

信息网络研究专题课之应用层

一、文献信息

1. 论文题目: Assessing Viewer's Mental Health by Detecting Depression in YouTube Videos

2. 作者: Shanya Sharma, Manan Dey

3. 发表途径: NeurIPS

4. 发表时间: 2019

二、问题意义

情绪低落的人倾向于沉思并观看令人沮丧的视频,以便获得认可或找到相同想法的同伴,这种行为模式在确定观看者的心理状态方面可能很有用。将对他观看的视频类型进行分析,以找出令人沮丧的内容和模式,这可用于监视推荐并根据观看活动限制触发视频的推荐。如果这些模式显示出不健康的精神状态征兆,我们的工具旨在通过警告用户其情绪变化点来提供早期诊断,并将其引导至正确的门户以寻求帮助。

在本文中,开发并测试了将机器学习技术应用于通过成绩单捕获的 YouTube 视频内容的功效,并确定视频是否令人沮丧或令人沮丧。本文还引入了一种现实生活中的评估技术,通过计算评论的 CES-D 分数,基于视频上发布的评论来验证我们的分类。

三、思路方法

本文先从用户至少 24 天的缓存中对观看历史的视频来提取关键的信息,再对视频进行抑郁/非抑郁的分类,之后对 CES-D 值进行预测,若预测值大于相应的阈值,则通过 youtube 的官方界面给予消息提示反馈给用户。

本文主要包括以下三层结构:

1. 安全的数据采集。
2. 从单个视频中提取特征及其内容分析
3. 随时间推移分析视频观看模式。

本文重点关注以下情况,用户关注具有相同情感内容的视频,长时间执行重复性的反省,并将此情况列为抑郁症状,文中我们将重点放在第二层,即基于视频中提取的文字特征来分析视频内容,以构建分类器,以识别视频是否可被称为抑郁或不抑郁。

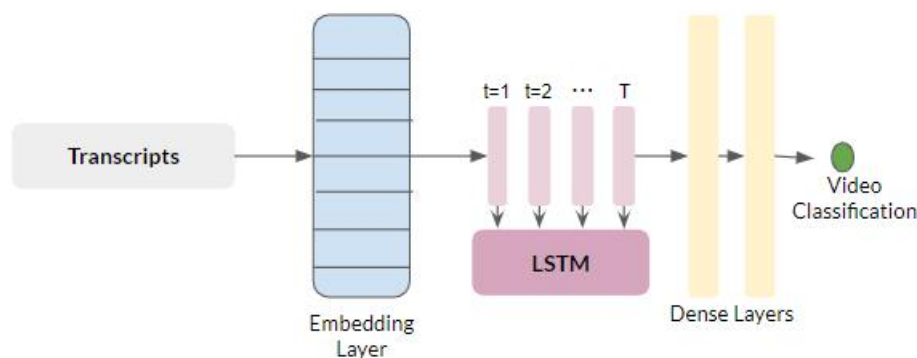
通过使用各种关键字提取视频并从中提取转录本来收集用于分析视频内容的数据。通过关键词标签的判别来对视频进行分类,关键字已经在文中附录呈现。

以下是对视频进行分类的两种深度学习算法:

长度为 N 的朴素贝叶斯算法: 本文使用从 Empath 模型中提取的特征,构建了多项朴素贝叶斯分类器,提供了是一组词汇类别的归一化分数。除了例如正面和负面情绪之类的预定义功能外, Empath 模型还提供了创建自定义的功能,还创建了与 CES-D 调查表诊断的症状有关的种子术语的类别,并将其包括在类别集中,便于我们之后对用户的评论做 CES-D 评估。

LSTM：长短期记忆，是一种特殊的 RNN（循环神经网络），主要是为了解决长序列训练过程中的梯度消失和梯度爆炸问题。LSTM 内部主要有三个阶段：1. 忘记阶段。这个阶段主要是对上一个节点传进来的输入进行选择性地忘记。2. 选择记忆阶段。这个阶段将这个阶段的输入有选择性地“记忆”。3. 输出阶段。这个阶段将决定哪些将会被当成当前状态的输出。LSTM 通过门控状态来控制传输状态，记住需要长时间记忆的，忘记不重要的信息；而不像普通的 RNN 仅有一种记忆叠加方式，适用于很多需要“长期记忆”的任务。

本文将经过预处理的每个注释通过嵌入层，该层填充有经过预先训练的 GloVe 300D 词嵌入权重，再将其输入到具有 196 个单位和 tanh 激活的 LSTM 网络中，并应用 0.2 的递归下降。然后，将 LSTM 的输出馈送经过两个密集层，中间之间有一个 0.5 的滤除层，从而输出视频的分类，本文使用了二进制交叉熵损失来加快权重更新速率和 Adam 优化处理来自适应的改变参数的学习速率。



为了验证和评估分类器产生的结果，我们为每个视频引入了一种称为 CES-D 的评分方法，以分析视频的压抑程度。分数是根据 CES-D 量表中考虑的各种症状（失眠，自恨，食欲等）得出的术语密度而得出的，在给定文本中以否定含义出现。

每个种子项生成的类别合并为一个集合，遍历集合中的每个单词并为其计算词频，并通过将每个单词的词频相加来计算合计词频，之后把存在正面和负面含义集合中的单词注释发送到 Empath 模型，分析得到正面和负面情绪词汇的得分。然后，将生成的正面和负面情绪的得分作差再与词频相乘得到 CES-D 值。如果对于评论中正面得分超过负面得分，则 CES-D 得分计算为 0。为了计算整个视频的 CES-D 分数，为了确定评论中的压抑强度，将前面计算的总和累加到所有评论中，并通过除以评论总数为最终得分。

四、实验结论

对于不同算法的仿真结果比较：Empath 模型 + NB 算法的分类准确性较低，Empath 模型与 TF-IDF 功能的组合所产生的分类精度与 LSTM 基本相同，而前者消耗的时间较少，实用性更高。

CES-D 仿真分析：本位通过对每个类别（情绪关键字）计算其 CES-D 得分以累积的方式分析了评论，在比较上述各种类别时，令人沮丧的视频与非令人沮丧的视频的 CES-D 平均得分之间的差异非常明显。非令人沮丧的视频的非令人沮丧的评论（CESD = 0）的百分比比较高，在此类评论中非抑郁类别比抑郁类别要高得多。然而，CES-D 分数的范围在抑郁视频中最大，

令人沮丧的评论（高 CES-D 分数）在抑郁视频中占多数。已经证明，在悲伤/沮丧中，人类倾向于倾向于可以找到与自己的心理状态相似的群体。因此可以推论出，经常观看高 CES-D 分数的视频（抑郁视频）是精神状态不健康的指标。

分类器效果评估：将评论获得的 CES-D 值认为是真实值，用于分类器的效果评估。本文在构造较少的域中测试了结果，向该模型提供 1500 个随机视频的副本，并将分类与通过评论获得的 CES-D 得分进行比较。得出为训练目的收集的抑郁视频的 CES-D 分数平均值，并把 CES-D 的阈值设为 20，大于 20 的 CES-D 得分被认为是令人沮丧的。通过实验可得出结论，基于这种方法的分类器的准确性为 84%，因此，本文的视频分类方法被证明是较为真实有效的。

五、启发思考

1. 本文获得的结果是使用文本特征的，对 YouTube 上的视频进行分类，再对用户的心理健康进行分析的。未来的工作可以基于音频和视频为特征，使用相应的深度学习算法和优化方法，对视频进行分类。

2. 也可换一个角度，仅对用户的评论做 CES-D 值评测，首先将视频分类，将正面评论赋予正值，负面评论赋予负值，通过关键词等来区分。再将上值经过某种算法在不同类型的视频中赋予以不同的权重，除以评论总数来得到最终的 CES-D 值，通过大数据分析得出区分心理是否抑郁的 CES-D 阈值，从而判断出人的心理健康状况。这其中考虑到用机器学习算法对评论的筛选，通过关键词来筛选出带有情感色彩的词汇，，因为很多无关情感的评论对于我们的心理分析是较为无效的数据。

3. 未来的工作中还可以使用唤醒维、效价维、支配维等将情感刻画为一个多维信号，与传统的离散情感模型相比，具有表示情感的范围广、能描述情感的演变过程等优点，单个模态的信息存在信息不全面、容易受噪声干扰等缺陷，多个模态的信息能够互相印证、互相补充，从而可以为情感判断提供更加全面准确的信息，提高情感判断的性能。

4. 本文验证分类器的准确性时还不够严谨，因为 CES-D 值和分类器都是根据关键词的判别来区分的，若想做出更精准的判别，可以随机找一些已知心理状态的人们来实际实验他们的 YouTube 的观看历史，这需要征求本人的同意才可实验，因此从道德上来说还需考量。