

《计算科学导论》课程总结报告

|  |  |
| --- | --- |
| 姓 名 | 柳昂 |
| 学 号 | 2007010114 |
| 专业班级 | 计算2001 |
| 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1月8日

**1 引言**

作为一个刚刚进入大学学习计算机的20级计算机小白，面对自己所选的计算机科学与技术专业，以及这个专业所要学习的东西，我是感到模糊和迷茫的。选择计算机专业，很大程度上是因为计算机专业是现在的热门专业，是身边人口中的“高薪职业”，是老师们建议选择的专业，自己对这个专业的认识也仅仅停留在百度百科的国家一级学科，除此之外，几乎一无所知。因此我在开学时就希望有一门课能够带领我真正系统性地认识计算机科学与技术这门专业，了解这个专业的特点，本质，以及以后可能遇到的分岔口，让我可以做出合理的选择。

因此，我对孙运雷老师的计算科学导论这门课十分期待，虽然孙运雷老师因为学科建设的事情把课程推迟了一段时间，但是这更让我珍惜这来之不易的课程。孙运雷老师有着很高的学历，发表过10余篇论文，拥有1项专利，资历深厚。在孙运雷老师的授课过程中，我对计算机科学与技术所代表的知识体系有了重新的认识，对这个专业所包含的学科方法论有了更深的了解。从某种程度上来讲这门课程的意义深于其他基础课程。

下面就开始详细总结一下我对计算科学导论这门课的认识、体会和在分组演讲时对所讲课题的一些问题的思考和延伸。

**2 对计算科学导论这门课程的认识、体会**

《计算科学导论》重点落在“导论”两字，是一本引导性教材，目的在于从科学哲学的角度用高级科普的形式为初学者提供一个了解和学习计算机科学和技术的导引。课程中所介绍的学科基本概念和基本知识，明确了计算机科学与技术是什么，点明它的定义、特点及范畴；而对每一个概念以时间为线的介绍方式有利于整个知识体系的建构。全书以学科方法论为切入点，其中内含的是哲学层次的人才成长规律及科学发展规律，这也是与我们自身发展关联最为密切的一部分。下面就我对学习了这门课程之后在科学与哲学方面的认识和其对于我自身的指导意义做出展开详述。

**2.1 计算科学**

计算科学有着狭义和广义之分。狭义的计算科学指称的就是计算机科学与技术，其研究内容覆盖了对计算问题的一般研究。计算机科学与技术的研究与学科发展历经十分曲，从20世纪30年代到60年代初，计算机科学与技术的研究与学科发展基本上处在萌芽状态。当时研究计算机科学与技术的科学家主要工作是围绕什么是计算开展理论探索，寻求计算的数学理论模型，弄清计算的极限。40年代时，在图灵和冯·诺伊曼等人的努力下，储存程序式通用电子数字计算机（图2.1-1）诞生了，这极大地推动了计算数学的快速发展。在此后计算机的发展过程中，硬件、软件不断与理论融合，计算科学变成了一个拥有完整体系，可以单独作为一个学科来开设的领域。而经过不断调整，涵盖计算这一领域研究工作的一级学科专业名称为计算机科学与技术。

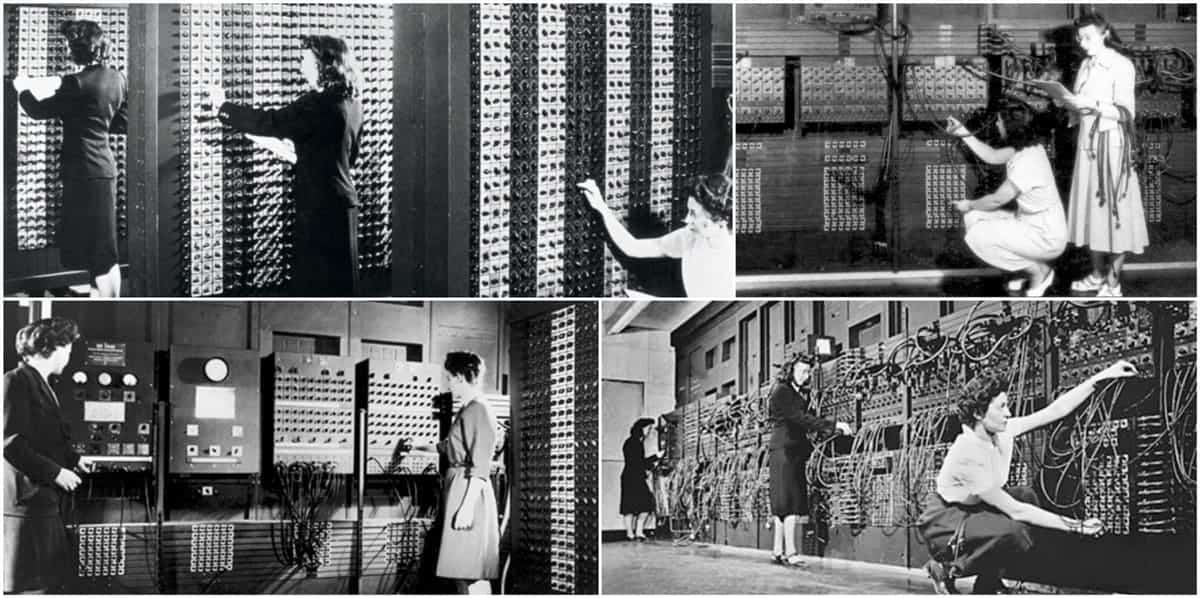


图2.1-1

支撑计算科学的基础知识体系是十分庞大的：计算模型与二进制、储存程序式计算机的基本结构与工作原理、数字逻辑与集成电路、机器指令与汇编语言、算法、过程与程序等等。而在孙运雷老师的带领下，我通过第二章的学习，对这些知识进行了初步的了解。但我明白，我所学的只是冰山一角，如今的计算科学枝繁叶茂，从计算的数学理论到计算机组成原理、器件与体系结构，从计算机应用基础到基本应用技术，面对如此多的内容，我深手学业压力之重，但也有着所承比他人更多的骄傲感。

相对于狭义的计算科学，广义的计算科学包含的内容要广得多，它不仅覆盖了计算机科学与技术的研究范畴，而且还包含着更多的内涵。我们知道，历史上一共有两大学科类形态，理论与实践。而计算，则成为了第三种学科类形态，但它并未独立存在和发展，依然在借助理论科学和实验科学以获得自身的发展。随着信息科学与技术的发展，在图形学与图像处理、数据库系统、人机界面、人工智能、虚拟现实等技术的支持下，人们已经开发了许多计算可视化的系统，使计算机可以模拟真实处理问题的实验场景或人的思维过程来解决许多问题。

**2.2 科学哲学与学科方法论**

学科形态和学科类形态属于科学哲学的范畴，那么什么是科学哲学呢？

在19世纪中叶以前，自然科学的发展一直是依附在哲学的母体之中的，在亚里士多德的时代，自然科学的发展主要从古典哲学中获得科学方法，演绎逻辑成为科学发展最为重要的逻辑基础，三段论成为判断科学知识真理性的重要方法。而慢慢的，随着自然科学的发展，科学的思维亟需要被拓展，便有了后来的学科方法论的探索。从归纳方法到逻辑经验主义和演绎逻辑主义，真前提的基础，基础的基础从何而来？这种对学科最原始的基础的可靠性或真理性进行研究的学问称为该学科的本体论，后分为“自然科学哲学”和“专门科学哲学”。

在自然哲学建立的初期，哲学曾给予自然科学极大的帮助与推动，但与此同时自然科学快速发展，哲学对于其的领导作用就大大降低了。我们常说，理论指导实践，可当理论走在实践的前端时，自然科学的根基是不稳定的，大厦将倾的局面必然发生。在解决种种问题的过程中，科学哲学应运而生，为科学提供着支撑。

这一过程于今日的科学发展依旧是有着重要意义的，就像老师上课所说的，如今的芯片制造行业，大多依靠着实践经验，缺乏理论的支撑，所以在发展上步履维艰，企业的芯片业发展有时仅仅依靠个人，人才的流动决定着企业于这个领域内的兴衰，因为最核心的东西出于实践，而实践源于个人，这样的行业发展是极其畸形，极其危险的。与个人而言也是一样，没有理论支撑的实践，终将遇见瓶颈。

**2.3 对自身的指导意义**

《计算科学导论》第三章对计算科学的意义、内容、方法做了详尽的介绍，第四章讲解了如何学习计算科学和健康成长，而孙运雷老师也在课上给我们介绍了很多我们学校的培养方案，以及这次学科建设成功之后我们的学科培养体系将会更加科学、完善。通过这些，我深刻的感受到面对体系如此庞大的专业知识，科学、系统、循序渐进的学习的重要性，大厦欲成必得环环相扣，步步稳扎稳打，这也就对我自身提出较高的要求。

《计算科学导论》对于很多专业内容一笔带过，这也说明了我更需要通过四年的学习来入门。计算机专业的知识更新周期很短，我们不仅要学习深奥的理论知识，更要面对大量实践内容的学习。理论与实践相结合，理论与实践的统一是计算科学一大特点，它决定了我在学习中要经常不断地在严密的逻辑思维与形象的实验操作之间转换学习方式。虽然这是一条布满荆棘的道路，一条充满艰辛的人生之路，但我必须在这条路上坚定的走下去，才能在未来学科的高速发展中有所作为。

**3 广告推送**

**3.1选题背景和意义**

随着互联网的迅猛发展以及互联网用户数量的大量增加，网络购物变得越来 越流行，因此以互联网为载体的互联网广告应运而生。同时信息的暴增也催生了 大数据时代的来临，使得互联网广告投放的粗放性和无目标性弊端越来越明显， 这时定向广告随之而来，而以定向广告为核心的基于大数据的网络广告推送服务 也因此诞生。因为传统的互联网广告推送将广告投放给所有的用户，其依据仅仅是广告主的需要，并没有依据网页特征和用户兴趣投送相关类型的目标广告，这样一方面给广告商带来了金钱上的浪费，另一方面也给用户带来了不便。鉴于目前大数据时代已经来临，向目标用户群投放相关广告成了互联网广告发展的新方向。

**3.2 广告推送的趋势：精准推送广告**

精准广告是基于大数据时代下新生的广告名词，是一种广告的创新形式。学界对此定义尚未统一。随着互联网、大数据算法的不断发展，精准广告也被称为精准投放，其意义在不同的语境下不断延伸。鞠宏磊等学者将大数据精准广告定义为“依托互联网网络及广告交易平台，运用大数据信息检索功能、受众定向分析功能，以及数据挖掘等技术对目标消费者的网络数据进行实时抓取与分析，并针对目标消费者的个性化特征以及需求，进行高度相关性的商业信息的传播与沟通方式”。学者于婷婷指明精准广告是借助信息数字技术，通过搜集和分析用户个人数据对其需求和偏好进行预测，并据此投放具有针对性 营销内容的行为。学者张寒对精准广告的理解更为详细，将互联网作为研究背景，指出互联网精准广告是依托广告交易平台、供应平台、数据管理等不同的平台，进行针对性投放的精准广告营销。综合以上学者观点，将精准广告定义为：基于个性化、时效性、准确性等特点对消费者行为数据进行分析精准定位，从而推送符合消费者购买意愿的互联网广告。

**3.3 为什么现在的广告推送可以如此精准？**

这一切归根于大数据下的用户画像、用户行为的分析。在正确的时间向正确的人提供正确的内容，以达到最佳的营销效果。我们之所以可以精准收到推送是因为企业有一套客户数据管理系统，实时动态记录和统计客户的兴趣爱好和行为习惯，建立实时动态的标签管理。

**“在京东搜索过有品牌的电脑，之后我的微信朋友圈、知乎都出现过京东或者苏宁的相关广告”**

通过统一的数据ID，进行跨平台的采集客户数据，如京东、淘宝、微信、知乎、官网、头条、抖音、百度等，将这些采集的客户数据统一管理在客户数据管理平台。企业可以根据客户数据管理平台提供的用户画像进行分群管理，再通过业务中台实现BI自动化营销。这就是你前面所看到的在A网站浏览过的记录，B网站会给你推送相关的营销内容。

**“广告都这么"恐怖"了吗？很慌，这样个人隐私还有吗？”**

它只是将我们的ID做全渠道的SDK集成，建立基于这个ID的统一的用户画像。就像每个人的电脑都有属于它的IP一样，我们上网浏览了什么，能查到的是我们的IP的浏览记录，通过IP的浏览情况做推送，所以您的隐私还是在各个平台数据库里面做过加密保护措施的。

对于大多数企业来说，想实现精准的推送可以通过数据中台CDP实现时局的管理。

* **CDP数据中台可以帮企业采集所有数据源，通过直接集成（或批量处理）将来自各个渠道的数据集合在一起，为企业提供完整的、不断更新的客户画像，实现深度的客户洞察。**

**1. 数据的挖掘能力是基础**  
**id的识别：**通过唯一标识符填写姓名、地址、电话、电子邮件等重要确实字段，以了解他们是谁。  
**匿名访问者的数据：**通过使用cookie、设备id或IP地址解析身份和丰富相关信息，将匿名访问者数据转化为资产。通过筛选我们的网络流量，企业可以更有效地向潜在的和现有的客户营销或参与其中。  
**数据的伙伴关系：**第三方数据交换和发布的一站式服务。动态更新旧客户记录，附加关于客户和前景的附加信息，并积极防止数据衰退，而不必管理和协调来自不同第三方数据供应商的多个数据集。

**2. 360度客户画像赋能营销**  
  
每一个客户都是不同的，基于客户的档案、多重的身份、特征标签、消费记录、互动记录，等各种各样的信息构成客户的360度画像。  
  
通过客户的画像，对每一群“相似”的人进行细分，不断地细化颗粒度。细分能力的颗粒度是精准营销的基础。  
  
例如：  
**（1）、人群特征的细分：**人口属性、消费习惯、会员等级、地域等因素在一定的时间和空间的情况下对特定特征的人进行分群。  
  
**（2）、行为特征的细分：**企业可以通过客户的消费行为和非消费行为进行特征分析  
  
企业可以通过客户的消费行为（如：消费金额、购买商品、购买时间、购买渠道、购买场所、品牌、品类、款式、历史消费记录等行为）给客户绘制消费能力画像。通过数据模型，还可以预测客户未来的消费能力，根据客户的消费习惯和消费能力做个性化和定制化的营销。  
  
客户的浏览记录、来源渠道、浏览时间、停留时长等都是非常重要得行为数据，通过更精细得行为分析，我们可以知道这个前来的客户是否是我们的潜在客户。当然这一切都基于我们收集和整理了大量的行为数据的基础上通过数据分析模型分析出来的，形成深度的客户洞察能力，判断客户的潜在消费能力和兴趣爱好，引导客户消费。  
  
**（3）、模型的组合和优化：**通过不同维度的分析和组合发现更多的客户特征，如人群消费能力和人群特征的差异、行为路径对人群特征的变化影响、时间周期内的人群变化情况等因素进行人群的分析和能力模型的优化，以便企业更深入的洞察客户。  
  
CDP让数据更有价值，它是企业的眼睛，帮助企业更好地去洞察和研究客户，同时CDP为营销赋能。

**（4）、全渠道的营销手段**邮件、官网、短信、微信、公众号、小程序、支付宝、广告平台等都是企业与客户沟通的渠道和桥梁，企业可以通过业务中台统一进行管理、计划和执行与客户的进行互动营销 **①.基于用户画像，实现千人千面的营销**根据CDP用户标签分群，精准地将个性化内容推送给不同渠道的特征人群，实现全渠道的千人千面的营销。此时强大的内容库的管理就显得尤为重要。海量的沟通内容、短信内容，H5等等。  
**②.基于客户整个生命周期考虑的营销**不止千人千面的精准营销，还可以基于客户整的生命周期进行营销。对于品牌，每个客户都会经历一个完整的生命周期。从认知、了解、兴趣到购买使用产品的全流程生命，再到客户的获取、激活、活跃、衰退、流失的整个生命周期的管理都是非常重要的营销点。客户的沟通、激活和活跃需要基于整个生命周期考虑，以及内容的策划、设计、沟通，基于人群、渠道、时间的发布营销。  
**③.客户的价值分群和等级**根据客户的消费行为和消费能力进行进行等级和价值的分群，对于一些高消费能力的客户和低消费能力的客户推送的优惠活动都会有所不一样，根据客户的特性可以制定更具特色的营销内容。  
**④.效果的实时跟踪和评估**企业可以通过数据看板，数据大屏的形式实时跟踪营销的效果，及时分析和评估各个渠道效果和转化情况。

**3.4 网络定向广告推送的实现**

**3.4.1 网络定向广告推送服务平台**

目前在互联网定向广告推送服务平台的研究方面，国内外已经有了相应的研究成果，主要是由数据管理平台(DMP)和需求方平台(DSP)组成的定向广告推送服务体系。数据管理平台DMP (Data Management Platform)是一个技术平台，它把分散在各处的用户数据整合纳入到一起，通过特定的技术算法，确定用户的行为特征，然后建立准确的用户分类，最后通过用户分类将广告投放到指定的人群。

DMP使得广告主和广告代理对其广告受众的数据进行控制，这些数据使得需 求方拥有更加明智的媒体购买和广告计划。DMP收集了网民的兴趣属性数据，并 对数据釆取分析及存储操作，之后按照不同的属性将数据中的用户群体进行分类, 并与广告需求方平台(DSP)进行数据对接° DMP细分处于广告购买决策周期中 的各类消费者群体，以便向拥有相关需求的用户群体投放精准的定向广告。

DMP的主要功能如下：

1. 用户数据整合：吸纳整合用户数据，以便进行集中分析。
2. 数据分析：数据整合到一起之后，通过平台中大数据分析模块对用户 数据进行分析计算，解析出用户真实的消费意图，进一步对特定的人群进行细分， 为后续的推广活动做准备。
3. 数据迁移：一旦通过数据管理平台完成数据的收集、整合和分析，可以 进一步运用相关数据，与展示广告、搜索广告等广告生态圈的供应商们进行联合 工作。

可扩展性：面对大量各种各样的数据，互联网供应商和广告发布方需要 对它们照看和处理，DMP应当具备处理如此海量的数据以及并发分析的能力。

需求方平台DSP (Demand Site Platform)是一个广告竞价投放平台，早在2011 年的时候就已经在国内悄悄兴起，这标志着互联网广告发展越来越成熟。DSP首先接收数据管理平台中的分析结果数据，然后找到拥有投放意识的广告主，最终实现广告竞价投放的功能。

**3.4.2 网络定向广告种类**

定向广告的关键点是找到广告发送对象，即把广告发送给谁。为了有效地鉴 别发送对象类别，目前所使用的方式主要有两种：一种是基于网页内容的上下文 定向广告，另一种是基于用户上网行为的行为定向广告。

上下文定向广告(Content Targeting),又称作内容定向广告。互联网上的信息 复杂多变，我们认为如果某个网络用户浏览了一张网页，那么他很可能对该网页 上的信息感兴趣。上下文定向广告便是基于这样一个原理，分析网页内容，提取网页关键词，然后与广告之间通过关键词进行相似性比对，选出与网页内容最相 似的广告，最后投放给浏览该网页的用户。例如，如果您正在访问一个关于去美洲旅行的网站，突然看到一个弹出广告提供飞往加拿大的特价机票。这就是一个 典型的上下文广告。

行为定向广告(Behavioral Targeting)是指通过分析和跟踪用户行为，然后判 断其偏好和需求特征，最后选择恰当的营销信息与之进行交流。具体地说，行为定向广告分为三个步骤：首先收集用户的行为信息，主要是指用户浏览过的大量 历史网页；然后将内容相似的网页聚合到一起形成多个不同类别的网页集合，即为用户兴趣；最后选出最有价值的用户兴趣，并将代表用户兴趣的特征词与广告 关键词进行相似性比对，进而选出与用户兴趣最匹配的广告投放给用户。

**3.4.3 网络定向广告投放技术**

（1）上下文定向广告投放技术

针对上下文定向广告，通常认为网页内容即为用户的兴趣所在，其主要目的 是选出与网页主题相关度最高的广告，强调广告与网页内容关联度的重要性。它 需要对网页的内容和结构进行具体分析，自动地从网页内容中提取出网页主题， 然后快速有效地从广告库中选择出与网页内容最相关的广告。先前的研究人员已 经开展了一些卓有成效的研究工作。

Ribeiro-Neto和A.Lacerda等人提出了一种广告匹配算法，该算法利用遗传规 划算法来生成了一个匹配函数。实验表明该匹配函数在一定程度上提高了文本 广告与网页内容的匹配准确度。

Murdock等人又提出了一种新的上下文定向广告投放算法，该算法是基于机器 翻译的，其被用来解决文本广告与网页与之间的词汇失配问题。实验表明，与 基准方法相比较，机器翻译的方法使得广告匹配准确度得到了提高。

（2）行为定向广告投放技术

与上下文定向广告不同，行为定向广告是通过跟踪和分析用户的网络行为，进而提取出用户最有价值的兴趣点，最终向其投放与之兴趣匹配的广告信息。整 个过程的关键在于准确地分析出用户的兴趣点，分析用户的兴趣点首先要获取用 户浏览过的大量历史网页，然后通过相似网页聚合技术将内容相似的网页聚合到 一起形成一个网页集合，即为用户的一个兴趣点。而分析用户兴趣点的关键在于 能将内容相似的网页准确地聚合到一起，相似网页聚合准确性越好，提取出的用 户兴趣越精确，进而用户兴趣与广告匹配的准确度越高。

C.Wang, P.Zhang等人的研究工作充分证实了投放与用户兴趣相关的广告可以增加广告点击率，此外还能极大改善用户体验。P.Chatterjee等人也指出广告的针对性越强，效果就越明显。他们的研究工作证明了用户兴趣与广告相关的重要性。

Ribeiro-Neto等人提出了十种网页相似性聚合策略。这些策略都以向量空间 模型为基础，在这个模型中所有网页都被表示为同一空间中的向量，网页之间的 相似度由两个向量的余弦值来衡量。虽然所有网页都被映射到了同一个欧式空 间，但是由于向量空间模型的局限性使得网页相似性聚合过程仍然存在着严重的 误差问题。为了解决这个问题，作者提出了基于词汇扩展的聚合方法，该方法主要对目标网页的词汇进行相似性扩展，实验表明网页相似性聚合准确性得到了一定程度的提高。

另一种解决相似网页聚合的方法是通过从网页中抽取出能代表网页特征的关 键词，并将这些关键词集合表示为代表网页的向量，进而通过向量空间模型、词 汇相似度计算方法以及聚类算法进行网页相似性聚合。J.Goodman等人提岀了一种 用于提取网页关键词的算法，该算法考虑了词的长度、词的TF-IDF值、词在网页 中岀现的位置以及停用词等诸多因素，之后根据这些因素综合决定词的重要性，并根据这些词的重要性进行排序，最终选出能准确表述该网页特征的关键词。研究表明这是一种较为有效的相似网页聚合方法。

针对行为定向广告中的用户兴趣价值计算过程，浙江大学提出了一种全新髙 效的用户行为分析算法，该算法通过分析网页集合中的网页访问时间，进而计算 出用户兴趣的特征新鲜度和特征离散度，然后综合考虑这两个值得出用户兴趣的 价值权重，再按照价值权重高低对用户的所有兴趣点进行排序，最终选出比较有 价值的用户兴趣。实验表明这是一种较为高效的用户行为分析算法。

**3.5 未来AI广告展望**

未来，通过深度学习使用算法来模拟人脑的神经网络，从而快速地处理受众群体的资料数据，再针对用户习惯推送定制个性化广告。未来的人工智能将实现广告的自动整合，人工智能将为广告主提供广告文案、图片、视频、色彩搭配等资源，并根据该品牌的消费群体、竞品信息等与广告投放相关的数据，从而自动生成与广告主需求相匹配的广告。AI将利用数据库中的资源，事实地根据受众群体的消费习惯和行为特征，为不同的广告主创造组合出各种不同广告方案。而且广告主不止会收到一个版本的广告，AI将根据其需求和不同消费群体的特征提供具有独特性的广告方案。不仅如此，广告主还可以利用其积累的客户数据并结合人工智能，针对不同消费群体，定位符合其确切品味的产品，甚至还可以使用这些数据来预测未来将推出的产品是否符合用户的喜好。这样，广告推送便会更加容易精准。

以前，广告的成功度是通过CTR（点击率）和展示次数来衡量的，但对于人工智能所构建的个性化广告成功与否并不适合用这些来衡量。因为未来的广告并不需要消费者最终在品牌网站上完成交易。相反，我们将会看到嵌入在个性化视频广告中的互动聊天机器人，它将出现在视频中并通过互动挖掘出用户的真正需求，进而引导消费者进行购买。该智能互动机器人将伴随消费者完成从看到产品一直到订购付款的全过程。随着广告方式的改变，衡量AI智能广告成功与否的关键将会变成消费者在观看广告上花费的实际时间。随着这些创新在广告行业的普及，人工智能将不再只出现在科幻小说中，完全个性化的AI定制广告正在逐步成为现实。

**附录：**

*观察者*

****

*学习强国*



*哔哩哔哩*

*CSDN*

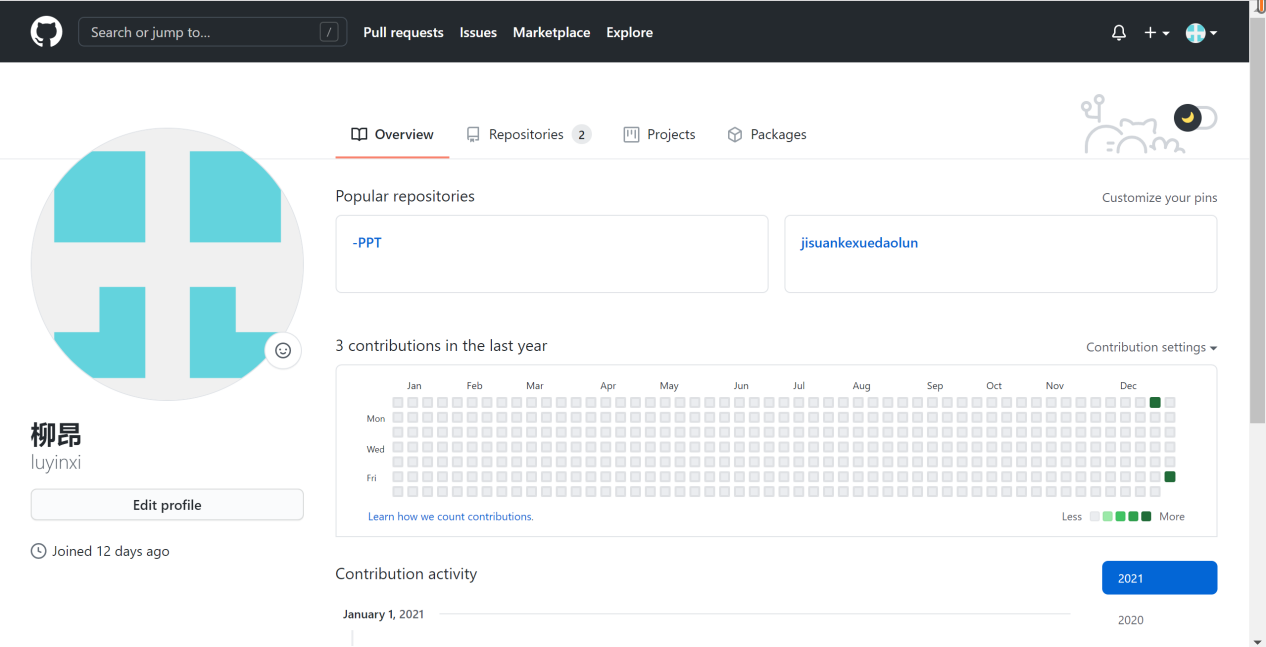


*[https://i.csdn.net/#/uc/profile](https://i.csdn.net/" \l "/uc/profile)*

*博客园*

*<https://home.cnblogs.com/u/2269337/>*

*GitHub*

**

*参考文献：*

[1]邢林森，基于Spark广告推送系统的设计与实现

[2]张良悦，大数据精准广告的伦理问题和路径分析:基于消费者感知的实证研究

[3]陈顺风，基于大数据的网络广告推送服务研究

[4]聚麦数智，知乎：《为什么现在的广告推送如此精准》

[5]duozhishidai,，CSDN：《为什么说广告的未来，需要人工智能来改变》