P1

老师好，下面由我们小组介绍新疆棉事件舆情分析及可视化。

2021年3月24日，瑞士良好棉花发展协会与HM、匡威、阿迪达斯等协会会员陆续发表“不采购新疆产品或原材料”的声明引发了中国网民广泛的舆论声讨与商业抵制。这是继2012年以来全球最大的商业抵制活动。它即包含了普通大众、媒体、企业、外交都等多主体间的激烈互动，也展现了出当前网络空间下，线上舆论对线下行为的巨大影响力。

本作品旨在利用各种数据分析方法，以外媒与中国民众为研究对象，可视化新疆棉事件的舆情，直观展示舆论走势与民众的呼声。

chapter1 P2

第一部分是以Allsides新闻网站为例，对事件源头的分析。

P3

Allsides是一家汇集了美国各类媒体新闻报道的网站。其特点在于为每条新闻都标注了左翼、右翼或是中立的政治倾向标签。便于我们将每条新闻的语义内容与其体现出的政治倾向联系起来，分析新闻的客观性。

P4

我们选取了网站上所有关于新疆棉事件的新闻进行分析，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。可以看出，我们分析的新闻媒体来源丰富、政治倾向多样。

P5

对每条新闻分词、去重、去除停用词、词性划分等预处理后，我们分别对形容词和名词分类词频统计。其中，形容词主要用来反映人们关心事件的层面与态度，名词主要反映事件相关的主体。得到图像如下：

形容词的词频统计结果显示，美媒在主要道德层面上对中国进行了抨击并持否定态度。名词词频统计结果显示，高频出现的词大多与中国和美国相关，可见报道对中国和美国进行了大量荒谬、别有用心的对比。

P6

最终绘制的新闻报道的词云图如下。

chapter2 P7

第二部分中，我们运用深度学习的方法对新闻进行语义分析。

P8

对语料库中的单词编码后进行word2vec处理，将单词向量化，并利用词频逆向文本频率算法确定了每条新闻中每个单词的权重。最后使用k-means算法将语义聚类。

P9

根据误差平方和曲线，确定将所有新闻按语义分成0，1，2三类，其中第一类占比最多。

P10

当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时发现：三类分别为左翼占主导、右翼占主导、中立占主导。也就是说聚类结果与政治立场具有极高的一致性。然而，聚类过程仅通过语义进行，并未考虑政治立场。因此可以得到结论：美媒报道的内容受其所处政治立场影响，报道存在极大的主观性。

charpter3 P11

第三部分中，我们基于微博评论分析国内舆情，了解中国民众的反应。

P12

首先从时间维度的统计数据看舆论的传播力与反应趋势。

传播力角度看：相关话题的点赞与转发的统计数最高可达10万数量级，信息传播范围广。话题热度持续长达3～4天。比较Kwak等人的结论“超一半的转发行为发生在Twitter发布1小时内，75％发生在1天内”可以看出，本次舆论的传播周期较长。

反映趋势上看：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。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。

P13

微博数反映话题的关注度，而转发量可以看作网民对舆论的传播意愿。从3月10日到4月20日，两者随时间的累计动图如下。

P14

P15

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，相比于其他地域以表达情感为主的评论，新疆本地人的微博中对事件细节的阐述更多。

P16

“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

P17

这些地区也正是HM等相关品牌在中国的主要市场。

P18

之后我们围绕微博评论内容进行可视化分析。

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。企业的话题下，除了对相关品牌虚伪致歉信的不认可，民众还关注明星、代言人的解约行为。可以看出在类似的重大舆论事件面前，关键意见领袖的影响力比较大。

P19

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：中国、新疆、棉花、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。

P20

“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。

P21

“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。企业、人民、政府等关键词与中国相连，反映出中国全民一心的精神。

P22

因此，在舆论对社会影响力越来越大的当下，我们更应学会使用先进的舆论分析技术应对海量信息、加深舆论的认识水平。既要高效识别各类新闻的语义与立场，有戳破谣言的幌子能力，又要能及时了解到民意，并在合适的时机正确引导舆论的走向，防止正面舆论被谣言绑架。为建设现代化国家和法治政府营造良好的舆论氛围。

以上就是我们展示的全部内容，感谢您的观看。

框架：新闻源头分析->舆论发酵后最终民众反应

新闻源头分析（意义：宏观高效）：1分类词频统计，观察新闻报道涉及的n主体与adj媒体态度

2语料库中的单词编码后进行word2vec处理，每篇报道的内容就以词向量的形式表示。通过逆向文本频率算法对使用attention机制加权，对加权后的词向量聚类。根据误差平方和曲线确定聚成3类效果最好。通过交叉分析语义聚类结果和政治倾向，我们惊奇地发两者存在相关的关系。因为新闻内容的主体是事实报道，不同政治倾向语义的不同恰恰说明了事实报道内容的不同，也就是说存在因为政治倾向导致的选择性报道，即作为源头的新闻并不客观。

深度学习与聚类的原因：1、焦点事件的新闻报道多，采用机器学习手段分析更高效，可以节省人力。2、分析的目标在于根据新闻内容与政治倾向的相关性快速判断报道的客观性。3、如果发现存在不客观报道，我们可以利用前面聚类结果，进一步比较不同语义类别报道内容的区别。

选择性报道举例：

民众舆情分析：github开源的微博爬虫按主题获取数据。时间地理纬度统计数据看趋势、词频语义网络看重点和内容。

意义：统计数据看舆论的传播力和时间地理上的反应趋势、文本分析看民众的关注点与表达的意思。

指导实践：新疆棉之外的其他舆情事件中，可能存在谣言、立场等原因导致舆论分化出几个对立的观点。时间地点维度的统计数据有利于把控负面舆情发展趋势，文本内容上的分析有利于快速了解舆情的内容，为找到合适的时间、合适的内容击破谣言，引导舆论的走向有很大的帮助。