第一部分是对外媒的分析：

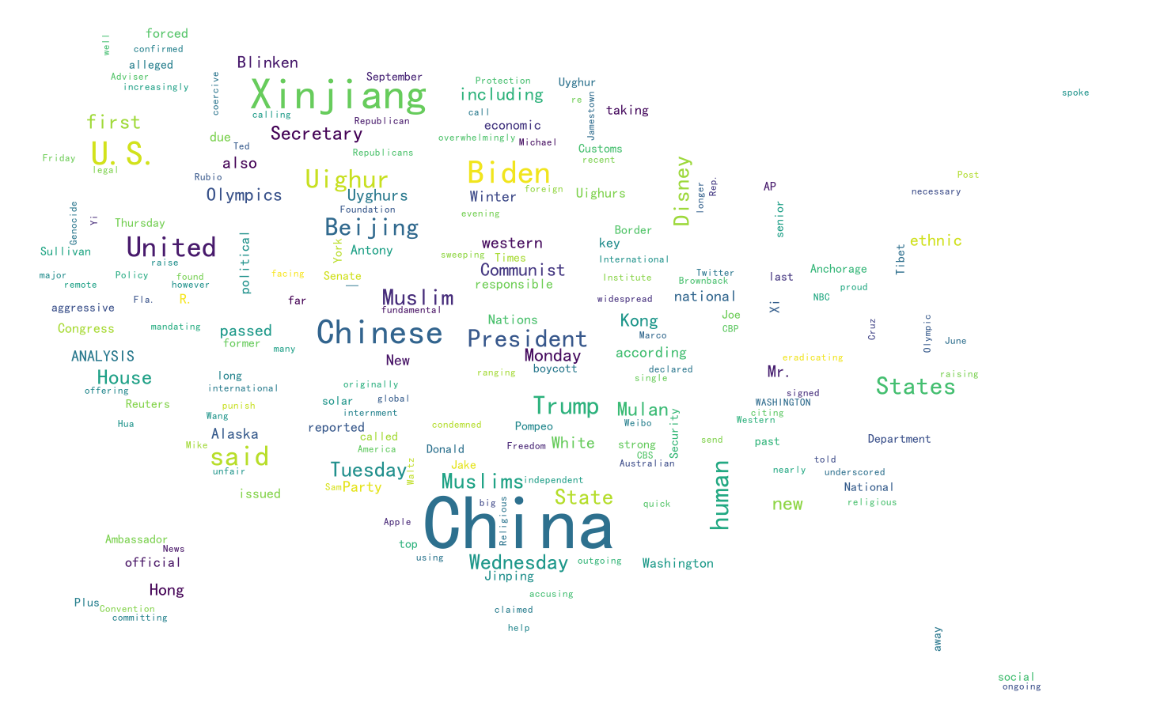
利用allsides网站上带有政治倾向标签的新闻信息，从源头对美国媒体报道的新疆棉事件的相关新闻进行文本分析，分析外媒对于这一事件报道的倾向与客观性。

我们选取了网站上所有关于新疆棉这一事件的新闻进行分析，一共46条新闻，对于报道每条新闻的媒体、每条新闻体现出的政治倾向做如下的可视化。

首先，对每条新闻进行分词、去重、去除停用词、词性划分等预处理，并对词频进行统计，进行可视化。考虑到不同词性的词能体现出的层面有所不同，因此分别对形容词和名词进行词频统计，以形容词为例，得到的图像如下所示：

根据形容词的词频统计发现，美媒在主要道德层面上对中国进行了抨击并持否定态度；同理，根据名词的词频统计结果，我们发现报道的主体以中国和美国的各方面为主，可知报道对中国和美国进行了大量荒谬、别有用心的对比。

最终，我们生成了新闻报道的词云图。

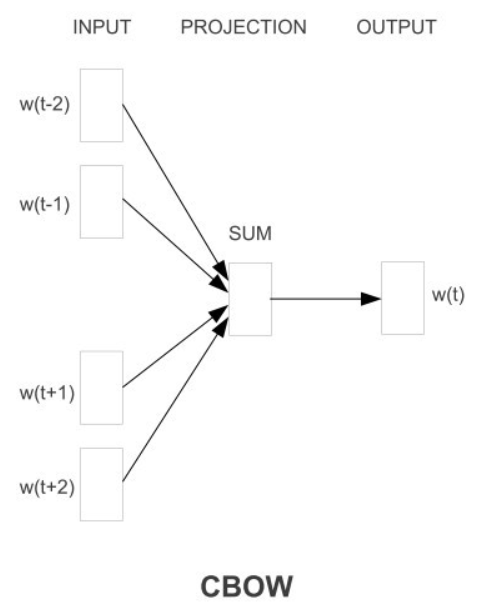


第二部分我们利用深度学习对新闻进行语义分析

P9

我们利用深度学习对语料库中的单词进行了word2vec处理，将单词向量化，并利用Tf-Idf算法确定了每条新闻中每个算法的权重。

对于word2vec，我们使用了CBOW算法实现，即在训练过程中需要完成利用上下文对中间词进行预测的任务。考虑到语料库的大小，我们选择了以2个词作为窗口大小。



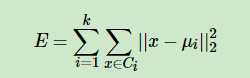
将单词向量化后，我们利用tf-idf算法确定新闻中每个单词的权重。Tf指的是词频，即每个词在该条报道中出现的频率，频率越大说明该词在文档中越重要，权重也应当越大，因此tf值与权重应正相关；idf指的是逆向文档频率，即假设整个文档有n篇文章，而一个词语在k篇文章中出现，则idf值为：**idf=log2(n/k)。因此，idf值越大说明该词对这条报道的贡献越小，权重应该越小，因此权重应与idf反相关。**

**因此，权重的公式应为：**

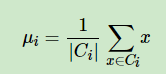
**1619840832(1)**

**P10**

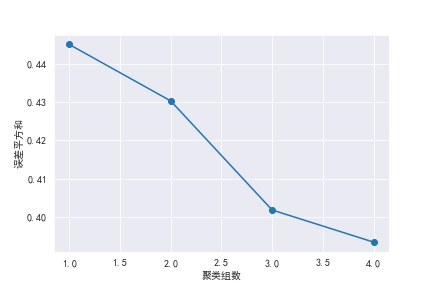
**将单词向量化、每条新闻向量化后，我们对新闻进行了K-Means聚类分析。K-Means聚类是常用的一种聚类方法，其思想在于对于给定的样本集，按照样本之间的距离大小，将样本集划分为K个簇。让簇内的点尽量紧密的连在一起，而让簇间的距离尽量的大。目标是最小化平方误差E：**

****

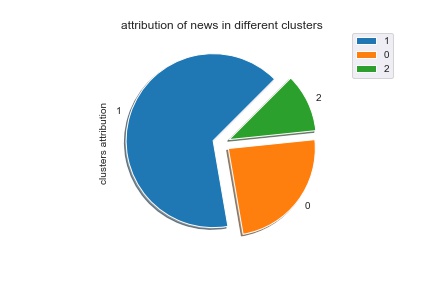
**其中的μ是簇的均值向量，成为质心：**

****

**因此，根据误差平方和曲线**

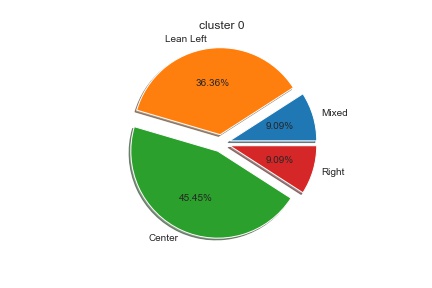
****

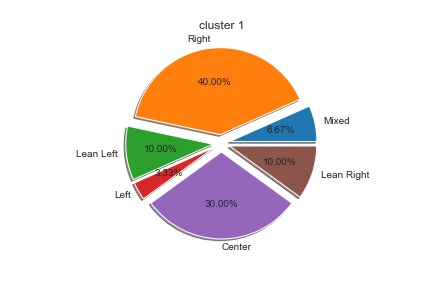
**确定应聚类为三类，得到的聚类结果为（分别为第0类、第1类、第2类）**

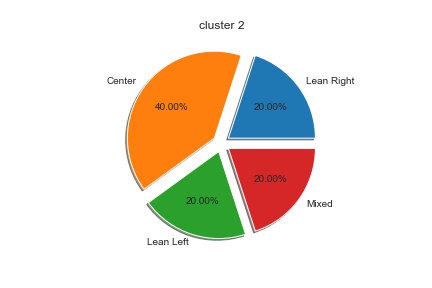
****

**P11**

**接下来，我们将聚类结果与政治倾向联系起来，分析每类新闻中政治立场的分布：**

****

****

****

**发现三类分别为左翼占主导、右翼占主导、中立占主导，发现聚类结果与政治立场具有极高的一致性。然而，聚类过程中仅仅是通过语义进行的，而并未考虑政治立场，因此可以得到结论：美媒的报道受其所处的政治立场的影响，报道存在极大的主观性，不够客观。**

**第三部分是对**国内舆情的分析

**我们通过分析微博相关话题的评论来了解中国民众的对新疆棉事件的反应。**

时间趋势上看：在3月24日HM发布相关报道前，各大品牌相关的微博话题热度一般，只有商家推出活动引发的部分评论。而24日之后，随着各大品牌陆续发布抵制新疆棉的公告，网民抵制国外品牌的声明出现爆炸式增加。

点赞可以看作微评论，集中反映了网民的关注度。日点赞数统计图显示：中国网民短期反应较为强烈，其中，新疆棉、HM、匡威相关话题在24日点赞超过百万，热度最高.但最长高度关注期只能持续3～5天。

转发量可以看作网民对舆论的传播意愿。日转发数统计图显示：对多数话题，转发量与实时热度成正比，且网民的对品牌抵制的转发意愿更强。

微博数随时间累计动图如下：

**地理趋势上看，“新疆棉”相关的话题在新疆本地的反应最为强烈。结合评论内容可以看出，新疆人对国外品牌的无端造谣非常反感。**

**P17**

**“HM”与“优衣库”话题的抵制微博多集中于东部经济较为发达的地区，这些地区也正是HM等相关品牌在中国的主要市场。**

**P18**

**P19**

**统计相关话题的高频词绘制词云发现：**

**新疆棉话题下集中了反对外媒外企诽谤、抵制相关品牌、支持新疆发展、支持国货的讨论。在其他企业的话题下，除了对相关品牌不真诚致歉信的讽刺，还更多的关注了**明星、代言人的解约行为。可以看出在重大舆论事件面前，关键意见领袖的影响力比较大。

**P20**

**对高频词进行语义网络分析，可以更清楚的看出评论的语义倾向。“新疆棉”语义分析网络中：声明、棉花、中国、新疆、抵制、强迫、劳动等是除了“新疆棉”以外的关键节点，主要反映出网民**对抹黑中国强迫劳动企业的抵制。发展、棉花、新疆、国货等高频词反映出民众对国货的肯定与信心。

P21

“匡威”语义分析网络中：匡威、合作、终止三个高频词与其他高频词都有相连，反映出中国民众支持代言人向匡威终止合作的主要语义。

P22

“HM”语义分析网络中：HM、新疆、中国、市场等是最重要的关键词，网民普遍认为HM的举动会因为失去中国市场而亏损。此外，声明和道歉、是非等关键词隐含了网民对HM的公告与道歉声明都有自己的判断，和对HM道歉声明的不认可。企业、人民、政府等关键词都与中国相连，反应出中国全民统一战线。

P23

以上就是我们报告的全部内容，感谢您的观看。