POS调研

1. 传统方法MFCC-HMM[[1]](#footnote-1)
2. 主要步骤

分帧（类似于N-gram）->特征转化（MFCC[[2]](#footnote-2)）->状态（HMM[[3]](#footnote-3)）->音素。传统方法中无论是提特征还是获取状态，都有不同的方法，但是使用最广泛的还是 MFCC和HMM模型。其中关键点也是两种方法的实现。

1. 开源实现（python）

特征提取[[4]](#footnote-4)（亲测可用）

HMM[[5]](#footnote-5)（viterbi方法）

1. DL方法

还没调研，感觉训练数据是个梗。

1. 企业开发SDK

其中很多是在线的工具，所以可能不太使用。科大讯飞有几个是离线的，我下载了一个windows版本，是基于C语言。如果是使用C#，可能还要在找一找。

1. 科大讯飞[[6]](#footnote-6)
2. 腾讯[[7]](#footnote-7)
3. 百度[[8]](#footnote-8)

1. https://www.zhihu.com/question/20398418/answer/18080841 [↑](#footnote-ref-1)
2. http://www2.cmpe.boun.edu.tr/courses/cmpe362/spring2014/files/projects/MFCC%20Feature%20Extraction.pdf [↑](#footnote-ref-2)
3. http://digital.cs.usu.edu/~cyan/CS7960/hmm-tutorial.pdf [↑](#footnote-ref-3)
4. https://github.com/jameslyons/python\_speech\_features [↑](#footnote-ref-4)
5. http://scikit-learn.sourceforge.net/stable/modules/hmm.html [↑](#footnote-ref-5)
6. http://www.xfyun.cn/sdk/dispatcher [↑](#footnote-ref-6)
7. http://gcloud.qq.com/product/6 [↑](#footnote-ref-7)
8. https://cloud.baidu.com/product/speech.html?track=cp:nsem|pf:pc|pp:speech|pu:speech|ci:|kw:63094 [↑](#footnote-ref-8)