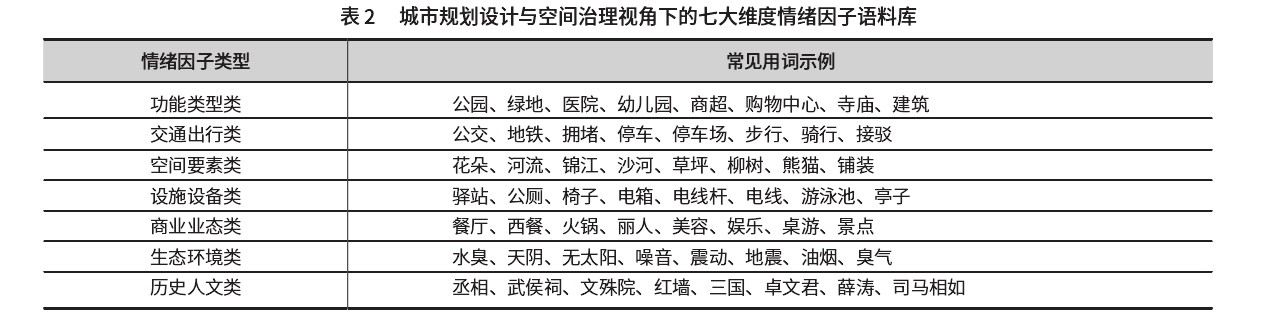
**第五章 历史城区情绪因子提取及空间回归分析**

这个是lda么？我这块还没太看明白流程

可以再讨论一下，大概就是需要您整句拆词后从已有的一个语料库里面对应识别出来的主题词，然后描述性统计词频、词云图

需要1. 高频词语提取

然后我按照高频词语类型去做地理分布图

这个语料库如图，好像也是作者自己写的