课程总结报告

计算2002 25 吴景新

1. 引言

“计算问题一直是伴随着人类的发展而发展，人类的发展离不开计算学科的发展。计算学科的发展促进着人类历史的发展。从原始社会的结绳而治到中国古代使用的算盘，再到齿轮计算器，再到现在的计算机，计算学科在发挥着越来越重要的作用。在现代的生活中，计算机几乎被应用到了方方面面，为人们的生活带来了更多的便利，更是节省了很多的工作量。计算机依然是一个年轻的学科，表现了很强的生命力，充满了无限的机遇和挑战。”

现在的大学计算机设有计算机科学与技术、软件工程、网络工程三大专业。其专业的主要课程有C语言程序设计、编译原理、操作系统、电工与电子技术、计算机网络、计算机系统结构、计算机组成原理、接口与通讯、离散数学、数据结构、数据库原理、数字逻辑、线性代数、专业导论、专业英语等；专业选修课有：概率论与数理统计、汇编语言、计算方法、人工技能、软件工程、信息检索与利用、JAVA语言、多媒体技术、分布式与并行处理、计算机控制系统、计算机图形学、面向对象技术、模糊逻辑与运用、嵌入式系统、数据挖掘、数字图像处理、算法设计与分析、通讯原理、网络操作系统、网络与信息安全、新技术专题等。这些课程几乎涵盖了计算机科学体系所有的知识与技术，它们为计算机科学的发展提供了大量的基础理论扎实、专业素质好、实践能力强、知识面宽的人才，同时这些人才为社会经济的发展做出了重大的贡献。

“计算机科学导论”就是为了给日益繁多而且日渐复杂的计算机学科提供一份总纲。也就是说，带给学生关于计算机学科的整体概念，为他们未来的道路指明方向。于此可知，一本关于计算机学科的导论有多么重要，又有多么复杂。计算机科学专业的道路十分艰巨。难怪有些计算科学大师感慨：“计算科学是年轻人的学科。”计算机导论是计算机专业学生必修的一门专业基础课程。也是计算机专业的入门课程。课程内容主要介绍计算机领域相关基础知识。

1. 认识、 体会

大学里的学生都是来自全国各地，由于经济和教育水平的差距，有些学生在中、小学的时候就已经接受系统的计算机基础教育，甚至有些学生还参加过全国计算机大赛，这些学生具备了一定的计算机知识和操作能力。然而有些学生尤其是来自偏远地区的，由于当地的教育设施有限，这些学生对计算机一些基本知识不是很了解。针对这种情况，如果相同的内容按照统一进度来教学的话。而基础较差的学生有可能会跟不上进度，进而影响整个教学进度，以及对这门课程的兴趣。由于这门课程是对计算机专业各个领域的一个总结性概括，其原理性、知识性比较强，讲起来比较枯燥、乏味，会影响学生对本课程的兴趣，从而影响整个教学质量。

但最重要的不是能在计算机导论中学习记住多少，而是学会孙老师提倡的计算科学方法论与知识体系。或许我们中的一些人工作不是计算机相关，但计算科学的知识体系是我们一生中最宝贵的财富。因为科学的方法论指导即人们正确认识世界改造世界的根本方法，是客观事物本质和规律的一般认识原则, 其可以划分为理论层次的科学方法论和经验的科学方法论。

科学的方法论主要包括以下三点：实验——数学方法、科学归纳方法、直观——演绎方法。实验——数学方法即通过长期的实践，用数学语言表述事物的状态、关系和过程，并加以推导、演算和分析，以形成对问题的解释、判断和预言的方法。科学归纳法，在科学研究中运用归纳方法提出和建立假说，在实验基础上抽象和概括事物之间关系的一种科研方法。至于直观——演绎方法，直观是指通过对客观事物的直接接触而获得感性认识的一种方式。中国按其不同涵义分别译为直观和直觉。在笛卡尔认为，哲学理论要具有可靠性，从少数一些简单明了的普遍原理出发，通过层层演绎，最后获得关于其它一切具体事物的结论。所以确定普遍原理的可靠性要用直观方法的确定性来保证。这样一来，一向默默无闻的直观就变得异乎寻常地重要起来。而谓演绎法或称演绎推理（Deductive reasoning）是指人们以一定的反映客观规律的理论认识为依据，从服从该认识的已知部分推知事物的未知部分思维方法。是由一般到个别的认识方法。演绎法是认识“隐性”知识的方法。

之所以我在这提出人类史上最著名的三个方法论，就是为了论证《计算科学导论》对于方法论的指导作用。我们常常做的是，遇见情景，提出问题，找出思路，给出结果，不断试错，得到正解，总结方法，建立数学模型。实验——数学方法强调的是给出思路，建立数学模型。科学归纳方法着重于在正解的情形下，通过相似或相反的方法，归纳总结出核心关键的步骤。至于直观——演绎方法，用直观方法的确定性，来验证结果的唯一性，从而可以演绎或反推核心的方法。举个例子，就像某个黑匣子，输入数字，得到数字，不断重复这个行为，摸透行为机制运算原理，得到最终的核心算法。或许用此可以模拟人脑关于美的识别算法。

《计算科学导论》不止想教会我们方法论，更想培养我们的科学思想。方法论或许会过时，失去效力。但只要有科学的思想体系，方法论就可以不断推陈出新，永不过时。典型的科学思想有：数学科学中的极限思想、自然科学中的互补思想、生命科学中的进化思想、社会科学中的和谐思想、思维科学中的系统思想、哲学科学中的转化思想等。计算学科中的科学思想，也许叫做计算思想？核心是关于实践行为的效率分配。宛如计算机中的并行运算，串联运算。其指导我们学会分配时间，高效地完成好几件事；也指导我们逻辑思考，条理清晰地思考清楚一件事的上下因果，从而抓住关键，一举制胜。不止于此，它教会我们向计算机思考，冷静而又理性。美国科学社会学家默顿认为：普遍性、公有性、无私利性和有条理的怀疑性构成科学的思想气质。我国的蔡德诚教授则把科学精神归纳为“六要素”，即：客观的依据、理性的怀疑、多元的思考、平权的争论、实践的检验、宽容的激励。但无论怎么划分，我们倡导的一般的科学思想概括起来就是：一个科学的认识，一套科学的方法，一个科学的程序。一个人只要善于使用这把利器，无论有多少困难险阻，都足以迈步向前。

1. 思考

我们组的课题演讲报告是商品推荐。其最主要也是最关键的协同过滤算法。协同过滤又可分为评比（rating）或者群体过滤（social filtering）协同过滤以其出色的速度和健壮性，在全球互联网领域炙手可热。最初的协同过滤系统的创建者已经不可考据，只留下创立的集体名，或许这就是日后日渐复杂的发明创造由集体产出而非个人单打独斗的一个剪影吧。最早的协同过滤系统有2个（Tapestry（1992） 、 GroupLens（1994））

Tapestry（1992）

这是最早应用协同过滤系统的设计，主要是解决Xerox公司在Palo Alto的研究中心资讯过载的问题。 其运作机制大致如下：

个人决定自己的感兴趣的邮件类型；个人旋即随机发出一项资讯需求，可预测的结果是会收到非常多相关的文件；从这些文件中个人选出至少三笔资料是其认为有用、会想要看的；系统便将之记录起来成为个人邮件系统内的过滤器，从此以后经过过滤的文件会最先送达信箱；以上是协同过滤最早的应用，接下来的里程碑为GroupLens。

GroupLens（1994）

这个系统主要是应用在新闻的筛选上，帮助新闻的阅听者过滤其感兴趣的新闻内容，阅听者看过内容后给一个评比的分数，系统会将分数记录起来以备未来参考之用，假设前提是阅听者以前感兴趣的东西在未来也会有兴趣阅听，若阅听者不愿揭露自己的身分也可以匿名进行评分。 和Tapestry不同之处有两点，首先，Tapestry专指一个点（如一个网站内、一个系统内）的过滤机制；GroupLens则是跨点跨系统的新闻过滤机制。再来，Tapestry不会将同一笔资料的评比总和起来；GroupLens会将同一笔资料从不同使用者得到的评比加总。

GroupLens具有以下特点：1.开放性：所有的新闻阅听者皆可使用。若有不同的评分机制也适用于GroupLens。2.方便性：评分结果容易诠释。3.规模性4.隐密性：如果使用者不想让别人知道他是谁，别人就不会知道。在GroupLens之后还有性质相近的MovieLens，电影推荐系统；Ringo，音乐推荐系统；Video Recommender，影音推荐系统；以及Jster，笑话推荐系统等等。乃至于YouTube、aNobii皆是相似性值得网络推荐平台。不同的是经过时间推移，网络越来越发达，使用者越来越多，系统也发展得越来越严密。

与之相类似亦或是发展出来的是电子商务的推荐系统。最著名的电子商务推荐系统应属亚马逊网络书店，顾客选择一本自己感兴趣的书籍，马上会在底下看到一行“Customer Who Bought This Item Also Bought”，亚马逊是在“对同样一本书有兴趣的读者们兴趣在某种程度上相近”的假设前提下提供这样的推荐，此举也成为亚马逊网络书店为人所津津乐道的一项服务，各网络书店也跟进做这样的推荐服务如台湾的博客来网络书店。 另外一个著名的例子是Facebook的广告，系统根据个人资料、周遭朋友感兴趣的广告等等对个人提供广告推销，也是一项协同过滤重要的里程碑，和前二者Tapestry、GroupLens不同的是在这里虽然商业气息浓厚同时还是带给使用者很大的方便。 以上为三项协同过滤发展上重要的里程碑，从早期单一系统内的邮件、文件过滤，到跨系统的新闻、电影、音乐过滤，乃至于横行互联网的电子商务，虽然目的不太相同，但带给使用者的方便是大家都不能否定的。

协同过滤是利用集体智慧的一个典型方法。基于用户(user-based)的协同过滤主要考虑的是用户和用户之间的相似度，只要找出相似用户喜欢的物品，并预测目标用户对对应物品的评分，就可以找到评分最高的若干个物品推荐给用户。而基于项目(item-based)的协同过滤和基于用户的协同过滤类似，只不过这时我们转向找到物品和物品之间的相似度，只有找到了目标用户对某些物品的评分，那么我们就可以对相似度高的类似物品进行预测，将评分最高的若干个相似物品推荐给用户。

但这协同过滤算法也存在不少的缺点。由于是基于用户对产品的评分，所以对新用户进行推荐或者是对用户推荐新产品，精度不高；随着用户量的增多，计算量呈线性增加，影响系统的性能。并且受到信息获取技术的约束，例如自动提取多媒体数据(图形、视频流、声音流等)的内容特征具有技术上的困难；如果两个不同的产品恰好用相同的特征词表示，这两个产品就无法区分；如果一个系统只推荐与用户的配置文件高度相关的产品，那么推荐的只能是与用户之前购买过的产品非常相似的产品，无法保证推荐的多样性。缺点也很明显，主要是技术上的问题。如果推荐做不到精准，做不到高用户体验，那么其实对于用户而言，反而会觉得不胜其烦，会导致用户因此流失。这也不难理解有些用户喜欢简约设计的产品，干净整洁。

真正想解决这些问题，就需要一个个对症下药。针对新用户进行推荐，广泛地推荐日用品或性价比高的商品，收集足够的数据后，再渐渐的调整商品推荐的主流。对用户推荐新产品，讲究一个不经意。利用用户的好奇心，将感兴趣的用户吸引进来。再根据他们的着重喜好更改商品用途或描述。至于处理计算量，要么增加处理器的数量或性能，要么减少数据的冗杂量，增加权重，合并，删除相似或一致数据，要么设置人物性格模板，根据用户的选择套用相似的固定模板。而为了保证推荐的多样性，就不能一个劲地推荐已经推荐过的商品，提供多种多样，甚至毫不相关的商品，保证商品的多样性与随机性，充足开发出用户的隐性需求，以做到高用户体验。至于更高一筹的简约设计的用户界面，需要对用户的心理了如指掌。万一用不好，就是口碑流失的后果，所以大多数商家求稳，不求有功但求无过，使用花里胡哨，词条繁多但实用且有用的描述。要解决用户关于不同风格的产品喜好问题，可以适度隐藏描述词条，调整显示界面，一切就完美了。还可以一键选择简洁版，详细版，就更贴心了。

1. 总结

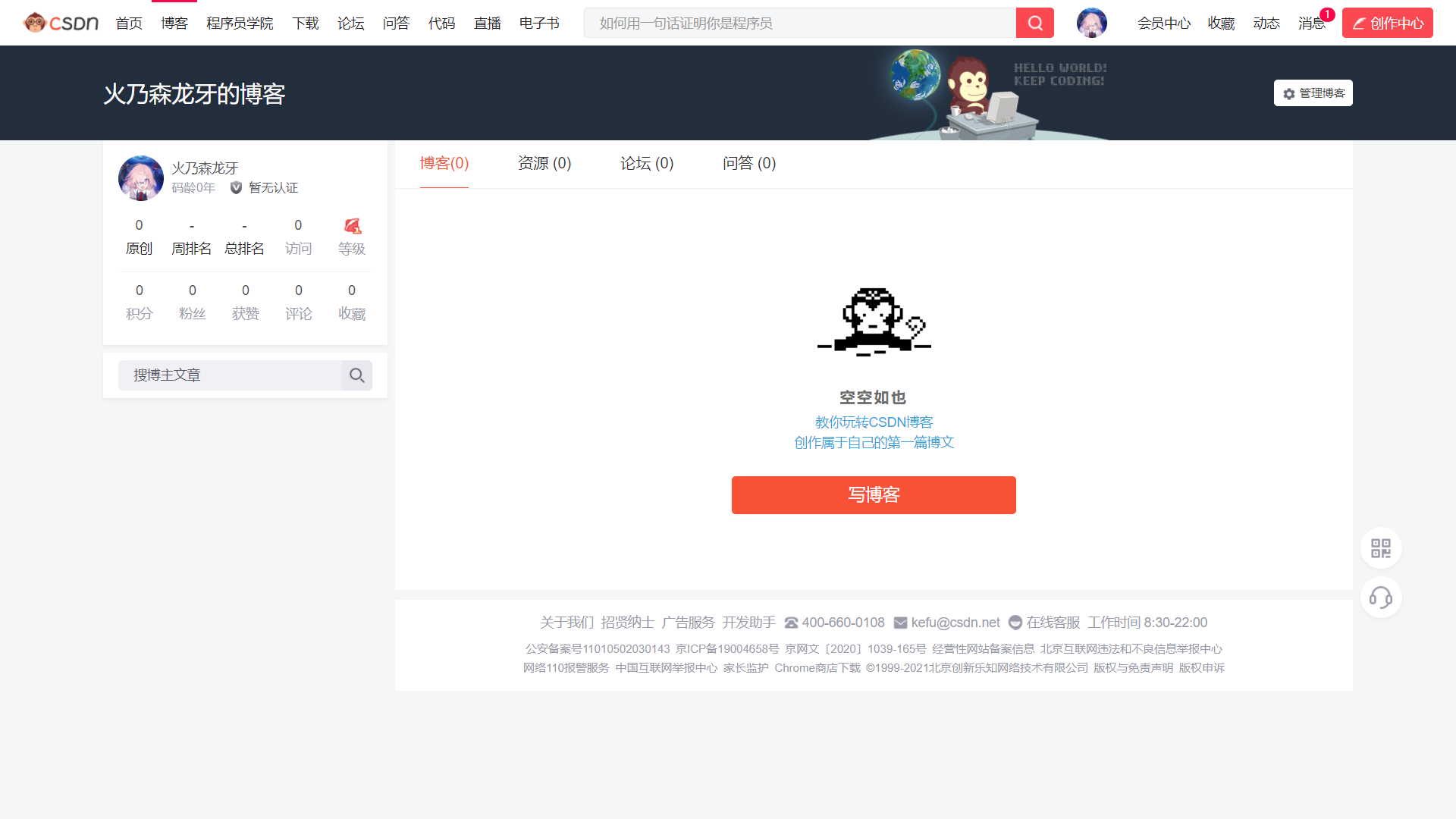
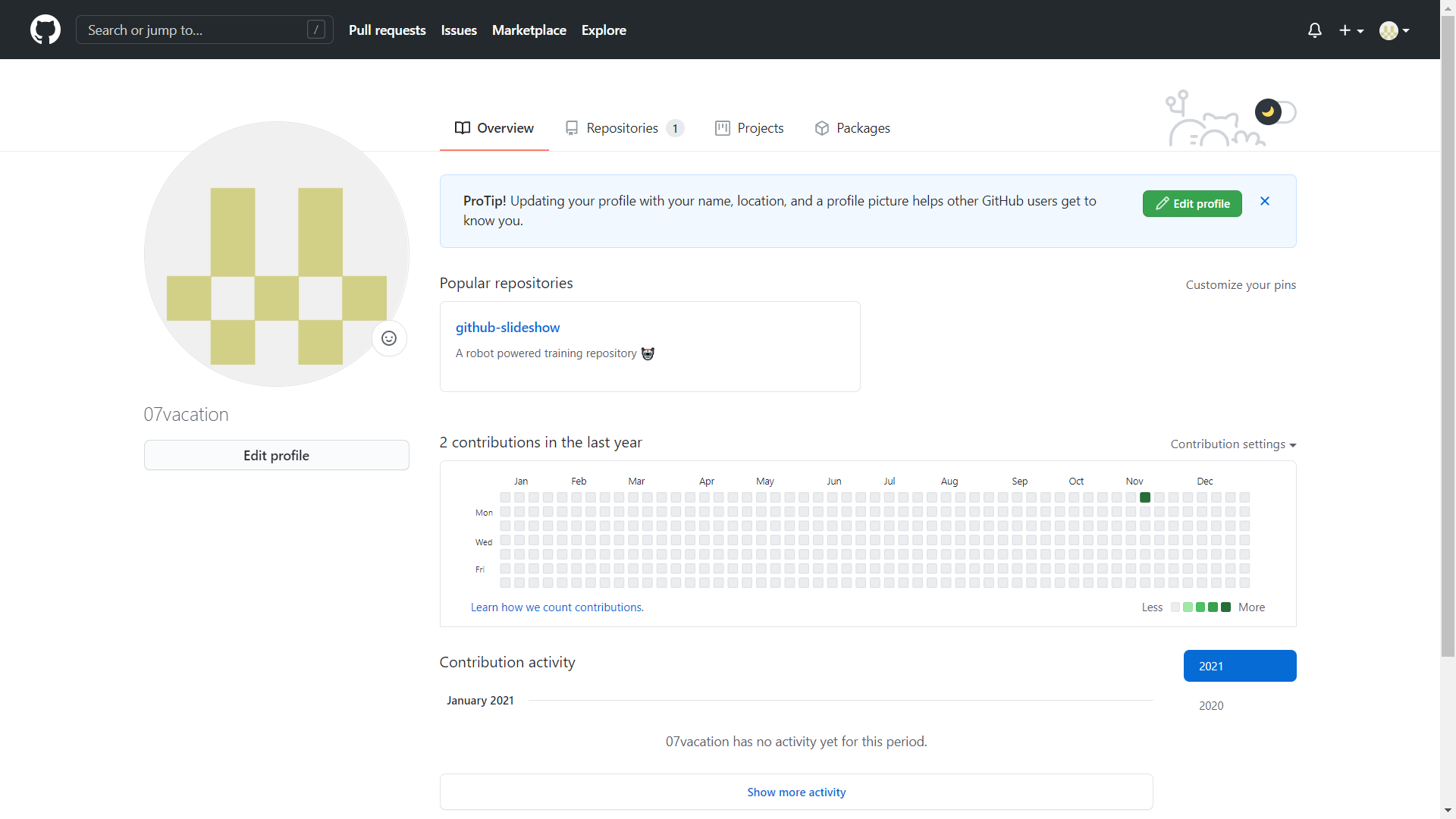
只有认识世界，才能改造世界。非常感谢《计算科学导论》这本书，它让我能站在一个更加专业的角度去重新认识这个世界。其所提及的科学的思想哲学可以用来解决费马定理，丢番图方程，无理数，超越数等无数问题。科学计算方法是获得理解，主要是通过分析在计算机上实现的数学模型。毕达哥拉斯学派所言： "all things are numbers"，万物皆数。所以万物皆可以用计算来模拟。而我们所需要的是，科学的思想哲学。或许这就是《计算科学导论》所希冀的，养成逻辑思维，培养理性，之后，万物皆有解。

1. 参考文献

哈佛大学公开课：计算机科学导论 、河北工业大学计算机科学与软件学院 某人计算机科学导论论文 、lili安 CSDN博客 、 知网 基于时序行为的协同过滤推荐算法 、 知网 基于协同过滤的推荐算法研究 、 知网 协同过滤推荐技术综述 、 科普中国·科学百科 、 《计算机科学导论》 、 MBA智库·百科 、刘建平Pinard 博客园、万物皆数

1. 附录

Github 07vacation https://github.com/07vacation

 <https://blog.csdn.net/vacation07?spm=1000.2115.3001.5343> 账号vacation07



<http://muchong.com/bbs/index.php> 账号07vacation