



中央财经大学

Central University of Finance and Economics

个人总结与规划

学年学期： 2019-2020 第二学期

课程名称： 数据挖掘

课程代码： 3320062

任课教师： 马景义

姓 名： 徐菡

学 号： 2017310719

班 级： 金融实验班 17

写在前面：这篇文章前半部分描述感受较多，可能与本作业要求不太相关，老师如果时间不够可以略过。重点部分已用黑子标出，直接阅读重点请点击此链接：[本文重点](#)

总结与规划

作为一名中财首批金融科技专业的学生，在大二专业分流刚刚进入这个班级时对未来的工作发展是充满希望和信心的，因为那时像“互联网金融”和“金融科技”都是非常热门的词汇，一方面，我经常在公众号和知乎上看到关于大数据技术在金融领域的应用前景，甚至看到过“金融+科技”是最有前景的专业这种论断；另一方面，我每每在数据库中浏览金融有关论文以及在行研库中查找报告时，都会看到很多将“大数据”、“数据挖掘”、“机器学习”与“金融”、“投资”结合的一些研究，如“大数据金融风险控制”、“利用机器学习方法识别财务欺诈风险”、“运用数据挖掘技术开发量化投资策略”等，给人一种金融科技已发展得十分成熟的错觉。

然而，“纸上得来中觉浅，绝知此事要躬行”，在两年的时间内，我渐渐看清了现实，对金融科技这个专业以及自身的发展有了更清楚的认识。金融科技应该最早是在美国提出的，主要指“由大数据、区块链、云计算、人工智能等新兴前沿技术带动，对金融市场以及金融服务业务供给产生重大影响的新兴业务模式、新技术应用、新产品服务等”（来源于百度百科）。但事实却是，相比于欧美日这些发达国家，

特别是美国，中国的金融体系是非常不发达的，市场也是不健全的，中国的金融科技目前主要是在网络支付、网络银行、消费金融等方面（网络借贷即 P2P 在前些年出现一些爆雷事件后现在监管趋严），而众所周知的阿里、腾讯这些涉及第三方支付的企业作为金融科技的大头其本质是互联网公司而非金融机构，招聘也更多地集中在计算机、软件工程、统计、数学等专业而非金融或者是所谓的金融科技专业，工作内容也主要是。而反观目前国内的金融机构，包括银行、证券、保险、信托等，就我从各种渠道上获取的实习生和员工招聘信息上来看，对象主要是金融、经济等专业，而描述的工作内容也都限于传统金融行业，与数据、编程这些基本无关。我从金融从业者口中了解到，国内的金融目前还主要是合规金融，不论是一级市场上的直接融资还是二级市场上的投研分析、风险管理等工作，都主要围绕财务和法律，内容多是撰写各种报告，虽然国内也有一些量化基金，但与传统金融相比体量非常小，而且从业者大多是计算机或者数学的博士。

大二下学期我有幸到加州大学伯克利分校进行了为期五个月的交换，我从在工程运筹学院学习金融科技专业的一名学姐那里得知，她们虽然学习专业叫做金融科技，但是重点其实是数据而非金融，除了金融科技专业，还有数据科学专业、运筹学专业等许多专业的课程设置都比较类似，并且学生最后的工作去向也都是湾区的数据岗。这其中比较重要的原因可能是中国学生在美国几乎不太可能找到金融机构的工作（因为语言、肤色、背景、性格等多种原因），而数据

岗则对中国学生较为友好，门槛相对来说没有那么高，只要学历和经历过关、再稍加刷题冲过笔面试就有可能拿到一个像样的 offer。那些在美国学习金融工程、金融科技专业的学生，除了极少数就读于哈佛、MIT、斯坦福等学校的十分拔尖的人才以外，其实大部分都不太可能在华尔街的基金公司找到工作，他们一部分人毕业后即回国，从事着的与数据不相关的纯金融工作，而另一部分则在美国成为了码农或者从事数据分析之类的岗位。

而国内又何尝不是如此呢？金融行业人才济济，进入门槛之高堪称各行业之最，并且在未来几年里这个门槛只可能升而不会降。多少金融学子嘴里喊着要去基金公司当量化研究员，但是却未曾想过每年招人为个位数的基金公司，怎么回放着清北计算机博士和已有多年工作经验的业内人士不要而去选择你一个刚毕业的技术小白。金融相关专业的应届生们绝大部分最后都去了招人最多的银行开始当柜员，虽然这也不失为一种稳妥而体面的工作，但未免与我们当初刚进大学时的想象差距过大，而我们在大学时学习的专业知识、编程技术确无用武之地。

从目前的情况来看，在排除量化投资研究员这个选项之后（因为短期内实现可能性过低），我需要在金融和数据二者中做一个选择，而**我最终选择了数据**，主要原因如下：一方面，传统金融行业对财务和法律知识要求较高，我在相关方面是比较匮乏的；而另一方面，很

多金融工作岗位都需要和人打交道，对自身资源的要求加高，不管是投行还是银行、证券的行研，都需要面对客户，与我的性格（比较内向）和家庭条件（不属于自带资源的那种）不符。而数据则相对来说则主要是看重技术和能力，即通过自己的努力是可以找到一份不错的工作甚至达到一定的成就的，并且我在学习了数据挖掘、大数据等课程后发现自己对这方面也有不错的天赋和较高的兴趣。

在确立了以后从事数据相关工作的目标后，我开始梳理相关职业所需的技能以及自身的能力，以确定努力方向，弥补不足。我了解到数据相关的职业又可以分为偏业务和偏技术的，偏业务的有数据分析，数据运营，商业分析，偏技术则是数据挖掘、数据建模算法类的等等，对数学的要求较高。我很快就确定了偏技术的方向，即数据挖掘算法工程师类，这类岗位对相关知识技能的要求非常高，即进入门槛相对较高，这对我来说意味着更大的挑战以及更多的机会。要想成为一名合格的数据挖掘算法工程师，需要掌握的知识和技能大致包括：编程语言（Python，R，Java 等），大数据处理框架（Hadoop 等），操作系统（Linux），数据库（SQL），统计知识，数据结构与算法，机器学习算法等。

然后是对自己的能力进梳理：

(1) 我在本科阶段学习了 Python, C++, Java, R 这些编程语言、数据库以及数据结构与算法, 但大都只是浅尝辄止, 许多方法都未进行深入学习, 并且在熟练程度也有待加强;

(2) 在数学和统计方面, 本科阶段也学过一些基本的课程, 但是这些对于工作来说肯定是不够的, 相关的数学和统计知识还得在需要的时候进行学习或重温;

(3) 在操作系统和大数据处理框架这些比较底层的接近于计算机知识技能方面我是比较匮乏的, 这些都是工具, 在实践的过程中学会怎样应用即可;

(4) 最后就是最重要的机器学习算法了。在数据挖掘这门课中, 我第一次跳脱了万年“调包侠”, 学会了自己写算法。在揭开“黑匣子”的面纱后, 我发现自己写算法也不是一件多么恐怖的事情, 从最简单的线性回归开始, 到选择变量的最优子集回归、主成分回归、偏最小二乘回归、Lasso 和岭回归等统计学习算法, 再到分类问题中的分类树、Adaboost、随机森林, 还有主成分分析、聚类分析等, 其实都是有迹可循的, 只要理解了算法的原理和推导过程, 掌握了基本的矩阵运算和相关的 Python 函数, 就可以自己对算法进行实现。虽然 sklearn 库中的源代码都很长, 但是许多都是注释和帮助类的, 以及一些不常用到的参数, 问题都是由简至繁的, 只要掌握了基本的原理和推导, 搭建好了基本的框架, 后面就可以逐步对代码进行丰富, 以完善细节或实现更多的功能。(以上是我学习数据挖掘这门课以及完成几次作业后的收获和感想)。虽然我现在能推导出来并用 Python 进

行实现的算法还是比较简单的算法,但是可以通过之后的学习和不断的实践对这一方面进行加强,可以尝试自己实现更复杂的算法,如 svm 等,并用真实数据进行实验。

有了对自己能力更清晰的认知,我便可以朝着数据挖掘算法工程师的目标不断努力了。首先,我认为本科阶段的这些知识肯定是不够用的,需要在研究生阶段对计算机和数据挖掘算法这些方面进行进一步加强。硕士学位虽然重要,但对我来说更重要的是我更充裕的时间来丰富自己的技能和相关经历(如参加一些数据科学的竞赛),为之后真正进入工作做更充分的准备。不过我清楚的是,即使做完这些仍然是不够的,因为真实的工作与理论研究不同,不是自己写出一个看似漂亮的算法就是成功,因为算法是要用来解决实际问题的,是要经受得住实际数据的考验的,也就是说怎样在实际的场景中选择合适的算法、合适的参数,根据业务场景与数据特点对算法进行改造或者调整以得到最优的效果,这就要求我们不仅要理解算法的原理和过程,还要对实际业务和真实数据要有清楚的认识和理解。这些都是需要在之后的学习、比赛和实习中不断去探索的。学无止境,尤其是在日新月异的今天,工作的环境和内容会不断更新换代,新的算法也会层出不穷,只有具备扎实的基础和丰富的经验,我才能离一名成功的数据挖掘算法工程师更近一步。

这篇文章花了较多篇幅在描述对现实的感受和对未来的规划上，也算是在大三结束，所有专业课都学完之际与自己进行对话。需要感谢所有给我传道授业解惑的老师，尤其是领我进入数据科学大门的马景义老师。如果当初没有选这门课，我可能会在迷茫中度过大学生活，而不是如此果敢的选择另辟蹊径。“山重水复疑无路，柳暗花明又一村”，虽然经历过迷茫，但是幸运的是已找到方向。我相信我也可以和统数学院的同学一样成为一名数据挖掘算法工程师。