



中国石油大学 (华东)
CHINA UNIVERSITY OF PETROLEUM

《计算机科学导论》课程总结报告

学生姓名: 于金龙

学 号: 2007010326

专业班级: 计科 2003

学 院: 计算机科学与技术学院

课程认识 30%	问题思考 30%	格式规范 20%	IT 工具 20%	Latex 附加 10%	总分	评阅教师

2020 年 12 月 26 日

1 引言

计算科学导论一书，由于我们大一新生尚缺乏学习后续课程必要的基础知识，而该课程在学时数上的安排又偏大，从而导致在内容上增加了不少本来应该由其他课程（包括实验课程）承担的教学内容，如计算机操作命令与汉字编码等，也增加了一些不应该进入教学计划与课程体系的内容，如某些高级语言和某些数据库系统语言及其应用等。其中有一部分时进入高年级后如果有必要，能够很自然，很容易地自学掌握的知识。由于本书写法贯彻让不同水平的读者都有收获的文学创作原则，此所谓“深者得其深，浅者得其浅”的作品境界。因此，在以后学习的各个阶段，都应该时时翻阅本书，一边得到新的理解。

2 对计算科学导论这门课程的认识、体会

计算科学导论给我的整体认识就是全面。第一章引论中介绍计算科学的由来，学科方法论，一般的科学思想方法等。我印象最深的是孙老师估算“青岛市一共有多少钢琴调音师”，一开始我是懵逼的，知道孙老师用科学的方法引导我们思考这个问题，详细的步骤我大致还记得：第一步是了解青岛市的总人口有多少，第二步是假设每 $3/4$ 个人为一个家庭，假设 10 个家庭中有 1 个家庭有 1 架钢琴，从而估算出青岛市钢琴总数，第三步是每个钢琴调音师都会有固定的客户，由于钢琴是精度要求比较高的乐器，假设每个家庭每个月都需要维修钢琴，在钢琴调音师理想工作的情况下，估算每个钢琴调音师每个月能维修的总钢琴数，第四步则是将总钢琴数与每位钢琴调音师每月所能维修的钢琴数相除，最终估算出青岛市的钢琴调音师数量级。也就是从那一节课开始，我对计算科学导论产生了浓厚的兴趣，在通过计算机解决问题时，不仅要考虑数据与数据之间的联系，更要对社会环境有一个认识，例如在估算“青岛市有多少钢琴调音师”时，就需要了解青岛市的总人口，青岛市的家庭组成，钢琴调音师这一职业的情况，钢琴对调音的需求等。

第二章计算科学的基本概念和基本知识中内容则要丰富得多，包括计算模型与二进制，存储程序式计算机的基本结构与工作原理，算法，过程与程序，高级语言，程序设计技术与方法等，在孙老师离开期间，李昕老师曾给我们介绍过二进制，包括将一个数进行进制转换的除 2 取余法，除 10 取余法，而真正让我认识到二进制魅力的则是李昕老师玩过的猜数字游戏，我记得大致流程是这样的，首先让我们心中想一个在 1-63 范围内的数字，然后依次放出几张 PPT，如果有就答是，没有则答否，最后真的猜对了，一开始我们猜测难道这是一种二分法，最后李昕老师揭露了其中的原理——二进制，每一页的数字都有一个共性，它们的二进制的某个数都为 1，通过回答是否，就可以判断出当前位置为 1 还是为 0，最后将二进制转换为十进制，就可以做到百发百中。而另一位老师则是给我们讲解了或与门电路，并示范了将两个数相加时，电路时如何执行的。通过两位老师的拓展，让我对二进制有了深刻的认识，二进制的 0 与 1 代表假与真，通过电路可以进行代数运算，而通过且，或等操作又可以进行逻辑运算，而计算机的理论基础正是逻辑与代数。

3 进一步的思考

我觉得计算导论让我最受益的环节应该是分组演讲，不仅锻炼了我们的能力，更拓宽了我们的视野。

在上大学之前，我的老师家人仅仅是告诉我计算机是最赚钱的专业，这似乎就是他们认为我应该选计算机专业的全部理由了，而我在上导论之前也是这么认为的。知道孙老师告知我们将会进行一个分组演讲，给我们出了 1732 道题，其中包括安全，操作系统，大数据，分布式计算，服务计算，互联网技术，机器学习，计算机视觉，计算机图形，计算机网络，计算理论，架构，开发与运维，量子计算机，区块链，人工智能，人机交互，软件工程，数据安全，数据库，深度学习，搜索引擎，算法，网络，物联网，系统架构，信息安全，学习路线，用户界面，游戏设计，云计算等大类。看完这些题目，我觉得计算机专业是宇宙第一专业似乎一点也不夸张。

我们所选的分组演讲课题是“声纹识别”，也称为“说话人识别”。相比于指纹识别，声纹识别似乎略显陌生，严格来讲，虽然声音并不具有真正意义上的纹理，但每个人的发音器官包括声带，声管等在大小和形状上都会有所差异，同时由于性别，年龄和地域的影响，使得我们每个人都有着不一样的声音。广义上来讲，所有可以区分每个人不同声音的特征，都可以称为“声纹”。由于这些特征的存在，声纹和指纹一样，衍生出各种使用的技术。

3.1 声纹识别分类

声纹识别分两种，即 1: 1 和 1: N，其中我对 1: 1 的理解是一一对一，说话人事先在声纹库中录入了自己的声音，说话时只需要说一句话就可验证自己的身份，比对时，将验证音频与注册音频通过算法对比，就可实现 1 对 1。1: N 则是在一个声纹库中已注册了 n 个人的声纹特征，搜索时，只需要目标人员的声纹特征，就可以与声纹库中的 n 个声纹特征通过算法对比，找到目标人员或者缩小目标人员的范围。如下图：

说话人识别的研究始于 20 世纪 30 年代，早期的工作主要集中在人耳听辨实验和探讨听音识别的可能性方面，在上世纪 60 年代，Bell 实验室的 L.G.Kestor 等人通过研究语谱图发现，同一个人所发同一个音的语谱总是比不同人发出相同音的语谱更相近，据此他用目视语谱图的方法进行说话人识别，并于同年文章中提出“声纹”的概念。随后在 1963 年，Bell 的实验室的 S.Pruzansky 提出的基于模块匹配和统计方差分析的说话人识别方法，引起信号处理领域许多学者的注意，兴起了说话人研究的高潮。

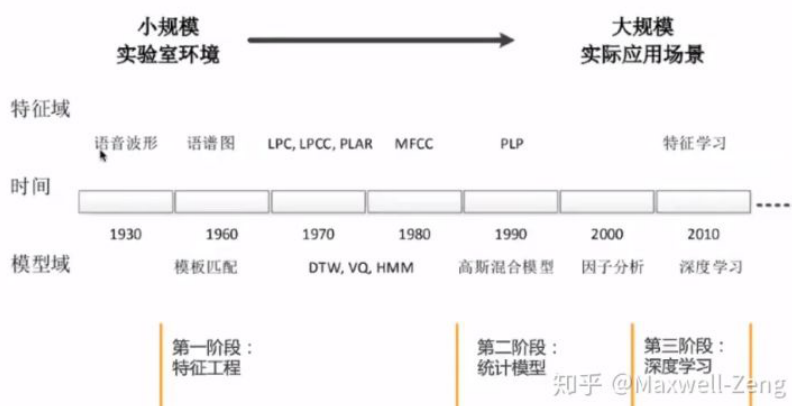
从 20 世纪 70 年代末至 80 年代末，说话人识别的研究重点转向对声学特征参数的处理及新的模块匹配方法上。研究者相继提出了线性预测代数 (LPC)，线性预测倒谱系数 (LPCC)，梅尔倒谱系数 (MFCC)，和感知线性预测系数 (PLP) 等说话人识别特征参数，与此同时，动态时间规整法 (DTW)，矢量量化法 (VQ)，隐马尔科夫模型 (HMM)，人工神经网络法 (ANN) 等技术也被陆续提出，并被广泛应用到说话人识别中，进一步提高了说话人识别性功能。

20 世纪 90 年代以后，尤其是 D.Reynolds 对高斯混合模型 (GMM) 做了详细介绍后，GMM 以其简单，灵活，有效以及较好的鲁棒性（健壮性），迅速成了当时与文本无关的说话人识别中的主流技术，将说话人识别研究带入了一个新的阶段。2000 年，D.Reynolds 在说话人确认任务中提出了高斯混合模型—通用背景模型 (GMM—UBM) 结构，为说话人识别从实验室走向实用做出了贡献。

► 工作方式

1:1 比对 — 说话人确认 *Speaker Verification*。确认两者是否一致，即你是不是你

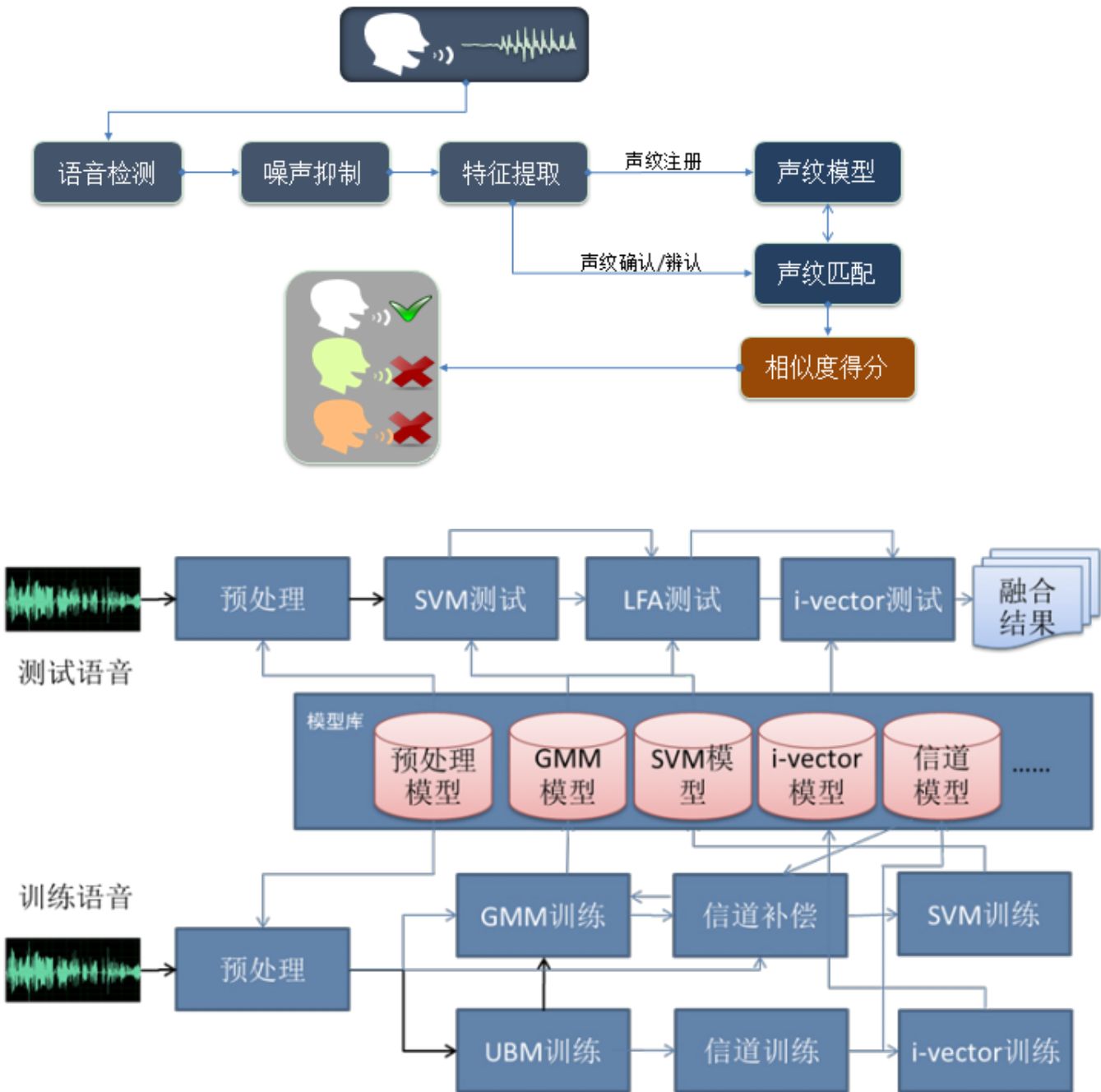
1:N 比对 — 说话人辨认 *Speaker Identification*。搜寻出哪儿个是你，即你是谁



进入 21 世纪，在传统的 GMM+UBM 方法上，Campbell 等人发现将 SVM 用于高斯混合模型十分有效。进一步，P. Kenny、N. Dehak 等人先后提出了联合因子分析（JFA）和 i-vector 模型，将说话人模型映射到低维子空间中，克服了 GMM+UBM 系统中高斯分量互相独立的局限性，提高了系统性能。为进一步提高模型的区分性能力，相关的区分性训练方法也应运而生（PLDA 等，即得到向量后，用什么方法对两个向量进行相似度比较）。

2010 年开始，随着计算机计算能力的增强，运用深度学习方法解决说话人识别问题变得越来越受学术界的重视。可以利用深度神经网络（DNN）自动进行特征提取，也可以在传统的 i-vector 的基础上将 DNN 作为后端分类器对所提取的特征进行分类，或者直接搭建一个端到端的说话人识别网络。在 2014 年，Ehsan Variani 等人用 DNN 对频谱图自动进行特征向量提取，并将提取出的向量命名为 d-vector。这个网络十分简单，就是将一段话每一帧的频谱图分别送入 DNN，然后将输出层的前一层作为 d-vector，然后对每一帧的 d-vector 进行平均，最终所得即为代表这句话说话人的 d-vector。然后 2015 年，Yu-hsin Chen 等人将卷积神经网络

络应用于文本依赖的说话人识别中，并取得了不错的效果（目前为止，两大主流之一就是使用 ResNet 进行说话人识别）。紧接着，在 2017 年，David Snyder 等人提出了著名的 x—vector，其是在 TDNN 结构上提取出来的，其网络中的时间池层在输入语音上进行聚合，从而捕获说话人的长期特征。这使得网络能被训练来区分来自不同长度的语音片段的说话者（这是另一主流方法，从 18 年开始，各种比赛 VOXSRC，NIST SRE 等的基线方法就是 x—vector，现在很多性能卓越的方法就是在此基础上改造的）。



以上为常见的几种生物特征识别及其特征对比，其中最为熟悉的应该是指纹识别，人脸识别。

特征	指纹	掌型	视网膜	虹膜	人脸	静脉	声纹
易用性	高	高	低	中等	中等	中等	高
准确率	高	高	高	高	高	高	高
成本	高	非常高	非常高	非常高	高	非常高	低
用户接受度	中等	中等	中等	中等	中等	中等	高
远程认证	不可	不可	不可	不可	不可	不可	可以
手机采集	部分可	可以	不可	不可	可以	不可	可以

美国圣何塞州立大学-国家生物特征测评中心，“Fundamentals of Biometric Technology” at http://www.gngr.sjsu.edu/biometrics/publications_tech.html for good discussion and comparison of biometrics discussion and comparison of biometrics

易用性方面：指纹识别应该是最高的，而声纹识别需要录入不小于 2 秒的有效文本内容，在注册与验证时都至少需要 2 秒，相比于指纹识别的一点即识，声纹识别的易用性明显不足。

准确率方面：生物特征识别，由于每个人都有独一无二的基因组成，在不同基因与环境所影响下形成的生物特征是独一无二的，理论上来说，每种生物特征识别的准确率都非常高，但是，声纹作为一种物理现象，不仅是受发声者本身的声带影响，更易受到外界环境的影响，通俗的来讲，就是声作为一种波，会发生波常遇到的干涉与衍射现象，当不同的波相遇时，就会加强或减弱，导致注册或验证时发生错误，因此，只有在理想的，安静的环境下，并且说话人发音准确才能使准确率提高到百分之九十八以上。

成本方面：声纹识别只需要手机自带的麦克风，而指纹识别则需要手机上加一块电容版。

远程认证方面：声纹识别最区别于其它生物识别特征的优势就是可以进行远程认证，只要有麦克风，就可以通过网络传到另一端，在另一端进行声纹识别。

3.2 声纹识别的应用与前景

声纹识别技术目前在公检法，金融等领域里应用比较广法，主要有以下几个应用场景和解决相应问题：

公检法：

1. 重点人员声纹采集和建库

建立重点人员声纹数据库，在 110 接警、重点区域范围、重大活动期间等，一旦发现重点人员、黑名单人员声纹信息，即进行预警，有效进行事前预防，目前行业做的比较好的有快商通、科大等，快商通的声纹采集器已通过公安部质量检测。

2. 侦察破案

利用声纹识别技术海量筛查优势，进行“案查人”、“人查案”、“案查案”与“人查人”等多种排查方式，缩小侦查范围，提高办案效率。

3. 反电信诈骗

利用声纹鉴定技术对电信诈骗等案件中的涉案语音进行个体、团伙的识别，确定犯罪嫌疑人身份，为侦查破案、案件诉讼提供技术支撑。

4. 治安防控

利用“语种识别”，“内容识别”，“声纹特征识别”等声纹综合分析技术，对重点人员进行布控，一旦出现立即进行关注控制。

金融行业：

1. 登录，支付场景

采用声纹识别技术，自动匹配用户个人身份信息，完成登陆、支付的身份验证，一般采用文本相关的方式，既 8 位随机动态数字串或者固定文本。

2. 业务核验身份

采用声纹识别技术，在业务沟通中完成用户身份核验，在自动匹配业务办理的信息，进行比对，完成业务办理的身份核验，一般采用文本无关方式，如开卡开户。

3. 信贷场景

采用声纹识别技术，在信审环节对用户身份进行识别，并查验是否为黑中介（黑名单用户），完成信审身份审核，采用文本无关的方式。

4. 金融洗钱

采用声纹识别技术，在判定出疑似洗钱行为后对用户进行电话远程身份验证以及自动对用户信息核对，完成可疑用户身份核验，采用文本无关的方式。

3.3 声纹识别的发展趋势

声纹识别研究朝着深度学习和端到端方向发展

语音作为语言的声音表现形式，不仅包含了语音语义信息，同时也传达了说话人语种，性别，年龄，情感，信道，嗓音，病理，生理，心理等多种丰富的副语言语音属性。以上这些语言语音属性识别问题从整体来看，其核心都是针对不定时长文本无关的句子层面语音信号的有监督学习问题，只是要识别的属性标注不同。

近年来，声纹识别的研究趋势正在快速朝着深度学习和端到端学习方向发展，其中最典型的就是基于句子层面的做法。在网络结构设计，数据增强，损失函数设计等方面还有很多工作去做，还有很大的提升空间。

1. 提升声纹识别系统的短时语音情况

在实际应用中，由于对基于语音的访问控制需求的不断增长，提升声纹识别系统在短时语音情况下的性能变得尤为迫切。短时语音中说话人信息不足以及注册和测试语音的文本内容不匹配，对于主流的基于统计建模的声纹识别系统是一个严峻的挑战。

2. 改进现有的深度说话人学习方法

目前采用的深度说话人识别方法首先利用神经网络提取前端的帧级特征，然后通过池化映射获得可以表示说话人特性的段级向量，最后采用 LDA/PLDA 等后端建模方法进行度量计算。相对于传统的 i-vector 生成过程，基于深度学习的说话人识别方法优势主要体现在区分性训

练和利用多层网络结构对局部多帧声学特征的有效表示上。如何进一步改变现有的深度说话人学习方法是现阶段的一个研究热点。

3. 深度对抗学习在声纹识别中的应用

生成式对抗网络 (GAN) 的主要目的是用在数据生成, 降噪, 等很多场景里面。它还被用在领域自适应里, 形成一个新的分布。第三个广泛的应用是生成对抗样本, 这会对分类系统产生很大的困扰。很多研究者用对抗样本攻击机器学习的系统, 在原始数据上增加一些扰动, 生成样本, 经过神经网络后就有可能形成完全不同的结果。这个思想在图像处理领域非常活跃, 会造成错误识别, 引起自动驾驶, 安全等领域的研究人员的广泛关注。

4. 深度嵌入学习是进行声纹识别和反欺骗的一个重要途径

说话人识别和欺骗检测近年来受到学术界和业界的广泛关注, 人们希望在实际应用中设计出高性能的系统。深度学习的方法在该领域得到了广泛的应用, 在说话人识别和反欺骗方面取得了新的里程碑。然而, 在真实复杂的场景下, 面对短语音, 噪声的破坏, 信道失配, 大规模等困难, 开发一个健壮的系统仍然是非常困难的。深度嵌入学习是进行说话人识别和反欺骗的一个重要途径, 在这方面已有一些显著的研究成果。如之前的 d -vector 特征和当前普遍使用的 x -vector 特征。

这里是简单列表的样例: (如果需要标号自定义或者自动标记数字序号, 请自行搜索语法)

4 总结

通过对计算导论的学习, 让我对计算机这门学科的认识从这是一门赚钱的专业上升到了这是一门对世界有重大影响的专业, 在孙老师布置的 1700+ 道题目中, 与计算机相关的大类就有分布式计算, 服务计算, 互联网技术, 机器学习, 计算机视觉, 计算机图形, 计算机网络, 计算理论, 架构, 开发与运维, 量子计算机, 区块链, 人工智能, 人机交互, 软件工程, 数据安全, 数据库, 深度学习, 搜索引擎, 算法, 网络, 物联网, 系统架构, 信息安全, 学习路线, 用户界面, 游戏设计, 云计算等, 而每个大类下又可细分, 计算机的未来似乎比我想象的要多, 希望通过大学四年的学习, 能掌握更多的专业知识, 以便投身其中的一门。

5 附录

参考文献:

- [1]CSDN. 喜欢打酱油的老鸟. 一文看懂“声纹识别 VPR”
- [2] 阿里聚安全. 探秘身份认证利器——声纹识别
- [3] 知乎.Maxwell—Zeng. 说话人识别 (声纹识别) 发展史简单介绍 par [4] 吴玺宏. 声纹识别听声辨认. 计算机世界. 2001. (8).
- [5]《计算科学导论》, 赵志琢著, 1998.
- [6]Kersta L G. Voiceprint identification[J]. Nature, 1962, 196(4861): 1253-1257.
- [7]Pruzansky, S.. “Pattern-Matching Procedure for Automatic Talker Recognition.” Journal of the Acoustical Society of America 35 (1963): 354-358.
- [8]Atal B S, Hanauer S L. Speech analysis and synthesis by linear prediction of the speech wave[J]. The journal of the acoustical society of America, 1971, 50(2B): 637-655.

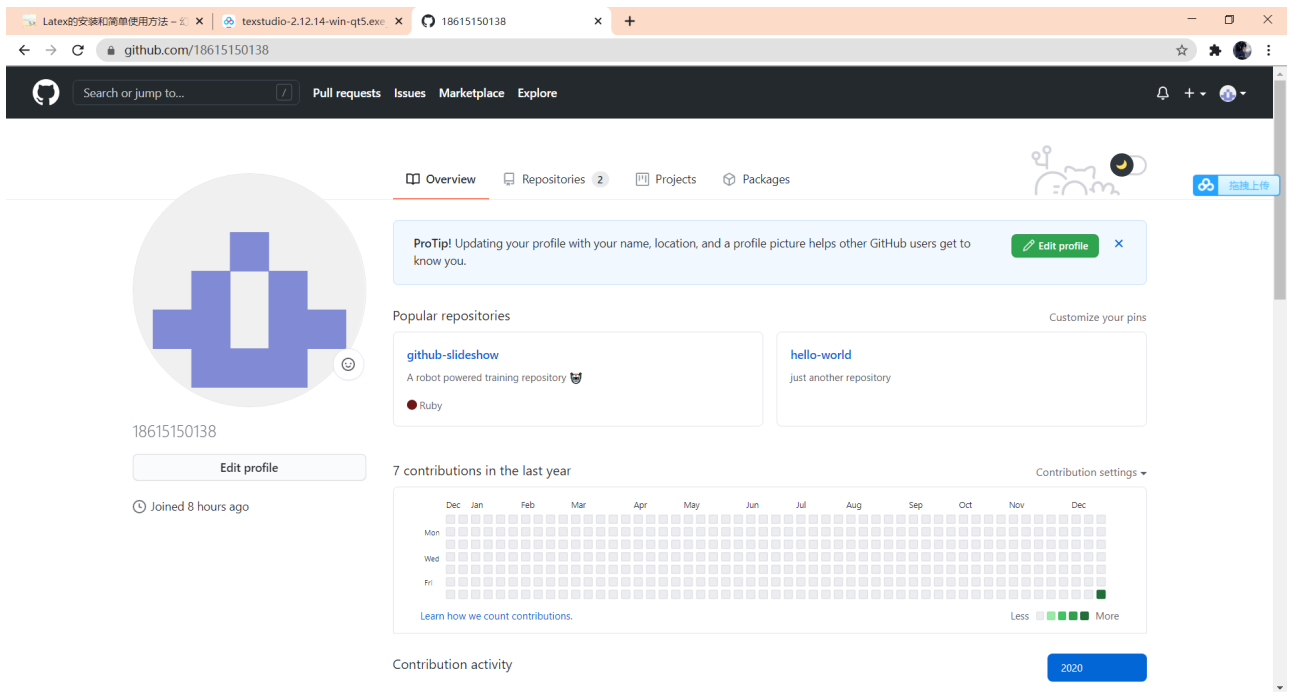


图 1: Github 账号

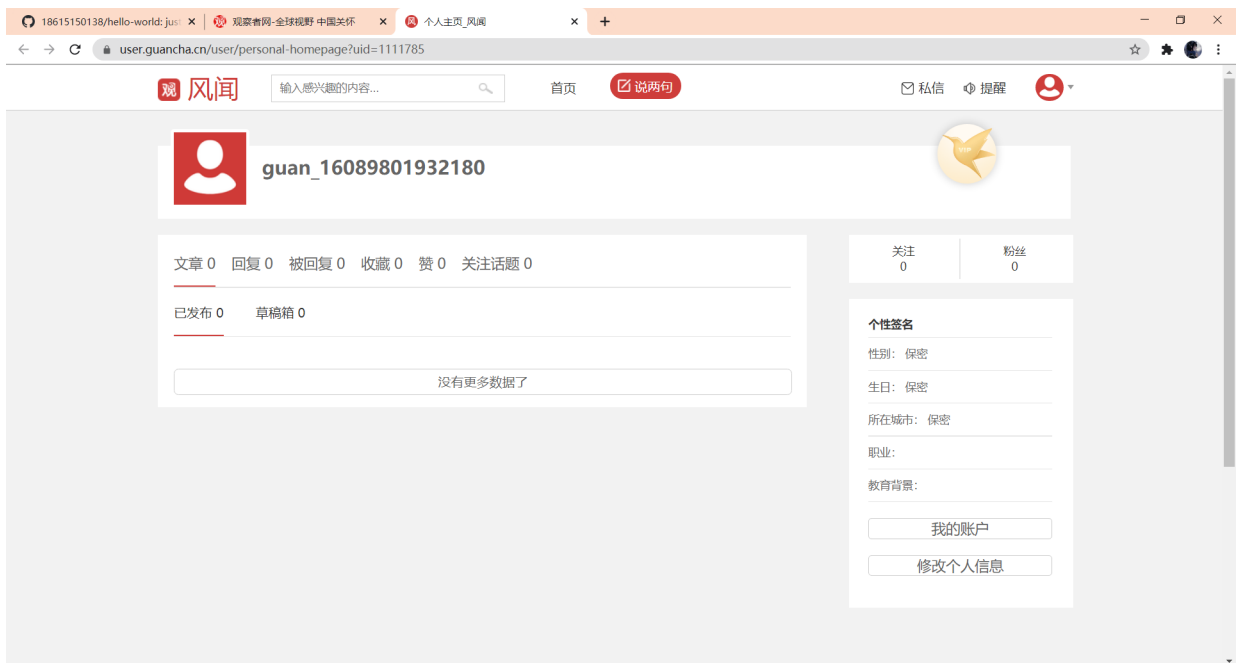


图 2: 观察者账号



图 3: 学习强国账号



图 4: 哔哩哔哩账号



图 5: CSDN 账号

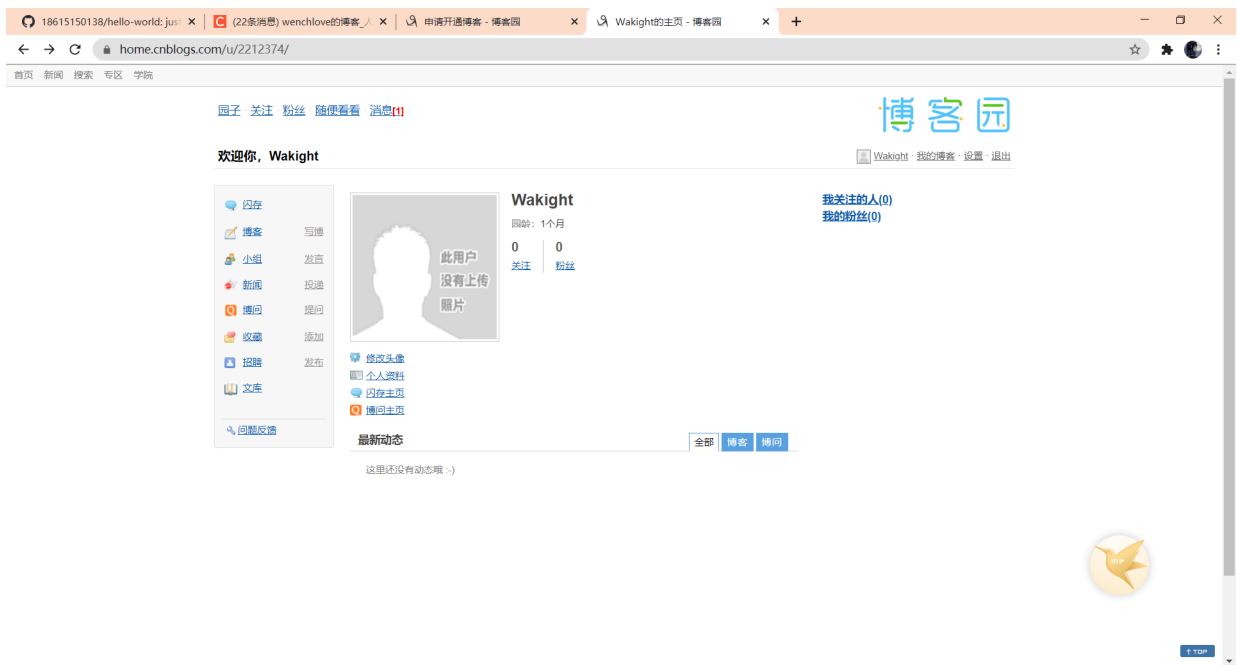


图 6: 博客园账号



图 7: 小木虫账号