**直播视频特征提取项目合约**

**特征列表：**

音频：

1. 关键词统计：欢迎，谢谢，你好
   1. Feature：关键词出现的次数
2. 声音场景分类：安静，说话声，背景声
   1. Feature：每个场景的duration
   2. 注：Shenggao自己可以完成样本标注和模型训练，应该不需要其他RA协助；背景声(音乐)可能跟说话声同时存在，需要注意

视频：

1. 人脸的基本表情分类：neutral, happiness, sadness, fear, disgust, surprise and anger
   1. Feature：每个表情的duration
2. 手部监测：
   1. Feature：手部出现的duration

输出文件格式：CSV

1. 一个CSV文件，包含所有视频（属于同一个session的视频要先合并起来）
2. 所有feature都按每分钟提取（每一行对应一分钟）

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Room | Session | Performer | Minute Index | Feature 1 | Feature 2 | ... |
|  |  |  | 1 |  |  |  |
|  |  |  | 2 |  |  |  |
|  |  |  | 3 |  |  |  |

**完成期限**：

* 15/11/2016前完成视频的两个feature提取；
* 05/12/2016前完成音频的两个feature提取；
* 每周五简单更新一次进度

**提交内容：**

* 程序：里边需要有必要的注释
* 文档：关于程序里用到的算法说明，以及程序具体执行的步骤
* 提取的特征表：对每个直播session（通过room，performer，session三个id定义），每一分钟内各个特征的信息，如上表所示。

**支付说明：**

* 所有费用，一起一共是10000新币；NUS方面支付7000新币，UNC方面支付3000新币。
* NUS方面只能以RA的方式支付, 暂定按月计算小时，定期结算的方式支付到声高和张力太太两个账号。之后视情况定，可以增加其他的支付账号；UNC方面的支付待定；
* 若不能在15/11/2016提供视频特征，或不能在05/12/2016提供音频特征，每晚一周，扣1000新币；
* 支付确保在任务完成后一个月内完成，如果逾期可适当支付拖延补偿（5%，即500新币）。



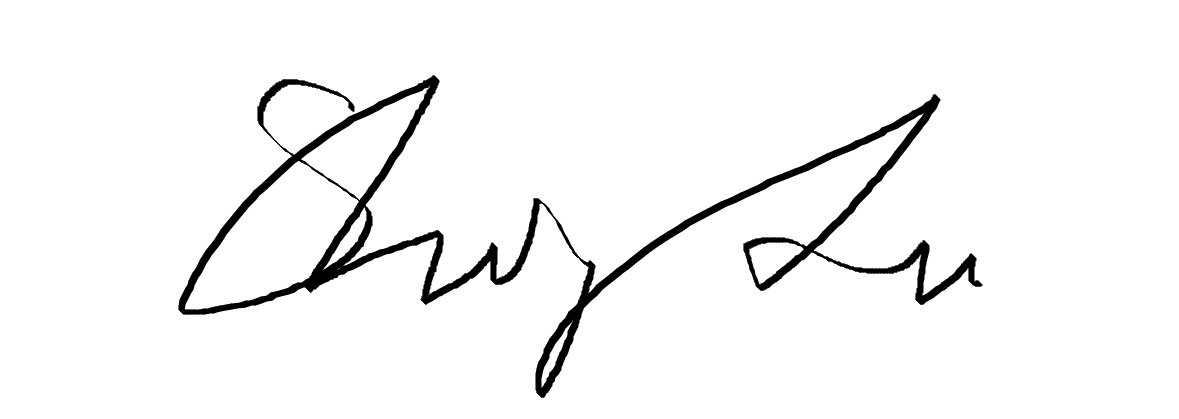
Dai Yao

Assistant Professor of Marketing

NUS Business School

National University of Singapore

Email: dai.yao@nus.edu.sg



Shijie Lu

Assistant Professor of Marketing

UNC Kenan-Flagler Business School

Email: Shijie\_Lu@kenan-flagler.unc.edu

<声高，张力签名>

1. Speech to text, 再加上关键词统计：
   1. **欢迎，谢谢**，你好...

>> 只需要能识别欢迎和谢谢就好。如果能比较容易实现完整的speech to text，那也算上吧，这样text, audio, video的数据都全了。如果需要花太多时间可以先不做。

1. 场景分类：
   1. 可以大致分为3类：
      1. **安静**
      2. **说话声**
      3. **非说话的声音（主要就是音乐，但也可能是背景声音等）**
   2. 说明：
      1. 这3类的分类器需要我自己先标注一些样本，训练模型，比较耗时，准确度估计可以达到70%以上
      2. 笑声一般比较少，而且夹杂在说话中，我自己训练模型估计很难区分，就不做独立的分类了。可以尝试从关键词中找“哈哈”等来代替。

>> 每分钟，安静的时间有多少；说话的时间有多少，其他时间(背景)有多少

>> 背景声和说话声可能重合

>> 标注的话，如果可以做个简单的工具，我可以找5-10个左右本科生RA一起参与标注。

1. Speaker Emotion分类：
   1. 根据说话的声音，分为两类：
      1. Neutral
      2. Emotional (在直播的环境下主要就是happy、exciting等)

>> emotion主要通过面部表情来分析，这里不需要

视频方面：

1. 统计人脸的pose

可以分为yaw， pitch, roll

>> 人脸的pose不是必须。如果有现成package可以调用，不需要标注和训练，那就做一下。

2） **统计人脸的基本表情**

**可以分为： smile，disgust, anger, surprise, sad, fear, neural**

统计：表情持续的时间

>> Marketing的literature里最近有一篇paper提到了两种方案，一种是7类(neutral, happiness, sadness, fear, disgust, surprise and anger)，一种是3类(positive, negative, and neutral). 最好能提取7类。

Lu et al (2016): <https://www.dropbox.com/s/1ud2h3dxcb6o2d4/Lu%20et%20al%20(mksc%202016%20-%20a%20video-based%20automated%20recommender%20(VAR)%20system%20for%20garments).pdf>

Lu et al (2016 - appendix): <https://www.dropbox.com/s/91vlmisssptkjt2/Lu%20et%20al%20(mksc%202016%20-%20appendix).pdf>

3） 皮肤部分的运动（手是否出现在屏幕之中）

>> 每分钟内，手出现的时间：比如一秒抽取1帧，如果一分钟内的帧数有20帧监测到手，那手出现的时间就是0.333分钟

>>>>> 另外，对各个feature，最后的数据都应该是minute-by-minute的数据，而且需要能跟索引文件里的三个id对应。