

实验二：基于 HMM 的语音识别与强制对齐实验

文档简介

《语音信号数字处理》课程的第二次实验课采用课下自主实验为主与课上讲解为辅的形式进行。

课下自主实验需要在 12 月 8 日之前完成，需完成文献阅读，并撰写实验报告，完成以下所述实验任务并回答问题，通过网络学堂提交（截止时间 12 月 8 日 23:59）。该实验旨在通过让同学们结合课程讲义第四部分“隐马尔可夫模型”、第五部分“语音识别”及第七部分“基于隐马尔可夫模型的语音合成”的内容，对 HMM 相关算法有更深入的理解，同时对基于 HMM 的语音识别和强制对齐有入门级别的直观认知。

课上讲解将于第九周周四（11 月 7 日）上午正常上课时间在国际一期 C504进行，课上将进行实验流程和细节的讲解。并请所有同学备好各自的笔记本电脑（及耳机）。

本文档为课下自主实验部分的大纲文档，核心内容包括三项基本任务：

- (1) 文献阅读并回答问题；
- (2) Viterbi 算法的实现，包括前向(概率)算法、后向(概率)算法、Viterbi 算法；
- (3) 基于 HMM 的强制对齐实验。

中英文术语对照表

Markov Process：马尔可夫过程

First Order Markov Process：一阶马尔可夫过程

(First Order) Markov System：(一阶)马尔可夫系统

Hidden Markov Model (HMM)：隐马尔可夫模型

Gaussian Mixture Model (GMM)：高斯混合模型

Automatic Speech Recognition (ASR)：(自动)语音识别

Text-To-Speech (TTS)：(文语转换)语音合成

基础介绍与资源

MFA：常用的语音文本强制对齐工具包，将在实验任务二中使用。

PyTorch：最常用的深度学习框架（之一），将在实验任务三中使用。

实验任务

基本任务：查找阅读文献资料并回答问题（50）

查找并阅读隐马尔可夫模型（Hidden Markov Model, HMM）的相关文献资料并回答问题。

- 1) 请查找并阅读至少 3 篇有关 HMM 模型的参考文献资料，其中：至少 1 篇为经过同行评审的正式发表的英文论文（会议或期刊均可）；不得多于 1 篇为知乎、博客、百科等非正式发表的科普性阅读资料或讲义。（10 分）
- 2) 以标准参考文献的格式列出所阅读的参考文献列表，并在回答以下问题时给出明确的引用。（5 分）
- 3) 什么是马尔可夫过程？一阶马尔可夫过程的特点是什么？（可结合图）对一阶马尔可夫过程及一阶马尔可夫系统进行明确的形式化定义。（3 分）
- 4) 什么是隐马尔可夫模型？（可结合图）对隐马尔可夫模型进行明确的形式化定义。（3 分）
- 5) 给出隐马尔可夫模型的 3 个典型问题的形式化描述。（4 分）
- 6) 给定隐马尔可夫模型 λ ，如何计算该模型生成某个观察序列 O 的概率？请给出数学公式及对应的推导。（5 分）
- 7) 给定隐马尔可夫模型 λ 、以及对应的观察序列 O ，如何得到该模型产生该观察序列的最优隐含状态的序列 Q ？请给出数学公式及对应的推导。（5 分）
- 8) 给你一堆观察数据 O_1, O_2, \dots, O_N ，并假设这些观察数据符合隐马尔可夫过程的假设，如何估计出一个最优的隐马尔可夫模型 λ ，该模型产生这些观察数据的概率最优。（5 分）
- 9) 什么是高斯混合模型？高斯混合模型如何与隐马尔可夫模型结合？（5 分）
- 10) 请结合相关文献，介绍一个隐马尔可夫模型在语音及自然语言处理领域的最新应用实例，包括但不限于语音识别、语音合成、文本分词等。注意需要解释清楚模型是如何应用到实际任务中的，以及是否存在方法上的改进。（5 分）

实验任务一：Viterbi 解码算法实现（20 分）

参考[实验任务一的实验要求（点击访问超链接）](#)和第四部分“隐马尔可夫模型”PPT 实现以下算法：

- 1) 前向算法（7 分）；
- 2) 后向算法（7 分）；
- 3) Viterbi 解码算法（6 分）。

注：HMM 发射（输出）概率矩阵，转移概率矩阵以及初始概率矩阵和观察序列已给定，算法输出的正确结果请参考所给的[参考实验结果 results](#) ([点击访问超链接](#))。

实验任务二：基于 HMM 的强制对齐实验（30 分）

参考[实验任务二的实验要求](#) ([点击访问超链接](#)) 运行每一步的脚本并在所提交的报告中回答所有相关问题。

- 1) 只要成功运行 align 命令，即获得基础分值 12 分。
- 2) 正确回答相关问题，将获得 18 分。
- 3) 完成两个附加题，将有加分。

注：连同加分，实验任务二得分上限为 30 分。

作业提交

完成所有任务后请将[实验任务一的 viterbi.py](#) 和[实验报告](#)打包压缩提交至网络学堂。

附录

HMM 经典文献

- L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," Proceedings of the IEEE, vol. 77, no. 2, pp. 257–286, 1989.

语音领域相关会议、期刊名称

- IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)
- IEEE International Conference on Acoustics, Speech & Signal Processing (ICASSP)
- Annual Conference of International Speech Communication Association (INTERSPEECH)

关于作业参考文献的规范要求(参考 ICASSP 论文模板要求)

- REFERENCES: List and number all bibliographical references at the end of the paper. The references can be numbered in alphabetic order or in order of appearance in the document. When referring to them in the text, type the corresponding reference number in square brackets as shown at the end of this sentence [2].
- The end of the document should include a list of references containing information similar to the following examples (Pay attention to all format details, including but not limited to

authors' name list, lower/upper-cases, punctuations, etc.):

[1] D. E. Ingalls, "Image Processing for Experts," IEEE Transactions on Acoustics, Speech, and Signal Processing (TASSP), vol. TASSP-36, pp. 1932-1948, 1988.

[2] L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," Proceedings of the IEEE, vol. 77, no. 2, pp. 257-286, 1989.

[3] Y. Lei, S. Yang, and L. Xie, "Fine-Grained Emotion Strength Transfer, Control and Prediction for Emotional Speech Synthesis," IEEE Spoken Language Technology Workshop (SLT), pp. 423-430, 2021.