

## 简介:

对广告进行识别是一种语义视频分类问题。特定平台上的广告是视听镜头的组合，其特征在于视听呈现，因此需要从视频镜头中提取的各种视听特征。抖音视频不遵循特定的视频样式，具有大的可变性和动态性质，呈现出具有挑战性的机器学习问题。

广告数据集包括5次采样、总长度为150小时的抖音视频中提取的视频镜头的标准视听特征，以270 fps的分辨率录制视频，分辨率为720 X 576。将视频数据处理为视频的时长、声音频谱、视频光谱、文字分布和画面变化等特征，以判断其是否为商用广告。最终的数据包含1个标签、230个特征。

## 数据集信息:

数据提供了129685份视频的信息，储存在commercial\_vedio\_data.csv文件中，其中labels为标签。

## 特征信息:

数据集共包230个特征，涵盖视频的时长、声音频谱、视频光谱、文字分布和画面变化等方面。

视频镜头用作生成样本的单元。视频使用RGB颜色直方图将连续视频帧之间分割成视频镜头。每个视频镜头中我们提取了7个音频特征（即短期能量，零交叉率，光谱质心，光谱通量，频谱滚降频率，基频和MFCC音频词包）和5个视觉特征（即视频镜头长度，每个视频镜头的屏幕文本分布，运动分布，帧差异分布，边缘变化率）。

音频特点： -

一般来说，为了吸引观众的注意力，广告具有更高的音频幅度，适当的背景音乐（相对更高的频率）以及从一种音乐到其他音乐或音乐到语音等的尖锐过渡。通过短时能量（STE），过零率（ZCR），光谱质心，光谱通量，光谱滚降频率和基频尝试区分出这些特点。所有这些短期音频特征都是在8000Hz采样频率下以20毫秒的音频帧大小计算的。在镜头上计算所有音频特征值的均值和标准偏差。MFCC音频词包已经成功应用于几种现有的语音/音频处理应用中。我们从150小时的音轨计算MFCC系数以及其他指标并聚集成4000组，形成音频词。

视频功能：

商业视频镜头通常具有较短的长度，快速的视觉过渡以及覆盖文本带的特殊位置。计算文本分布时，将视频帧将划分为大小为5 X 3（15个网格块）的网格，对网格块中存在的文本区域的分数求平均以获得文本分布特征。运动分布，帧变化分布和边缘变化率等特征也可捕捉商业镜头的动态特性。

由于上述处理，导致特征矩阵较稀疏，需要思考如何进一步处理。

包含大约63%的商业样本（阳性）。

不同特征及名称如下

标签： + 1 / -1（广告/非广告）

特征及其序号

拍摄长度

1

运动分布（均值和方差）

2 - 3

帧差异分布（均值和方差）

4 - 5

短时能量 (均值和方差)

6 - 7

ZCR (平均值和方差)

8 - 9

光谱质心 (平均值和方差)

10 - 11

光谱滚降 (均值和方差)

12 - 13

光谱通量 (平均值和平均值) 差异)

14 - 15

基本频率 (均值和方差)

16 - 17

运动分布 (40个分档)

18 - 58

帧差分布 (32个分档)

59 - 91

文本区域分布 (15个区间平均值和15个区间用于方差)

92 - 122

音频字组 ( 4000箱)

123 - 4123

边缘变化率 (平均值和方差)

4124 - 4125