**基于NLP技术的突发事件下群体化心理分析**

一、作品简介

本作品基于人们在微博重大突发事件下发表的评论进行情绪分析，拟通过算法创新和理论研究，更加准确的提取评论中的情绪信息，以及与其相关的情感极性，提高现有技术的准确性和普适性；通过对数据进行地区和群类的划分，更加精确地挖掘出不同地区、不同群类的用户在面对重大突发事件时的心理状态变化，使属性和用户所反映的真实感情倾向匹配度更高，为政府更好地体恤民情、制定措施提供较为可靠的依据。

二、作品意义与目标

随着现代社会的快速发展，人们的心理问题越来越受到重视。有研究表明三分之二以上的人存在或多或少的心理问题，那么毫无疑问，在遇到发生重大突发事件时，人们的心理状况将会受到更大的影响。据中国社会科学院社会学研究所社会心理学研究中心最新调查显示：新冠肺炎疫情出现以来，有79.3%的群众对疫情表示较强烈的担忧，其次有40.1%的群众有较强烈的恐惧，再次有39.6%的群众有较强烈的愤怒心理。随着疫情的持续，社会公众的心理问题成为了一个日渐重要的显性命题。

心理问题作为现下的一个热点，备受大众关注。市面上已经推出了如壹心理等一些较为成熟的心理方面的软件，但其本质都是提供一些基本的心理知识、付费心理咨询。本作品，立足于公众面对当前新冠肺炎疫情的心理状况，旨在分析出一场突然发生的重大事件会对公众的心理产生怎样的影响，影响的程度，怎样的人群易受到较为严重的影响等一系列问题，并将其可视化。而其中，我们将数据地区化、人员群类化，以保证数据与分析结果的针对性和有效性，获取重要信息的集中性和及时性。以此为依据，我们不仅可以在重大事件期间，提供较为精准的地区化、群类化的公众心理分析，还能够在突发重大事件刚发生时就预测出每类人群大概率会产生的心理诉求和心理状态变化，从而提供最为及时的心理引导和积极措施，做到防微杜渐。

三、应用场景

通过地区化、群类化地分析本次新冠疫情中公众的心理状况变化，提取出地区性、群类性的心理诉求和主要心理问题及概率，给政府提供相关的信息，以帮助政府采用一些有针对性地措施，帮助人们恢复疏解心理问题。

预测每次遇到类似重大突发事件人们的心理变化，分析出具体地区、具体人群由于突发事件刺激而可能引发的一定心理问题的公众数量增长，心理变化程度、方向等。从而有方向的为公众提供一定的心理方面的引导、使用积极的措施去改善等。

四、设计思路

我们首先采用Selenium爬虫和基本抓包工具抓取微博用户对于新冠肺炎相关事件的评论，根据获取的到的用户地理位置，首先将用户按地理位置划分开来。然后采用jieba工具进行分词、词性标注，提取情感词来定义情感词典。再使用SnowNLP进行情感分析，通过带有主观描述的中文文本，依据情感词典判断该文本的情感极性类别并根据词频给出相应的置信度，情感极性分为积极、消极、中性，在词性的基础上又将情绪的程度表现为具体的分数，每一个分数段能够统概一种情绪程度。

同样地，我们将提取到的信息按照受重大事件的影响程度不同的人群划分。以本次新冠疫情为例，我们可以将影响人群分为四级：第一级为确诊患者、疫情防控一线医护人员等；第二级为居家隔离的轻症患者、到医院就诊的发热患者以及疑似患者；第三级为与第一级、第二级人群有关的人，如家属、朋友、志愿者等；第四级为受疫情防控措施影响的疫区相关人群、易感人群和普通公众。每一类人群的心理诉求不同，因此可能产生的情感波动或者心理问题范围是也是不同的。再按照上文所述，进行一次情感分析，得出相应的数据和结论。

五、主要创新点

1、对比市面上已有的一些心理方面的软件，我们的作品主要针对于分析面对重大突发事件时人们的心理状态。重大突发事件就像一个超大型催化剂，在它的刺激下，公众的心理状态会有较为普遍的变化。而我们就是要通过大数据将这种变化具体的反映出来，并且能做到在下一次突发事件刚发生时就及时给出预测。

2、中国国土辽阔，各地区、各省、市，甚至是县都存在较大的差异，如果将所有人一概而论，一起分析，那么结果不具有太多的参考价值。以新冠疫情为例，本次新冠的重灾区在武汉，那么武汉人民的心理状态的波动应当是非常大的，我们不能用对全国的数据分析去诠释武汉人民的心理状态。因此，我们将所有数据按照地区分类后再分析，以达到点对点、面对面的地区化心理状态反映。各级政府也能够按照我们的分析针对本地区民众做出相应的措施，使决策更加合理有效，更符合民众心理诉求。

3、面对重大突发事件时，“直接受害者”和医护人员一定是受影响最为直接和最严重的群体，而普通群众可能没有那么大程度的影响。针对重大事件的不同群体，我们通过信息提取后进行分类，也给出了相应的分析，从而得到仅关于事件本身不同的群体所产生的最直接的情感反映和不同群体的心理诉求。

六、关键技术

1、评论数据抓取

首先需要抓取微博上人们关于新型冠状病毒肺炎在全国范围内蔓延的相关评论信息，采用Selenium爬虫和基本抓包工具实现。

**（1）审查网页元素，获取目标网站DOM树结构**

**（2）Selenium抓取评论信息**

**（3）抓取前100页评论并存储至CSV文件**

# 2.情感分析预处理及词云显示

情感分析的基本流程如下图所示，通常包括：

**（1）自定义爬虫抓取文本信息；**

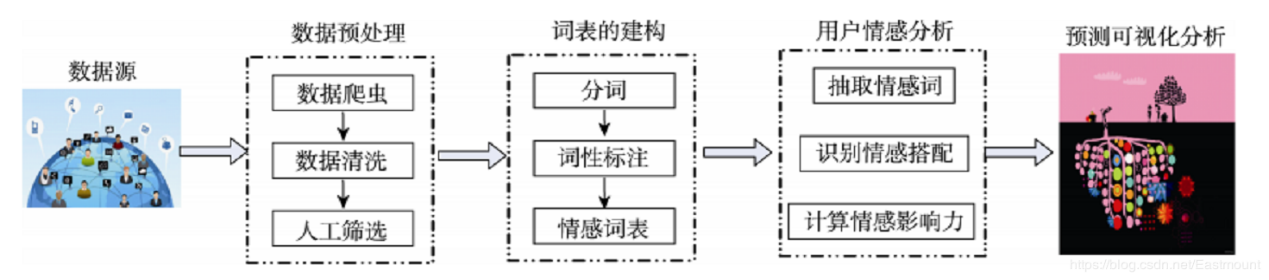
**（2）使用Jieba工具进行中文分词、词性标注；**

**（3）定义情感词典提取每行文本的情感词；**

**（4）通过情感词构建情感矩阵，并计算情感分数；**

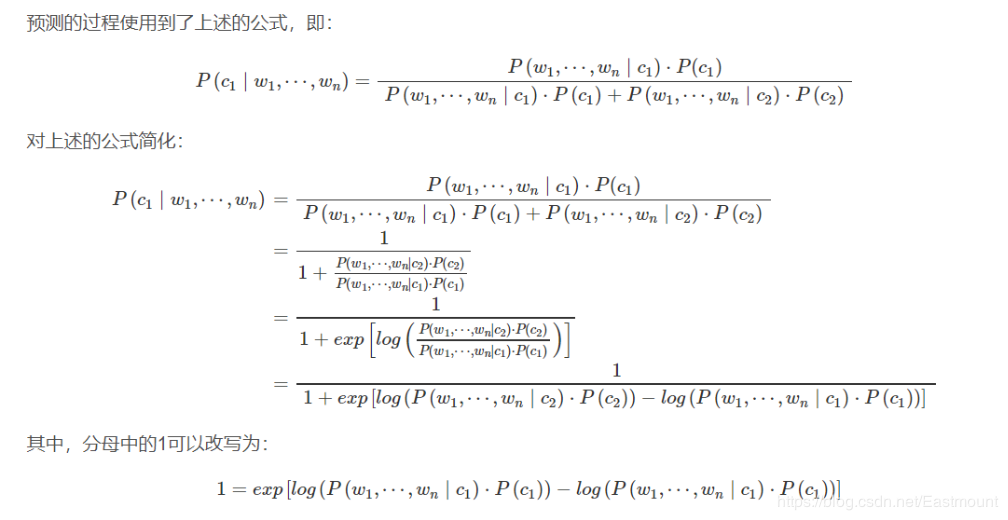
**（5）结果评估，包括将情感分数置于0.5到-0.5之间，并可视化显示。**

将抓取的评论信息复制至TXT文件中 ，每一行为一条评论，再对其进行中文分词处理。注意，这里仅仅获取序号1-1000的情感分数，而其他情感分析可以进行时间对比、主题对比等.



**3.情感分析**

SnowNLP情感分析也是基于情感词典实现的，其简单的将文本分为两类积极和消极，返回值为情绪的概率，越接近1为积极，接近0为消极。原理如下：



# 4.SnowNLP情感分析实例

**（1）情感各分数段出现频率**

**（2）情感波动分析**