



中國石油大學 (华东)  
CHINA UNIVERSITY OF PETROLEUM

## 课程总结报告

学生姓名： 李宇

学 号： 1907040113

专业班级： 本研人工智能

课程认识 30%	问题思考 30%	结构规范 20%	IT工具 20%	Latex附加 10%	总分	阅卷教师

2019年 12月 20日

# 1 引言

一个学期的计算科学导论课程已经结束了。在老师的指导下，我们完成了这门课程的学习。由原来的对计算机原理一无所知到现在的了解基本框架和学习方法，得到了许许多多的新知识。我们了解到学习计算机知识最开始应当从用哲学的思想方式去学习。我们了解到了计算机的起源，计算机的发展历史。我意识到计算机的发明是建立在无数古人付出毕生心血乃至生命大代价上的。了解到一次次的数学危机是如何使人们逐渐接近数学的真理，为计算机的发明打下基础的。我获得了许多的经验，如：重在学习过程的完整性，注重自我知识体系的构建与创新，注重学生个体能力的提升。这些能力不仅包括本课程技术知识应用的能力，还包括学习能力，表达能力，沟通能力等。

在高中时期，我从未接触到过计算机技术的学习，对计算机的组成、原理一无所知。那时，我虽然对计算机毫无理解，但却充满好奇，充满探索精神。从电脑进入平常百姓家时，我就下定决心无论前路有多大的困难，我都要掌握计算机使用方法。最开始学习时，我发现计算机的原理比想象中的复杂，计算机无法识别人类的思维，这也是人们要认清计算机工作原理的原因。计算科学导论作为计算机学院学生必备的基础知识。是我们在将来的路上的基石。它包括了在计算科学的基本概念和基本认识，计算科学的含义、内容和方法。可以不断学习到关于计算机的大体情况，不断获取，不断补充以前所不知道或不了解的行业动态。虽然还有很多地方掌握的不是很好，但以后我会通过不断地练习去慢慢掌握。

## 2 对计算科学导论这门课的认识、体会

这门课程最先说明了目前计算机的发展水平，发展方向。让刚刚接触到计算机的学生知道应该从哪些方面开始学习，最终让自己成为一个优秀的计算机方面的专业技术性人才。让我们明白计算机的基本构造和各个硬件之间的配合关系，使我们在之后的学习中如学习各类计算机语言时可以了解其中的细节关系从而更好的完成学业。这门课程也让我意识到了计算机的创造过程有多么的艰难，图灵，冯诺伊曼等人有着多么巨大的贡献和他们的伟大之处。这本书并没有告诉我们具体的应当如何学习计算机知识但却告诉我们应当在今后如何去学习，在全书的最初，他就告诉我们，一流的专业人才应具有高尚的品德和良好的人文素养，应具有坚实的专业基础和深厚的专业功底富有创新意识具有科学思想的方法三个必要条件。从第2章起，该书正式介绍计算机的组成和原理，我意识到计算机的原理源于数学，因此数学这门学科对于我们今后的学习有着至关重要的作用，我们明白了二进制的原理和计算方法，虽然目前我还不知道了解这些知识对我今后有什么帮助，可二进制作为计算机的最基本的知识在将来一定会发挥至关重要的作用。第3章则是从程序指令，图像识别模式识别人工智能等方面深度介绍计算机。在全部的课程学习中，目前超级计算机的发展情况是我最感兴趣的，并且这也说明了目前计算机在计算速度领域的发展水平。从课上内容得知，目前世界上最快的超级计算机是美国的顶点，计算速度达到了浮点运算每秒14.35亿亿次。产自中国的神威，太湖之光位列第二计算速度略慢。这两台超级计算机计算速度相差无几，并且目前世界500强的计算机中国占了一大半。在下一年的超级计算机第1名的争夺中，计算机将向每秒e次级计算冲击。中美双方都在为此进行充分的准备，在超级计算机方面两个国家的争夺也是两个超级大国之间斗争的一个体现。

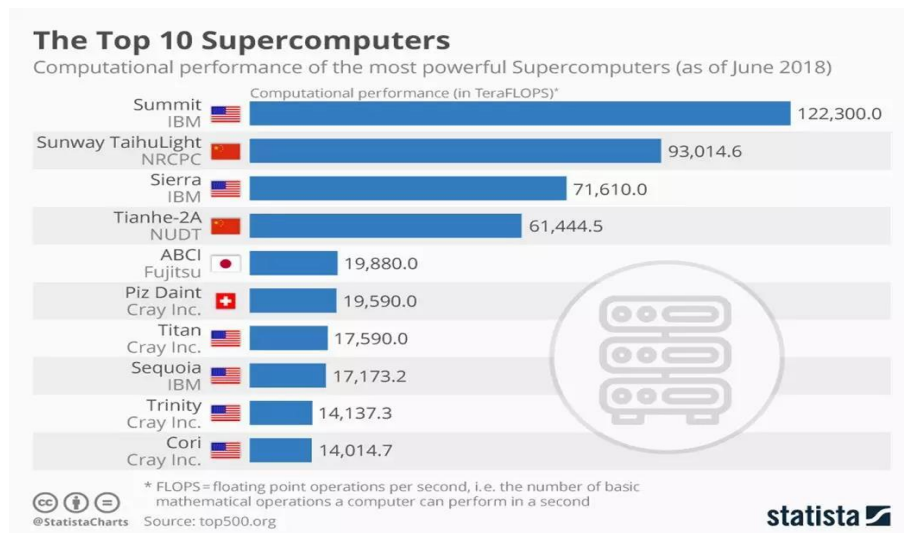


图 1: 超级计算机排名

在此，我重点介绍国产超级计算机：太湖之光。根据 6 月 20 日发布的世界最新高性能计算机 TOP500 排名数据显示，“神威·太湖之光”的峰值运算速度达每秒 12.54 亿亿次、持续运算速度每秒 9.3 亿亿次、性能功耗比 60.51 亿次/瓦，是世界上首台峰值运算速度超过十亿亿次的超级计算机，也是我国第一台全部采用国产处理器构建的世界第一的超级计算机。这些问题还需要 5 到 10 年的时间去解决，达到平衡状态。神威太湖之光的机身就安在一间约 1000 平方米的房间内，它有 40 个计算机柜和 8 个网络机柜组成，每个计算机柜比家用的双门冰箱略大，打开柜门 4 块由 32 块运算插件组成的超节点分布，其中每个插件由 4 个运算节点板组成，一个运算节点板又含两块申威，26010 高性能处理器。据统计一台机柜装有 1024 块处理器，整台神威太湖之光共有 40960 块处理器。申威 SW26010 是中国首个采用国产自研架构且性能达到世界一流的计算机芯片。SW26010 采用 260 核心众核架构乱序执行架构，频率 1.45GHz，整个处理器包括 4 个 MPE（Management Processing Element）管理单元、4 个 CPE（Computing Processing Element）计算单元及 4 个 MC 内存控制器单元组成，其中 CPE 单元又由 8x8 阵列的 64 核心组成，所以总计是 260 个核心（4x64=260）。与其他国产处理器相比，申威系在性能上完全走在了前列，在世界范围内都是靠前的，SW26010 的理论浮点性能高达 3TFLOPS，SW26010 可以直接访问主存，因此在实际使用效率上不见得就会比 intel phi 低多少，并且某些应用场合甚研制出高性能的芯片，打破西方国家的技术封锁。由此可见，处理器对于超级计算机计算的速度起着决定性的作用。我们需要不断提高我国的芯片制造水平，进而逐渐使我国从芯片制造大国成为芯片创造强国，便是我们这一代人的时代任务，也就是我们的努力方向。

### 3 进一步的思考

我的分组演讲课题是语音识别（ASR），语音识别技术定义是让机器通过识别和理解过程把语音信号转变为相应的文本或命令的高新技术。语音识别技术主要包括特征提取技术、模式匹配准则及模型训练技术三个方面。语音识别是人机交互的基础，主要解决让机器听清楚人说什么的难题。人工智能目前落地最成功的就是语音识别技术。语音之所以可以转化成文字，用

简单的话叙述如下：正如同英文单词是由英文字母组成的，声音也是由音素组成的，机器先收到并录入人们的声音，并将声音转化成一段长长的声波。随后机器将长长的声波分成极其小的一段，每一段叫做一帧。声学模型可以将已知的所有因素与录入的小段声波进行对比，并找出最有可能匹配的音素。随后将音速串在解码器中，与相应的文字进行对比。对比所用的算法和模型中行中，马尔可夫模型，是如今最常用最实用的。其他与语音识别相关的应用，如语音控制系统，智能对话查询系统等使用的方法，都与此相关。从上世纪50年代起，科研人员就对语音的音素，开始了相关的研究，上世纪70年代，马尔可夫模型在语音识别中的运用，使语音识别真正有了走向百姓人家的可能性。随着大数据，深度学习等技术的发展，让语音识别技术，在2010年开始快速发展，现如今主流的语音识别系统的准确率均超过了百分之90，业界领导者科大讯飞的语音识别率更是达到了百分之98，不仅是科大讯飞，包括百度，搜狗，谷歌等厂家的语音识别系统，均已达到了可以商用的级别。语音识别准确率之所以快速提升，马尔可夫模型发挥了至关重要的作用。

### 3.1 隐马尔科夫模型

HMM 是一种用参数表示的用于描述随机过程统计特性的概率模型，它是由马尔可夫链演变来的。所以它基于参数模型的统计识别方法。它是一个双重随机过程——具有一定状态数的隐式马尔可夫链和显示随机函数集。每个函数都与链中一个状态相关联。隐式过程通过显示过程所产生的观察符号序列来表示。“它把一个总随机过程看成一系列状态的不断转移。时刻  $t$  的状态用  $q_t$  表示，它可以是  $N$  种状态集合  $S = [s_1, s_2, \dots, s_N]$  中的任意一个。马尔可夫模型的特性主要用“转移概率”来表示。后一状态出现的概率决定于其前出现过的状态次序。即：状态  $q_t$  出现的概率为  $P_r[q_t | q_{t-1}, q_{t-2}, \dots, q_1]$ 。如果此概率只决定于前一个状态，即  $P_r[q_t | q_{t-1}]$ ，则称为一阶马尔可夫过程。它是研究中引用得最多的形式，即：  $P_r[q_t | q_{t-1}, q_{t-2}, \dots, q_1] = P_r[q_t | q_{t-1}]$ 。隐马尔可夫模型 (Hidden Markov Model, HMM) 则认为模型的状态是不可观测的 (这便是“隐”得名的由来)。能观测到的只是它表现出的一些观测量 (observations)。” [3]

HMM模型也有着很大的改建空间，如《隐马尔可夫模型在语音识别中的应用》中所言“HMM 技术之所以在语音识别中应用较为成功，主要是它具有较强的对时间序列结构的建模能力，尽管如此，HMM 仍然是有缺点和局限性的：(i) 对低层次的声学音素建模能力差，使声学上相似的词易混淆；(ii) 对高层次语音理解或语义建模能力差，使其仅能接受有限状态或概率文法等简单场合应用；(iii) 一阶HMM 假设很难直接用模型描述协同发音 (coarticulation)，因为HMM 假设输出是相互独立的，且依赖于当前状态。”

### 3.2 语音合成

语音合成 (text-to-speech) 与语音识别原理相同，使用到的技术、方法也大致相同。这种技术目前可以做到文本转换语音、音乐生成、语音生成、语音支持设备、导航系统等功能，就目前来看将来这项功能在未来最有可能用于为视障人士提供无障碍服务。语音合成一般可分为两个步骤：文本处理和语音合成 (前端和后端)。

首先第一个步骤文本处理，与语音识别中将音素逐步转化成文字的过程正好相反，语音合成要将文本转换成音素序列。但这个过程与音素转文字又有所不同，它有一些新的问题需

要解决，比如拼写相同但读音不同的词的区分、缩写的处理、停顿位置的确定、情感预测等等。“This paper has presented WaveNet, a deep generative model of audio data that operates directly at the waveform level. WaveNets are autoregressive and combine causal filters with dilated convolutions to allow their receptive fields to grow exponentially with depth, which is important to model the long-range temporal dependencies in audio signals. We have shown how WaveNets can be conditioned on other inputs in a global (e.g. speaker identity) or local way (e.g. linguistic features). When applied to TTS, WaveNets produced samples that outperform the current best TTS systems in subjective naturalness. Finally, WaveNets showed very promising results when applied to music audio modeling and speech recognition.” [2] 此论文说明WaveNet仍然是语音合成后端的最优方法。随着语音合成的不断发展，人们开始意识到当时语音识别的操作过于复杂，文字的生成速度过慢。自然，人们就想到，能不能将语音合成的两个步骤简化为一个步骤，这样一来，文字生成速度必然能大幅增加，之后科研工作者就开始了End-to-End模型的研究。《一个无条件的端到端神经音频生成模型》[1]中说明：“我们提出了一个新的模型，可以解决在原始声学领域的无条件的音频生成，这通常已经完成，直到最近手工制作的功能。我们能够证明，时间尺度的层次结构和频繁的更新将有助于克服对非常高分辨率的时间数据建模的问题。这使得我们可以直接从音频样本中学习数据流形。结果表明，该模型能很好地泛化和生成三个性质不同的数据集上的样本。我们还表明，该模型生成的样本是人类的首选。”

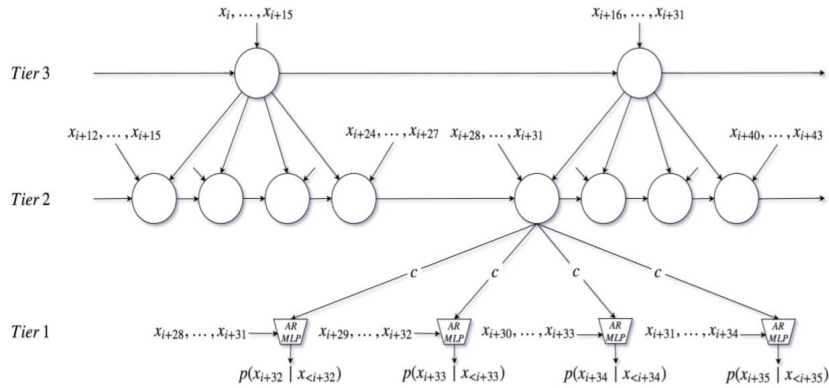


Figure 1: Snapshot of the unrolled model at timestep  $i$  with  $K = 3$  tiers. As a simplification only one RNN and up-sampling ratio  $r = 4$  is used for all tiers.

图 2: ent to end 初始模型

### 3.3 语音识别的商业盈利模式

现如今，使用场景越来越广阔。2014 年，全球智能语音识别市场规模达 2750 亿美元，2016 年达 4210 亿美元，同比增长百分之 53.1。有经济学家指出，这一市场规模到 2025 年将超过 10000 亿美元。即使如此，市场上的语音识别企业，大多仍然处于亏损状态，曾有数据显示，2018 年全年，近百分之 90 的人工智能公司处于亏损状态，而百分之 10 赚钱的企业基本是技术提供商。在

极少数可以盈利的企业中，讯飞便是其中之一，但它并不典型，它既是产品提供商也是技术提供商。C段业务住在成为讯飞输入的主力，打开讯飞的首页，我们可以看到它不同于其他语音公司的软件，它有着讯飞翻译机，学习本录音笔等多种智能硬件，软件中以讯飞输入法为主等多种软件共同实现营收，以讯飞听见为例，讯飞听见通过翻译时长需要购买的方式来创造收入，该公司之所以敢于采取这种营收方式，源于他对自己实力的自信。讯飞听见可以快速转写（约四十字每分钟），以及中英文实时翻译的功能都是业界领先的。由此见得，创造收入的前提是自身技术的优越。应用声纹识别技术来确认身份,高精度的身份确认,可以使军事系统的安全性大大增强。一些计算机产品,为了安全起见,也使用了声纹识别技术,例如在普通的移动存储设备上增加声纹认证功能,对电脑系统进行语音加密保护,用以保护个人隐私和军事机密。在保密方面，语音识别的优越性不言而喻。安保方面在未来也注定成为语音识别企业的盈利方面之一。随着智能化时代的到来，智能家居终将普及，家居智能的实现，离不开语音识别。想要增强智能家居消费者的体验感，舒适度。语音识别的精度，速度将会有着更高的需求。不仅是智能家居，未来大量的智能产品都将以语音识别为基础，这无疑会增加语音识别企业接受的投资，企业以此来提升自己的产品实力，也是公司最终实现盈利的方式之一。

声，聆音千里之臻，洞悉万象之源，重塑数据价值，强化机器学习，真实环境下语音识别率将全面提升，为更自由的人机语音交互提供保障，感你所言，知你所想，在纷繁的环境下，敏锐地感知你的声音，全方面无延迟地取景，每一声呼唤都闻声即映。领先的声学技术可以感知你的方向，倾听你的表达，消除空间混响，抵消声源回声，用灵敏的“耳朵”让人机交互更加随心所欲，声纹让交流专属独享，识别语调情绪，让沟通快速高效。聚合云端应用，让智能自由分发，智能汽车，玩具，机器人，智能家居，医疗，安防……语音识别将引领智能化时代，构建沟通的桥梁。繁华而有序的智能生活将在未来实现已成为了必然，在这种大趋势，语音识别的盈利问题自然会迎刃而解。

## 4 总结

一学期的计算科学导论结束了，书本上的知识让我们明白了计算机的发展历史，硬件做成，运行的基本原理。为我们今后的发展奠定了基础。两人一组的PPT介绍是课程相当重要的组成部分，我和搭档分工明确，经常一起探讨PPT应该怎么做，老师可能会问什么问题，以及共同思考演讲主题（语音识别）的相关知识。在这个过程中我们合作默契，并找到了不少提高效率的方法，最终的报告结果也比较圆满。在学习过程中，我用了十五天左右的时间完成了个人职业规划的内容。在明确自己的目标和将来努力的方向后，我感到自己学习的动力更加充足了。对未来发展的迷茫感减少了。我认为计算科学导论与新生研讨课的课程目的类似，都是为了了解自己专业整体形势，让我们明白自己因何而学，为何而学。

## 5 附录

gethub:个人网址<https://github.com/upc-ly?tab=repositories>

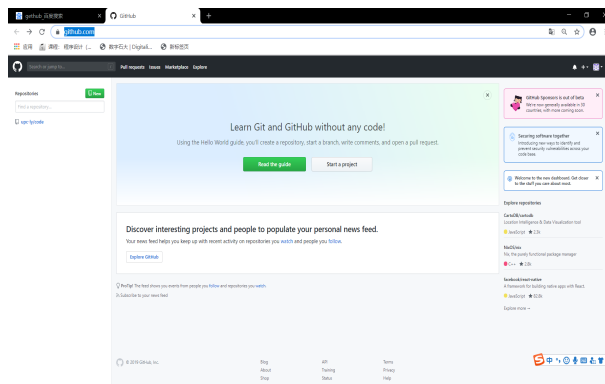


图 3: 网站截图

bilibili:



图 4: 截图

观察者：

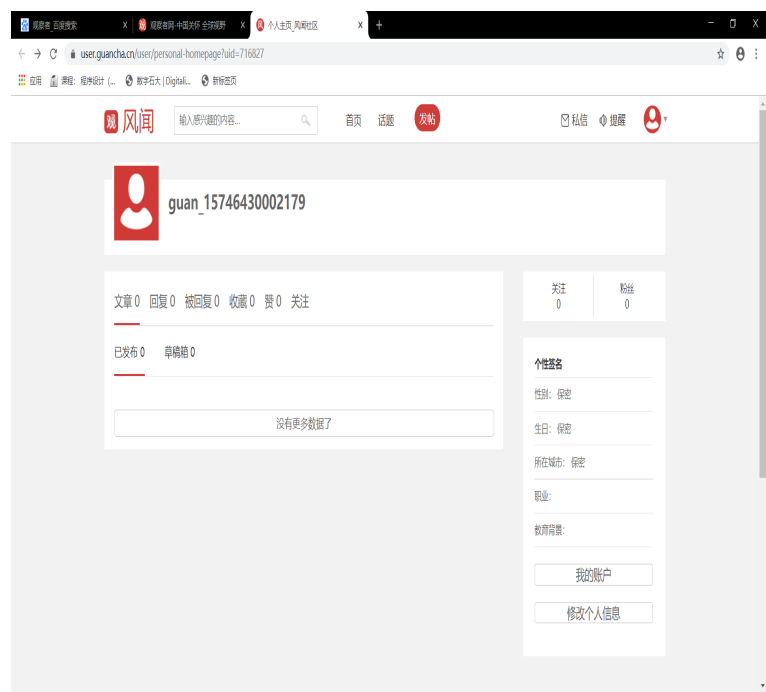


图 5: 截图



学习强国：

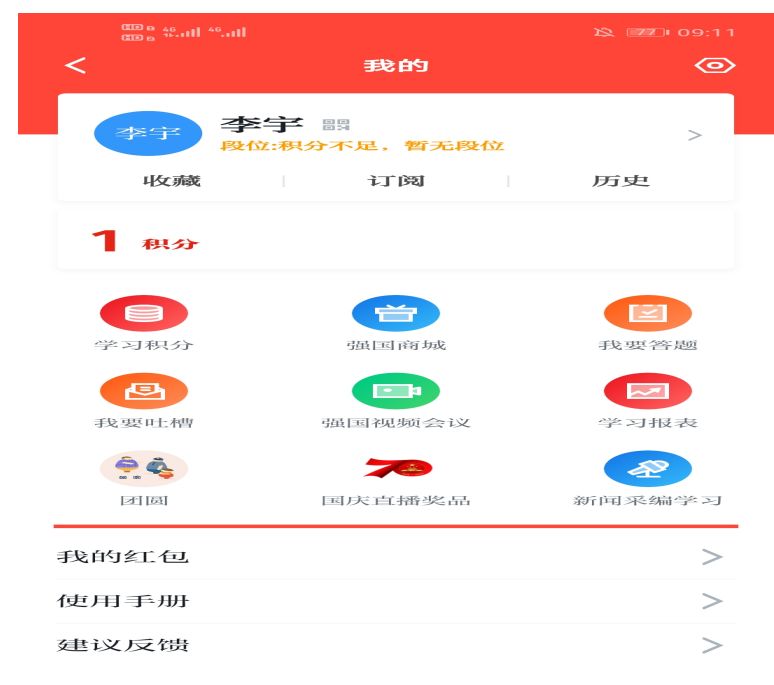


图 6: 截图

csdn:https://blog.csdn.net/qq45738761

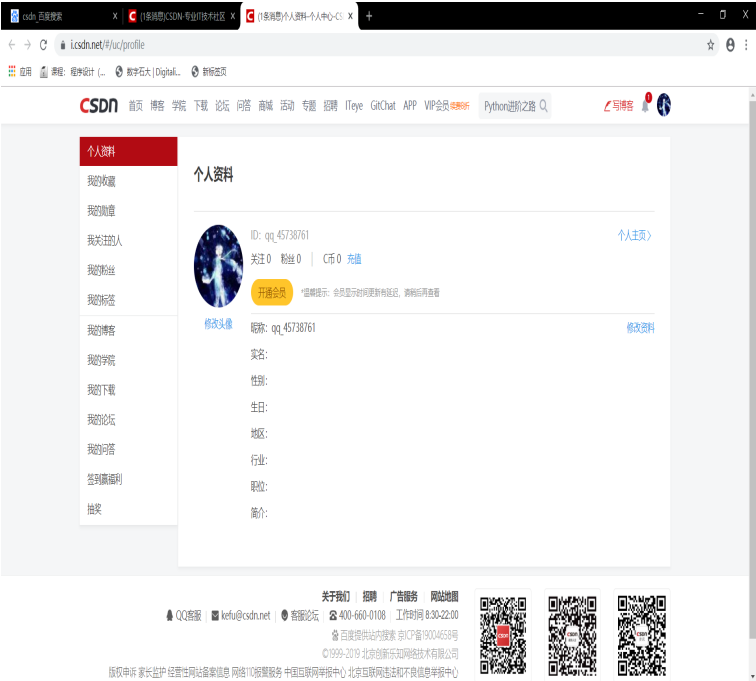


图 7: 截图

博客园: <https://home.cnblogs.com/u/1879174/>

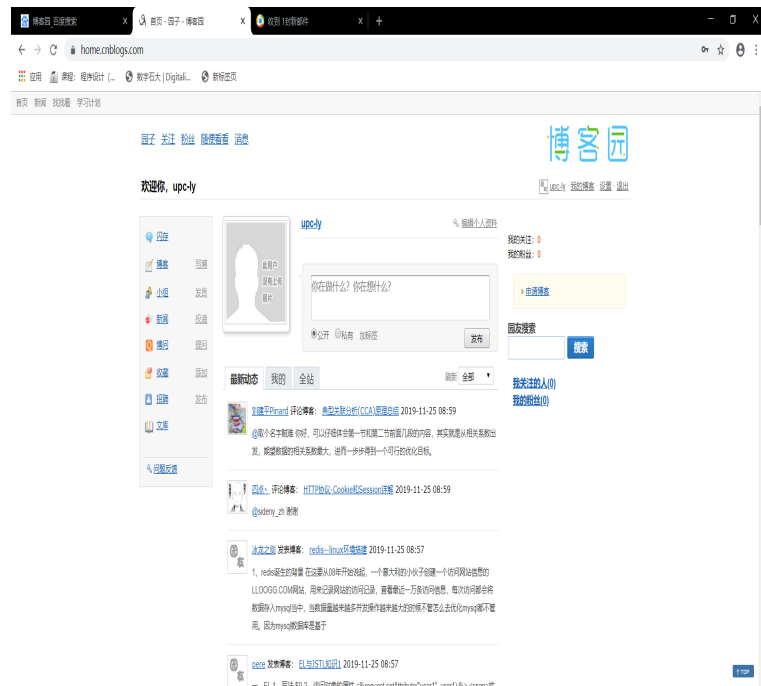


图 8: 截图

小木虫: <http://muchong.com/bbs/space.php?uid=19897831>

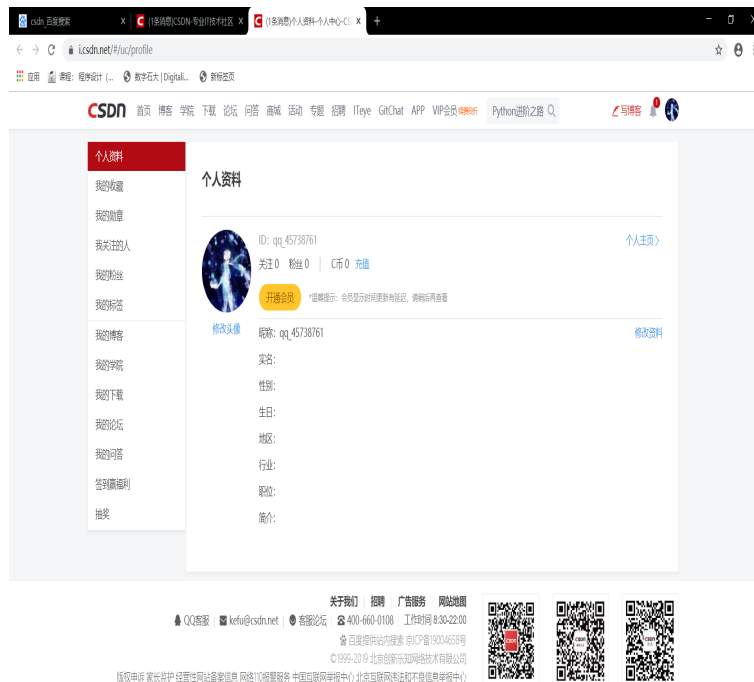


图 9: 截图

## 参考文献

- [1] SoroushMehri KundanKumar IshaanGulrajani RitheshKumar ShubhamJain JoseSotelo AaronCourville. An unconditional end-to-end neural audio generation model. 2017.
- [2] Aaron van den Oord Sander Dieleman Heiga Zen Karen Simonyan Oriol Vinyals Alex Graves Nal Kalchbrenner Andrew Senior Koray Kavukcuoglu. A generative model for raw audio. 2016.
- [3] 刘河生, 高小榕, 杨福生. 隐马尔可夫模型的原理与实现. 2015.