周报

这一周我看了《acoustic scene classification--classifying environments from the sounds they produce》,关于ASC有了一些新的收获和体会。

在这篇文章中提出了与声学场景分类相关的较新的研究——根据声音进行环境分类。文章从回顾以前在这方面的的研究着手，为ASC定义了一般框架，提出了其组件的不同实现方法。然后，文章中为解决数据挑战描述了一系列不同的提交算法，从而为ASC技术提供全面和公平的基准。也以此为目的记录了数据集和提出了用来评估比较提交方法的算法和统计显著性测试的执行标准。文章中使用基线方法，该方法采用MFCC，GMM和最大似然准则为基准。文章还评估了人类在执行相似分类任务时的分类精度。最好的执行算法的平均精度符合人类得到的平均精度，并且常见的几组分类都被计算机和人类错误识别。然而，所有的声学场景可以被一部分人正确分类，而有的场景被所有算法错误分类。

首先，文章介绍了声学场景分类（ASC，声音场景分类（ASC）指的是将语义标签与标识其生成的环境的音频流相关联的任务）和计算机听力场景分析（在环境中的声音会产生各种听觉场景，即发生在空间不同位置上的声音流所形成的“事件”分布。通过听觉系统的复杂处理，人在脑中会对这一场景中叠加的各种声音事件加以分类，即所谓的场景分析。对人的这种能力的研究及其功能建模，是计算听觉场景分析的任务；他对解决盲源分离及语音增强等工程问题 有指导意义）；提出了声学场景分类可以应用于上下文感知服务的设计，智能可穿戴设备，机器人导航系统和音频档案管理。ASC可能启用的未来技术的具体例子包括能够感知周围环境的智能手机，每次进入音乐厅时都会自动切换到静音模式。辅助技术，如能根据室内或室外环境的认识调整其功能的助听器或机器人轮椅，或储能自动将元数据分配给音频文件的声音档案。

从纯科学的观点来看，ASC代表了一个人类和机器都只能在一定程度上解决的有趣的问题。从一开始，一个声学场景或者语音环境的语义标签就是一个拥有不同的解释的任务，因为没有一个全面的分类涵盖能所有可能的类环境。研究人员定义了一系列的类，记录来自这些环境的样本，并将ASC视为可能类的封闭环境中的监督分类问题。此外，即使在预先定义的类中，声学事件或性质表征一定的环境通常是无限的集合，使得它很难得到规则的明确的计划声学事件或特征去区分场景。

然后文章介绍了ASC的历史背景。首次出现专门针对ASC问题的方法的文献，是由Sawhney和Maes在麻省理工学院（MIT）实验室所提出的1997技术报告。作者记录了一组包括一群人、声音、地铁和交通的数据集。他们利用从语言分析和听觉研究中得到的工具来从音频数据中提取了几个特征，采用神经网络和K近邻准则为特征和类别之间的映射关系建立了模型，并获得一个68%的整体分类精度。一年后，来自同一所大学的研究人员通过在几次汽车去超市的过程中头戴麦克风记录了一个连续的音频流，然后自动地将音频分割成不同的场景（如一个家，一个街道，和一个超市）。在分类方面而言，他们把从音频特征提取的合适的经验分布分配到隐藏的马尔可夫模型（HMM）。

文章中也提出了许多音频特征，他们包括：

1. 基于低水平时间和频率的音频描述符：有的ASC系统的使用在时域或其傅里叶变换域可以容易计算的特征。这些包括（其中包括）过零率，过零率可以测量单信号改变的平均速率，与单声道的主要频率有关；光谱质心，它可以测量光谱的质量中心，与亮度的感知有关。

2、能量带（能量/频率）：这类特征被各种ASC系统调用，通过在特定频率带宽上的幅度谱或能量谱计算得到。得到的系数可以测量在不同子带的能量数，也可以表示为子带能量和总能量之间的比率y，对信号中最突出的频率区域进行编码。

3、听觉滤波器组：能量/频率特性的进一步发展包括通过模仿人类听觉系统反应的滤波器组分析音频帧。Sawhney和Maes用Gammatone滤波器来实现这个目的[ 42 ]。克拉克森等则计算Mel比例计算滤波器系数（MFCs）[ 12 ]，而Patil和elahili [1，PE] 采用所谓的听觉的光谱.

4、倒谱特征：MFCCs是倒谱特征的其中个例，也有可能是ASC分析中最常用的特征。他们是通过Mel比例计算滤波器系数对数算法的离散DCT变换得到。单词“倒谱”是“光谱”和“象征”的变形组合，表明这一类的特点是采用傅里叶变换相关的信号频谱的计算。倒谱特征捕捉规范谱包络的声音，从而，总结其粗谱内容。

5、空间特点：如果声景已经被多个麦克风记录，那么其特征可以从不同频道中提取，然后与适当的声学场景相匹配。在一段立体声录音中，最常用的特征包括双耳时间差（ITD），可以测量左声道和右声道之间声源的相对延迟。和两耳间的水平差异（ILD），可以测量两种通道间振幅的变化。ITD和ILD都与在立体声场中声源位置有关。Nogueira等将空间特征包含在他们的ASC系统中。

6、浊音特性：当信号被认为含有谐波分量时，可以估计基频F 0或一组基本频率，并且可以定义特征组来测量这些估计的准确性 。就ASC来说，谐波分量可能对应音频场景中发生的具体事件，并且谐波分量的识别可以帮助区分不同的声学场景。盖革等人采用与系统中每一帧的基本频率相关的语音特征，[ 1，GSR ]。这个基于提取拟合特征的方法是被krijnders和Holt[ 1，KH]提出的，一系列的序列特征来源于音频信号中的听觉动机。首先，借鉴人工耳蜗特性的一个所谓的耳蜗图被计算来提供声学场景中在时域—频域中的表征。然后，每一次时域–频率范围的无音调与声学场景中音调事件的识别有关，造成色调配特征向量。

7、线性预测系数（LPC）：这类功能已经应用于语音信号建模为自回归过程的分析。自回归模型，样本一在给定时刻的信号S被表示为在前一时刻的样本的线性组合。



在组合系数确定模型参数和ͼ是剩余期限。有价值的建模得到的频谱包络之间的映射关系；因此，编码关于声音的一般光谱特性的信息。埃罗宁等人。在他们建议的方法中使用LPC特征[ 20 ]。

8、参数逼近特征：自回归模型是近似模型的一种特殊情况，在该模型中，信号被表示为j基函数与集合的线性组合。



当基函数被参数集参数化时，特征可以根据有助于逼近信号的函数来定义。例如，储等人。使用Gabor变换分解音频场景，该变换是每个基函数由其频率f、它的时间尺度u、它的时间偏移量及其频率相位来参数化的表示。识别非零系数的索引集对应于一组活动参数，有助于逼近信号并对音频场景中的事件进行编码。发生在特定的时间-频率位置。Patil和Elahili也提取特征参数从二维（2-D）听觉谱图和Gabor小波网络之间的卷积过滤器。

9、无监督学习：模型假设一组基函数定义了先验去分析信号。另外，基础可从数据或其他以无人监督的方式提取的特征中获得。Nam等采用弱监控波尔曼机（SRBM）自适应地从训练数据的MFCC中提取特征。SRBM是一个从输入图像中学习基础功能的神经网络，类似于人脑视觉接收器构建的表象性质。在ASC的背景下，SRBM自适应性的对训练信号的听觉的基本适应性编码并返回一系列从MFCCs中提取的特征序列和用于决定包含重要声学事件的时间单元的激活函数。

10、矩阵分解方法：音频应用矩阵系数的目的是描述声学信号的谱图作为一个捕捉典型的或显著的光谱元素的初等函数的线性组合。因此，它是无监督学习的一个特征。能够证明在分类中使用矩阵分解法正确的最主要的直觉是在声学场景识别中十分重要的事件标志在初等函数中应该被编码，并导致辨别学习。Cauchi采用非负矩阵分解（NMF）[ 8 ]，benetos等利用概率潜在成分分析他们提出的算法[ 6 ]。请注意，矩阵分解还输出一组激活函数，它们及时对基本函数的贡献进行编码，从而对属性进行建模。因此，这类技术可以考虑共同估计局部参数和全局参数。

11、图像处理功能：Rakotomamonjy和Gasso 为ASC的设计的特征提取函数的算法包括以下操作，首先，音频信号对应的每个训练场景使用Q值变换处理，Q值变换的返回频率表示对数间隔频带。然后，从Q值的表示中通过时间–相邻频点插值得到512\*512 #像素的灰度图像。最后，通过计算局部梯度直方图矩阵，从图像中提取特征。这是通过将图像分割成局部的补丁，定义一组空间定向的方向，并计算显示每个方向的边的出现。注意，在这种情况下，特征向量并不是从每一帧而是从Q值变换的时间-频率坐标中提取的

12、事件检测和声学单元描述符：heittola等人提出了一种基于ASC分类的系统，这种系统可以将基于检测信号的事件直方图的音景分类，在训练阶段，人工标注的事件发生（如汽车喇叭声、掌声、或篮球弹跳）是用来获取每个场景类模型。在测试阶段，采用隐马尔可夫模型识别未标记记录中的事件，并定义一个与来自训练数据的直方图相比较的直方图。这个系统代表了另一个常见的包括特征、统计学习和决策标准的框架，因为它本质上同时执行事件检测和ASC。然而，为了本教程的目的，声学事件可以被认为是由直方图描述其统计特性的高级特征。Chaudhuri等采用了类似的方法学习声学单元描述符（AUDs）和分类 YouTube多媒体数据。声学单元描述符由隐马尔可夫模型得出，用来将一段录制音频记录到事件序列。这个转录假定是由参数在不停的音景类别训练而成的N元语言模型生成。因此，在测试阶段，未标记录音的转录是按最大似然准则进行分类的。

以及各种模型

1、描述统计学(descriptive statistics)：是研究如何取得反映客观现象的数据(data tabulation)，并通过图表形式对所搜集的数据进行加工处理和显示(data visualisation)，进而通过综合概括与分析(statistical summaries)得出反映客观现象的规律性数量特征的一门学科。

2、高斯混合模型：高斯混合模型就是用高斯[概率密度函数](https://baike.baidu.com/item/%E6%A6%82%E7%8E%87%E5%AF%86%E5%BA%A6%E5%87%BD%E6%95%B0" \t "https://baike.baidu.com/item/_blank)（[正态分布](https://baike.baidu.com/item/%E6%AD%A3%E6%80%81%E5%88%86%E5%B8%83" \t "https://baike.baidu.com/item/_blank)曲线）精确地量化事物，它是一个将事物分解为若干的基于高斯概率密度函数（正态分布曲线）形成的[模型](https://baike.baidu.com/item/%E6%A8%A1%E5%9E%8B" \t "https://baike.baidu.com/item/_blank)。高斯模型就是用高斯[概率密度函数](https://baike.baidu.com/item/%E6%A6%82%E7%8E%87%E5%AF%86%E5%BA%A6%E5%87%BD%E6%95%B0" \t "https://baike.baidu.com/item/_blank)（[正态分布](https://baike.baidu.com/item/%E6%AD%A3%E6%80%81%E5%88%86%E5%B8%83" \t "https://baike.baidu.com/item/_blank)曲线）精确地量化事物，将一个事物分解为若干的基于高斯概率密度函数（正态分布曲线）形成的模型。 对图像背景建立高斯模型的原理及过程：图像灰度直方图反映的是图像中某个[灰度值](https://baike.baidu.com/item/%E7%81%B0%E5%BA%A6%E5%80%BC" \t "https://baike.baidu.com/item/_blank)出现的频次，也可以以为是图像灰度[概率密度](https://baike.baidu.com/item/%E6%A6%82%E7%8E%87%E5%AF%86%E5%BA%A6" \t "https://baike.baidu.com/item/_blank)的估计。如果图像所包含的目标区域和背景区域相差比较大，且背景区域和目标区域在[灰度](https://baike.baidu.com/item/%E7%81%B0%E5%BA%A6" \t "https://baike.baidu.com/item/_blank)上有一定的差异，那么该图像的灰度直方图呈现双峰-谷形状，其中一个峰对应于目标，另一个峰对应于背景的中心灰度。对于复杂的图像，尤其是医学图像，一般是多峰的。通过将直方图的多峰特性看作是多个高斯分布的叠加，可以解决图像的分割问题。

3、隐藏的马尔可夫模型（HMM）:是一种统计分析模型，隐[马尔可夫模型](https://baike.baidu.com/item/%E9%A9%AC%E5%B0%94%E5%8F%AF%E5%A4%AB%E6%A8%A1%E5%9E%8B" \t "https://baike.baidu.com/item/%E9%9A%90%E9%A9%AC%E5%B0%94%E5%8F%AF%E5%A4%AB%E6%A8%A1%E5%9E%8B/_blank)是[马尔可夫链](https://baike.baidu.com/item/%E9%A9%AC%E5%B0%94%E5%8F%AF%E5%A4%AB%E9%93%BE" \t "https://baike.baidu.com/item/%E9%9A%90%E9%A9%AC%E5%B0%94%E5%8F%AF%E5%A4%AB%E6%A8%A1%E5%9E%8B/_blank)的一种，它的状态不能直接观察到，但能通过观测向量序列观察到，每个观测向量都是通过某些[概率密度](https://baike.baidu.com/item/%E6%A6%82%E7%8E%87%E5%AF%86%E5%BA%A6" \t "https://baike.baidu.com/item/%E9%9A%90%E9%A9%AC%E5%B0%94%E5%8F%AF%E5%A4%AB%E6%A8%A1%E5%9E%8B/_blank)分布表现为各种状态，每一个观测向量是由一个具有相应概率密度分布的状态序列产生。所以，隐[马尔可夫模型](https://baike.baidu.com/item/%E9%A9%AC%E5%B0%94%E5%8F%AF%E5%A4%AB%E6%A8%A1%E5%9E%8B" \t "https://baike.baidu.com/item/%E9%9A%90%E9%A9%AC%E5%B0%94%E5%8F%AF%E5%A4%AB%E6%A8%A1%E5%9E%8B/_blank)是一个双重随机过程----具有一定状态数的隐[马尔可夫链](https://baike.baidu.com/item/%E9%A9%AC%E5%B0%94%E5%8F%AF%E5%A4%AB%E9%93%BE" \t "https://baike.baidu.com/item/%E9%9A%90%E9%A9%AC%E5%B0%94%E5%8F%AF%E5%A4%AB%E6%A8%A1%E5%9E%8B/_blank)和显示随机函数集。

隐[马尔可夫](https://baike.baidu.com/item/%E9%A9%AC%E5%B0%94%E5%8F%AF%E5%A4%AB" \t "https://baike.baidu.com/item/%E9%9A%90%E9%A9%AC%E5%B0%94%E5%8F%AF%E5%A4%AB%E6%A8%A1%E5%9E%8B/_blank)模型（HMM）可以用五个元素来描述，包括2个状态集合和3个概率矩阵：

1. 隐含状态 S

这些状态之间满足[马尔可夫](https://baike.baidu.com/item/%E9%A9%AC%E5%B0%94%E5%8F%AF%E5%A4%AB" \t "https://baike.baidu.com/item/%E9%9A%90%E9%A9%AC%E5%B0%94%E5%8F%AF%E5%A4%AB%E6%A8%A1%E5%9E%8B/_blank)性质，是[马尔可夫模型](https://baike.baidu.com/item/%E9%A9%AC%E5%B0%94%E5%8F%AF%E5%A4%AB%E6%A8%A1%E5%9E%8B" \t "https://baike.baidu.com/item/%E9%9A%90%E9%A9%AC%E5%B0%94%E5%8F%AF%E5%A4%AB%E6%A8%A1%E5%9E%8B/_blank)中实际所隐含的状态。这些状态通常无法通过直接观测而得到。（例如S1、S2、S3等等)

2. 可观测状态 O

在模型中与隐含状态相关联，可通过直接观测而得到。(例如O1、O2、O3等等，可观测状态的数目不一定要和隐含状态的数目一致。）

3. 初始状态[概率](https://baike.baidu.com/item/%E6%A6%82%E7%8E%87" \t "https://baike.baidu.com/item/%E9%9A%90%E9%A9%AC%E5%B0%94%E5%8F%AF%E5%A4%AB%E6%A8%A1%E5%9E%8B/_blank)矩阵 π

表示隐含状态在初始时刻t=1的[概率](https://baike.baidu.com/item/%E6%A6%82%E7%8E%87" \t "https://baike.baidu.com/item/%E9%9A%90%E9%A9%AC%E5%B0%94%E5%8F%AF%E5%A4%AB%E6%A8%A1%E5%9E%8B/_blank)矩阵，(例如t=1时，P(S1)=p1、P(S2)=P2、P(S3)=p3，则初始状态概率矩阵 π=[ p1 p2 p3 ].

4. 隐含状态转移概率矩阵 A。

描述了HMM模型中各个状态之间的转移概率。

其中Aij = P( Sj | Si ),1≤i,,j≤N.

表示在 t 时刻、状态为 Si 的条件下，在 t+1 时刻状态是 Sj 的[概率](https://baike.baidu.com/item/%E6%A6%82%E7%8E%87" \t "https://baike.baidu.com/item/%E9%9A%90%E9%A9%AC%E5%B0%94%E5%8F%AF%E5%A4%AB%E6%A8%A1%E5%9E%8B/_blank)。

5. 观测状态转移概率矩阵 B （英文名为Confusion Matrix，直译为[混淆矩阵](https://baike.baidu.com/item/%E6%B7%B7%E6%B7%86%E7%9F%A9%E9%98%B5" \t "https://baike.baidu.com/item/%E9%9A%90%E9%A9%AC%E5%B0%94%E5%8F%AF%E5%A4%AB%E6%A8%A1%E5%9E%8B/_blank)不太易于从字面理解）。

令N代表隐含状态数目，M代表可观测状态数目，则：

Bij = P( Oi | Sj ), 1≤i≤M,1≤j≤N.

表示在 t 时刻、隐含状态是 Sj 条件下，观察状态为 Oi 的概率。

总结：一般的，可以用λ=(A,B,π)[三元组](https://baike.baidu.com/item/%E4%B8%89%E5%85%83%E7%BB%84" \t "https://baike.baidu.com/item/%E9%9A%90%E9%A9%AC%E5%B0%94%E5%8F%AF%E5%A4%AB%E6%A8%A1%E5%9E%8B/_blank)来简洁的表示一个隐[马尔可夫模型](https://baike.baidu.com/item/%E9%A9%AC%E5%B0%94%E5%8F%AF%E5%A4%AB%E6%A8%A1%E5%9E%8B" \t "https://baike.baidu.com/item/%E9%9A%90%E9%A9%AC%E5%B0%94%E5%8F%AF%E5%A4%AB%E6%A8%A1%E5%9E%8B/_blank)。隐[马尔可夫模型](https://baike.baidu.com/item/%E9%A9%AC%E5%B0%94%E5%8F%AF%E5%A4%AB%E6%A8%A1%E5%9E%8B" \t "https://baike.baidu.com/item/%E9%9A%90%E9%A9%AC%E5%B0%94%E5%8F%AF%E5%A4%AB%E6%A8%A1%E5%9E%8B/_blank)实际上是标准马尔可夫模型的扩展，添加了可观测状态集合和这些状态与隐含状态之间的概率关系。

4、递归定量分析：Rome等人采用递归定量分析（RQA）去给声学事件的瞬时出现建立模型，这种技术被用来学习一系列参数，这些参数来自捕获时间序列的周期性的所谓的递归图参数和被用来学习在混沌理论背景下的动态系统。在ASC的背景下，递归定量分析的参数包括：复发测量在音频场景特征的自相似程度；与声音周期相关的决定论；和捕捉包含固定段的声音的层流罩。统计学习函数的输出是一组给训练集中每个声学场景进行建模的参数，然后，将这些参数集合送入支持向量机，以定义用于分类未标记信号的类之间的决策边界。

5、空间因子：由埃利萨尔德等人提出的系统建立在i矢量的计算基础上。这是一种最初在语音处理方面开发的用于定位扬声器验证问题的技术，它是基于使用高斯混合模型建模特征序列。在ASC的背景下，i向量是特地从梅尔倒谱系数衍生出作为高斯混合模型的参数。这导致一个低维表示总结声场景的特性和输入生成的概率线性判别分析。

还有几种决策衡量准则

1、一对一和一对所有：这些决策标准与多类支持向量机的输出相关联，用于将特征向量的位置映射到类中。

2、多数表决：每当一个全局分类必须从关于单个音频帧的判决中估计时，就使用这个标准。通常，音频场景是根据最常见的分配到每一帧的类别进行分类的。或者，也可以采用加权多数表决来改变不同帧的重要性。例如Patil和Elahili，分配较大的权重的音频帧包含更多的能量

3、最近邻：根据这个标准，一个特征向量被分配给与训练集最近的向量相关的类（根据一个度量，一般为欧几里得距离）。最近邻的一个推广是k-近邻准则，即考虑k个最接近的向量，并根据最常见的分类确定一个类别

4、最大似然法：该标准与生成模型相关联，从而将特征向量分配给其模型，该模型最有可能根据似然概率生成所观察数据。

5、最大后验概率（MAP）：最大似然分类的另一种方法是map准则，它包括关于给定类的边缘可能性的信息。举例来说，支持在移动设备中建立全球定位系统表明，在当前的地理区域，某些环境比其他环境更容易遇到。这个信息可以通过最大后验概率准则被包括到ASC算法中。

6、元算法：在有监督分类的情况下，元算法是通过并行地运行分类器的多个实例而减少分类错误设计的机器学习技术。其中使用不同的参数或不同的训练数据。然后将每个分类器的结果合并为一个全局决策。