

HPV舆论分析报告第二版

第一部分：使用深度学习分析公众舆论中HPV话题的关注热点和态度

1. 分析公众在HPV相关话题舆论中，对于安德森卫生服务模型所涉及的理论要素的关注度；

本目标旨在了解公众在HPV话题舆论中对于安德森卫生服务模型理论要素的关注程度，有助于研究者深入探讨公众对于健康问题的认知和看法。此外，还能为相关机构制定相关政策和计划提供重要参考。

方法和步骤：

1. 收集数据集。我们将采用网络爬虫技术从新浪微博平台上收集与HPV话题相关的微博评论。为了保证数据的质量和可靠性，我们将对数据进行清洗和筛选。具体地，我们将根据以下几个方面进行筛选：
 - 评论内容必须包含HPV相关的关键词或者话题标签；
 - 评论长度不能少于5个汉字；
 - 评论不能含有敏感或者不适宜的内容；
 - 评论不能是广告或者垃圾信息。

在收集到的评论数据中，我们将根据用户将其进行分类和组织，以便后续分析用户对于安德森卫生服务模型理论要素的关注倾向。

2. 预训练深度学习Bert模型。为了对评论数据进行分析 and 挖掘，我们将使用深度学习Bert模型进行预训练。具体地，我们将利用大型中文语料库和医学语料库对Bert模型进行预训练。通过预训练，我们将获取一个具有强表征能力的模型，可以为后续分析提供更加准确和可靠的基础。具体步骤如下：
 - 收集并准备数据集。我们会收集大量的互联网中中文文本和医学文献，对其进行处理和清洗。
 - 使用预处理工具对数据进行分词、句子分割、Tokenization等操作，生成可以用来训练Bert模型的数据。
 - 使用Pytorch、Huggingface等深度学习框架，搭建Bert模型并在预处理后的数据上进行训练。训练完成后，保存模型权重和参数，以备后续使用。
3. 我们将使用训练好的Bert模型来提取微博评论中的实体。具体步骤如下：
 - 对每个微博评论进行预处理，包括分句、分词、Tokenization等。
 - 对处理后的数据进行编码，并使用Bert模型进行特征提取。
 - 将提取到的特征输入到实体识别模型中，使用CRF等算法进行实体识别和提取。
 - 最终得到的实体包括人物、疾病、药品等与HPV话题相关的实体。
4. 根据安德森卫生模型对实体进行分类。将使用预先定义的规则和分类方法，将提取到的实体根据安德森卫生模型的理论要素进行分类。具体步骤如下：
 - 首先，定义一个基于安德森卫生模型的分类规则，将实体按照其所属的理论要素进行分类。
 - 对每个实体进行特征提取，并根据分类规则进行分类。

- 统计每个用户关注的理论要素及其关注程度，为后续分析用户态度提供依据。

2. 分析公众在HPV话题舆论中持有的情感态度。（中性态度，负面态度，正面态度）；

通过情感态度分析，了解公众对于HPV话题的态度倾向，包括中性态度、负面态度和正面态度。HPV疫苗是预防人类乳头瘤病毒感染和相关癌症的有效措施之一，但由于种种原因，如缺乏知识和误解，一些人可能不接受疫苗接种。因此，了解公众在HPV话题中对于疫苗的态度倾向，可以帮助医学工作者和公共卫生官员更好地了解公众对于疫苗的需求和关注点，有助于制定更有针对性的疫苗推广策略，提高疫苗接种率，进一步预防HPV相关疾病的发生。

方法：

使用预训练好的Bert模型，可以将用户在微博上的评论转化为计算机可理解的向量表示。然后可以使用这些向量作为输入来进行情感分析。

具体实现上，可以使用开源的中文情感分析库（如THULAC、SnowNLP等）或者基于深度学习模型的情感分析库（如TextCNN、LSTM、Bert等）。其中，使用Bert模型进行情感分析的方法较为高效和准确。

对于用户微博评论的情感分析，一般采用以下步骤：

1. 数据预处理：将采集到的微博评论进行清洗、过滤和分词处理，得到符合BERT输入格式的数据。
2. 使用预训练好的BERT模型：将预处理后的数据输入到BERT模型中，获取每个微博评论对应的BERT向量表示。
3. 情感分类：将BERT向量输入到情感分类模型中，对每个微博评论进行正面情感和负面情感分类。可以采用多种分类算法，如逻辑回归、支持向量机等。
4. 统计用户态度：将情感分析的结果分为中性态度、负面态度和正面态度。对每个用户的微博评论进行情感分类后，可以统计每个用户在HPV话题舆论中表达的不同情感态度的频率，分析出公众对于HPV话题的整体情感态度倾向。

3. 分析所关注不同安德森卫生模型的理论要素的人群，在HPV话题舆论中所持有情感态度之间的联系。（如安德森理论模型中要素相似的人群，所持有的情感态度）

可以为制定更有针对性的宣传策略提供参考。在HPV接种宣传中，不同的人群可能对于安德森卫生模型中的理论要素有不同的关注点。因此，通过对不同人群的聚类分析，可以了解到每个人群的特点和需要重点关注的问题。同时，还可以根据不同人群所持有的情感态度，制定更加个性化的宣传语言和方式，从而更好地促进公众的HPV接种意识和接种行为。

方法：

1. 根据第一阶段的方法，将用户根据对安德森卫生模型理论要素的关注倾向进行聚类，得到用户对安德森模型理论要素的关注倾向。根据第二阶段的方法，对用户评论进行情感分析，得到用户在HPV话题中的情感倾向。
2. 根据用户对于安德森模型理论要素的关注倾向进行聚类。根据安德森卫生模型的理论要素，我们可以将用户的关注倾向分成几类，比如“健康信念”，“社会支持”，“自我效能”，“知识认知”等等。针对

每个理论要素，可以统计每个用户在微博评论中对相关实体的提及频次，然后将每个用户对不同理论要素的提及频次向量化。

3. 对于向量化后的数据，可以使用传统的聚类算法，如K-means聚类、层次聚类，或者使用基于深度学习的聚类算法，如变分自编码器聚类。将具有相似向量的用户归为同一类别。聚类算法的目标是最小化同一簇内的距离，同时最大化不同簇之间的距离。经过聚类后，我们可以得到一些聚类簇，每个簇代表着一个用户群体，该群体对不同安德森卫生模型理论要素的关注倾向相似。
4. 我们可以比较不同聚类簇内的情感态度分布是否有显著差异。如果不同簇之间的情感态度分布差异较大，则说明在某些安德森卫生模型的理论要素上有所关注的人群，在情感态度上存在差异。这一信息可以帮助相关机构更精准地制定有针对性的宣传策略，提高HPV疫苗的接种率。

4. 分析不同安德森卫生模型的理论要素对于公众对HPV态度的影响程度。（即哪些因素的提升更能让人群获得积极态度）

HPV话题中的态度倾向与安德森卫生模型理论要素的关系，有助于我们更准确地识别和理解公众的态度，从而采取更加精准有效的干预措施，以提高HPV接种率。通过机器学习模型的可解释性分析，可以深入了解哪些因素对于公众在HPV话题中的态度具有更强的影响力，为政策制定者提供有力的指导。例如，如果模型分析出知识水平的提升对于促进公众对HPV的积极态度具有最大的影响，那么政策制定者可以考虑加强公众的HPV知识宣传和教育。

方法：

1. 建立随机森林决策树模型：随机森林是一种机器学习模型，可以用于分类和回归分析。在此研究中，我们需要建立一个分类模型，以分析安德森卫生模型的要素对于态度倾向的影响程度。我们需要先将数据集划分为训练集和测试集，然后使用训练集来训练随机森林模型，以预测测试集中每个用户的态度倾向。
2. 可解释分析：在这个步骤中，我们需要使用可解释分析方法来理解随机森林模型中每个变量的重要性。这可以帮助我们确定哪些安德森卫生模型的要素对于态度倾向具有重要影响。常用的可解释分析方法包括基尼重要性、平均减少不纯度和排列重要性等。
3. 分析卫生模型要素对于态度倾向的影响程度：在这个步骤中，我们将分析每个安德森卫生模型的要素对于态度倾向的影响程度。例如，我们可以分析“感知障碍”、“行为信念”、“态度”等因素对于态度倾向的影响程度。这可以帮助我们确定哪些要素能够对态度倾向产生积极影响。
4. 总结能对态度产生重大积极影响的卫生模型要素：在这个步骤中，我们将总结哪些卫生模型要素能够对态度产生重大积极影响。这些要素可能包括促进个人信仰和态度变化的行为信念、提高认知水平的教育和宣传、降低接种障碍的政策和服务等。