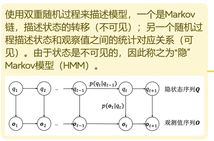
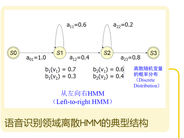
语音识别

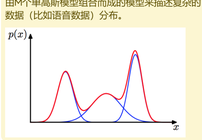
语音识别技术的展望

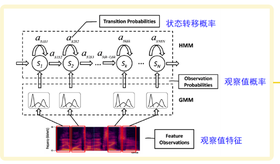
- 端到端语音识别
  - 基于CTC的语音识别算法
    - 基于神经网络的时序信号分类
  - 基于Attention的语音识别算法
    - 编码器-解码器模型（Encoder-Decoder Model）
      - 直接将声学特征映射到更大的单元序列中，而不是直接或近似地序列：都在序列
    - Attention模型的优点
      - 实现了端到端训练，能够将整个系统训练到最优
      - 与CTC相比，Attention不需要每帧独立的假设，
      - 在噪声环境下很难识别到好的效果，通常要比CTC的方法差
    - Attention模型缺点
      - 在噪声环境下很难识别到好的效果，通常要比CTC的方法差
      - 目前的解决方案：CTC和attention联合训练，采用CTC损失函数对于下层Encoder做强制对齐。
- 主要挑战
  - 复杂声学信息下的语音识别
    - 现实生活中，声学信息千变万化，训练集不能完全覆盖所有声学信息

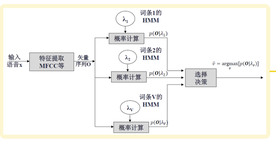
- 马尔可夫模型
  - 马尔可夫模型
    - 定义
      - 从一个状态到另一个状态的转换的随机过程
      - 系统在时间序列的状态与其在时间t0-t0状态相关，满足以下关系：
        - $P(q_t = s_j|q_{t-1} = s_i, q_{t-2} = s_k, ...) = P(q_t = s_j|q_{t-1} = s_i) \quad 1 \le i, j \le N$
        - 即称该随机过程为离散时间，离散状态的一阶马尔可夫过程，或简称为一阶马尔可夫链。
    - 马尔可夫链
      - 离散状态的马尔可夫过程
      - 使用双重随机过程来描述模型，一个是Markov链，描述状态的转移（不可见）；另一个随机过程描述状态和观察值之间的统计相关关系（可见），由于状态是不可见的，因此称之为"隐"Markov模型（HMM）。
      - HMM模型定义
        - 

      - HMM模型描述和参数
        - 状态集合
          - 状态 = {
            隐藏状态参数, // GMM或DNN的参数
            状态转移概率, // 到隐状态的转移概率
            持续时间统计 // 在该状态停留的时间分布
            }
        - 可观测值集合
          - 观测值 = {
            MFCC_1, MFCC_2, ..., MFCC_N, // 通常15个MFCC系数
            Energy, // 能量
            Delta_1, Delta_2, ..., // 一阶差分
            DeltaDelta_1, DeltaDelta_2, ... // 二阶差分
            }
        - 初始状态概率矩阵
        - 状态转移矩阵
        - 观测值概率矩阵
          - 表示在 t时刻，隐含状态是s_j的条件下，观察值为o_k的概率。
    - 马尔可夫链的问题与隐马尔可夫模型（HMM）
      - 问题1：模型概率计算问题
        - 定义
          - 到时刻为止部分观测序列为O1,O2,...Ot，且状态为Si为约束条件，在后面的剩余序列O(t+1),O(t+2)为的概率。
        - 前向算法(forward algorithm)
          - （1）初始化：对 $1 \le i \le N$, 有
            - $\alpha_1(i) = \pi_i b_i(o_1)$
          - （2）递推：对$2 \le t \le T, 1 \le j \le N$, 有
            - $\alpha_t(j) = \left[\sum_{i=1}^{N} \alpha_{t-1}(i) a_{ij}\right] b_j(o_t)$
          - （3）终止：$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$
        - 后向算法(backward algorithm)
          - 当前时刻t
            - $\underbrace{O_1, O_2, ..., O_t}_{已知} \underbrace{O_{t+1}, O_{t+2}, ..., O_T}_{?}$
            - 总是关注的是从t+1时刻到结束的这段序列，考虑了所有可能的状态转移路径。
          - （1）初始化：对 $1 \le i \le N$, 有
            - $\beta_T(i) = 1$
          - （2）递推：对$t = T-1, T-2, ..., 1, 1 \le i \le N$, 有
            - $\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)$
          - （3）终止：$P(O|\lambda) = \sum_{i=1}^{N} \pi_i b_i(o_1) \beta_1(i)$
      - 问题2：模型解码问题
        - 模型解码问题的描述
          - 问题2：模型解码问题
            - 给定一个最佳状态序列$Q^* = q_1^*, q_2^*, ..., q_T^*$，使$P(O, Q^*|\lambda)$为最大。
            - $Q^* = \underset{Q}{\operatorname{argmax}} P(O, Q|\lambda)$
            - 定义$\delta_t(i) = \max_{q_1, q_2, ..., q_{t-1}} P(o_1 o_2 \cdots o_t, q_1 \cdots q_{t-1}, q_t = s_i|\lambda)$为时刻t对应第一条路径$q_1, q_2, ..., q_{t-1}, q_t = s_i$产生$o_1 o_2 \cdots o_t$的最大概率。
        - Viterbi算法
          - 求解最佳状态序列Q?
            - （1）初始化：对 $1 \le i \le N$, 有
              - $\delta_1(i) = \pi_i b_i(o_1)$
              - $\psi_1(i) = 0$
            - （2）递推：对 $2 \le t \le T, 1 \le j \le N$, 有
              - $\delta_t(j) = \max_{1 \le i \le N} [\delta_{t-1}(i) a_{ij}] b_j(o_t)$
              - $\psi_t(j) = \underset{1 \le i \le N}{\operatorname{argmax}} [\delta_{t-1}(i) a_{ij}]$
            - 未观测到的状态序列Q? 的定义为
            - （3）终止：$P^* = \max_{1 \le i \le N} [\delta_T(i)]$
              - $q_T^* = \underset{1 \le i \le N}{\operatorname{argmax}} [\delta_T(i)]$
            - （4）路径回溯，最优状态序列：
              - $q_t^* = \psi_{t+1}(q_{t+1}^*), t = T-1, T-2, ..., 1$
      - 问题3：模型参数估计问题
        - Baum-Welch算法
          - 模型参数估计的思路
            - 用来解决在观测序列其隐状态序列未知的时候模型的算法，该算法是期望最大化（EM）算法的一个特例。
            - 设辅助函数$L(\hat{\lambda}, \lambda) = \sum_Q P(Q|O, \lambda) \log P(O, Q|\hat{\lambda})$.
              - 在步骤1 固定参数集λ去找出$\hat{\lambda}$。
              - 在步骤2最大化 更新模型参数$\hat{\lambda} = \lambda$ 直至$\hat{\lambda}$。
            - $\hat{\lambda} = \arg\max_{\hat{\lambda}} \sum_Q P(Q|O, \lambda) \log P(O, Q|\hat{\lambda})$

- 声学模型
  - 基于GMM-HMM的语音识别算法
    - 语音识别或离散HMM模型结构
      - 

    - 高斯混合模型（Gaussian Mixture Models）
      - 由N个单高斯模型组合形成的模型来描述复杂的数据（比如语音数据）分布。
      - 

      - 对于M个高斯模型的混合模型，o_t是观测数据等下第n个模型的概率值（混合系数），则模型总体概率分布为：
        - $P(o|\lambda) = \sum_{n=1}^{M} c_n N(o|\mu_n, \Sigma_n)$
        - $N(o|\mu_n, \Sigma_n) = \frac{1}{\sqrt{(2\pi)^D |\Sigma_n|}} \exp\left\{-\frac{1}{2}(o - \mu_n)^T \Sigma_n^{-1} (o - \mu_n)\right\}$
        - 其中，$\sum_{n=1}^{M} c_n = 1$，$N(o|\mu_n, \Sigma_n)$是第n个单高斯的观测分布概率函数。
      - 用连续的概率密度函数来描述连续的观察值向量（ot）的概率分布
      - 混合高斯模型（GMM）来表示ot的概率分布
        - $b_j(o_t) = \sum_{m=1}^{M} \frac{c_{jm}}{\sqrt{(2\pi)^D |\Sigma_{jm}|}} \exp\left[-\frac{1}{2}(o_t - \mu_{j,m})^T \Sigma_{jm}^{-1} (o_t - \mu_{j,m})\right]$
    - 基于GMM-HMM的语音识别框架
      - 

    - 基于HMM-GMM的语音识别算法
      - 基于HMM的孤立词识别技术
      - 
        输入语音
        MFCC
        输出概率
        HMM
        输入语音
        MFCC
        输出概率
        HMM
        输出概率
        HMM
        输出概率
        HMM
        输出概率
        输入语音
        MFCC
        输出概率
        HMM
        输出概率
        HMM
        输出概率
        HMM