**数据科学与大数据专业 专业导论课程报告**

**————大数据时代的思考**

班级：大数据分析2101

学号：202103151422

姓名：温家伟

**一、背景**

大数据时代已经悄然到来，如何应对大数据时代带来的挑战与机遇，是我们当代大学生特别是我们计算机类专业的大学生的一个必须面对的严峻课题。大数据时代是我们的一个黄金时代在讲座中秦永彬博士由一个电视剧《大太监》中情节来深入浅出的简单介绍了“大数据”的基本概念，并由“塔吉特”与“犯罪预测”两个案例让我们深切的体会到了“大数据”的对现今这样一个信息时代的不可替代的巨大作用。

**二、主要内容**

在前几年本世纪初的时候，世界都称本世纪为“信息世纪”。确实在计算机技术与互联网技术的飞速发展过后，我们面临了一个每天都可以“信息爆炸”的时代。打开电视，打开电脑，甚至是在街上打开手机、PDA、平板电脑等等，你都可以接收到来自互联网从世界各地上传的各类信息:数据、视频、图片、音频……这样各类大量的数据累积之后达到了引起量变的临界值，数据本身有潜在的价值，但价值比较分散;数据高速产生，需高速处理。大数据意味着包括交易和交互数据集在内的所有数据集,其规模或复杂程度超出了常用技术按照合理的成本和时限捕捉、管理及处理这些数据集的能力。遂有了“大数据”技术的应运而生。

现在，当数据的积累量足够大的时候到来时，量变引起了质变。“大数据”通过对海量数据有针对性的分析，赋予了互联网“智商”，这使得互联网的作用，从简单的数据交流和信息传递，上升到基于海量数据的分析，一句话“他开始思考了”。简言之，大数据就是将碎片化的海量数据在一定的时间内完成筛选、分析,并整理成为有用的资讯,帮助用户完成决策。借助大数据企业的决策者可以迅速感知市场需求变化,从而促使他们做出对企业更有利的决策，使得这些企业拥有更强的创新力和竞争力。这是继云计算、物联网之后IT产业又一次颠覆性的技术变革，对国家治理模式、对企业的决策、组织和业务流程、对个人生活方式都将产生巨大的影响。后工业社会时代，随着新兴技术的发展与互联网底层技术的革新，数据正在呈指数级增长，所有数据的产生形式，都是数字化。如何收集、管理和分析海量数据对于企业从事的一切商业活动都显得尤为重要。

大数据时代是信息化社会发展必然趋势，我们只有紧紧跟随时代发展的潮流，在技术上、制度上、价值观念上做出迅速调整并牢牢跟进，才能在接下来新一轮的竞争中摆脱受制于人的弱势境地，才能把握发展的方向。

首先，“大数据”究竟是什么?它有什么用?这是当下每个人初接触“大数据”都会有的疑问，而这些疑问在秦博士的讲座中我们都了解到了。“大数据”的“大”不仅是单纯指数量上的“大”，而是在诸多方面上阐释了“大”的含义，是体现在数据信息是海量信息，且在动态变化和不断增长之上。同时“大数据”在:速度(Velocity)、多样性(Variety) 、价值密度(Value)、体量(Volume）这四方面(4v）都有体现。其实“大数据”归根结底还是数据，其是一种泛化的数据描述形式，有别于以往对于数据信息的表达，大数据更多地倾向于表达网络用户信息、新闻信息、银行数据信息、社交媒体上的数据信息、购物网站上的用户数据信息、规模超过TB级的数据信息等。

在如此快速的到来的大数据革命时代，我们还有很多知识需要学习，许多思维需要转变，许多技术需要研究。职业规划中，也需充分考虑到大数据对于自身职业的未来发展所带来的机遇和挑战。当我们掌握大量数据，需要考虑有多少数字化的数据，又有哪些可以通过大数据的分析处理而带来有价值的用途?在大数据时代制胜的良药也许是创新的点子,也许可以利用外部的数据，通过多维化、多层面的分析给我们日后创业带来价值。借力，顺势，合作共赢。把自己的心门打开，像海绵般吸取积极、正能量的东西。了解了“大数据”的“大”之后我们也该了解它所具有的巨大价值。就目前来说“大数据”的来源主要还是互联网，来自互联网上的大多数不被重视信息都是具有巨大开发价值的，其具有巨大的商业价值，我们所缺少的只是一些数据分析等手段。例如:在如今，网购已经成为了一种风潮，网上也涌现了以淘宝、京东、亚马逊等一系列的购物网站。而在这些网站之中，顾客的浏览记录，购买记录等等都是一些巨大商业价值的信息。借鉴“塔吉特”的先例，我们可以利用“大数据”技术收集分析，就可预测需求、供给和顾客习惯等，做到精准采购、精准投放，达到利益放大的效果。从全球范围来看，很多人都把2012年看作是大数据时代的元年。[1]在这一年里，很多行业在大数据方面的管理、规划和应用已经觉醒。电商、金融、电信等行业数据有着长期的数据积累。事实上，很多互联网公司，例如亚马逊、Google、腾讯，更愿意将自己定位为数据企业。因为信息时代，数据成为经营决策的强有力依据，给企业带来了发展和引领行业的机遇。银行也同样拥有丰富的数据矿藏，不仅存储处理了大量结构化的账务数据，而且随着银行渠道快速渗透到社交网络、移动端等媒介，海量的非结构化数据也在等待被收集和分析。未来的金融业将更多地受到科技创新力的驱动，也越来越倾向于零售营销:对于金融业来说，大数据意味着巨大的商机，可强化客户体验，提高客户忠诚度。大数据技术的发展带来企业经营决策模式的转变，驱动着行业变革，衍生出新的商机和发展契机。驾驭大数据的能力已被证实为领军企业的核心竞争力，这种能力能够帮助企业打破数据边界，绘制企业运营全景视图，做出最优的商业决策和发展战略。金融行业在大数据浪潮中，要以大数据平台建设为基础，夯实大数据的收集、存储、处理能力;重点推进大数据人才的梯队建设，打造专业、高效、灵活的大数据分析团队;不断提升企业智商，挖掘海量数据的商业价值，从而在数据新浪潮的变革中拔得头筹，赢得先机。

在此，我想再重新审视一下关于大数据的历史沿革和现实意义。一、大数据的历史沿革及现实意义。首先，最早提出“大数据”时代到来的是全球知名咨询公司麦肯锡，麦肯锡称:“数据，已经渗透到当今每一个行业和业务职能领域，成为重要的生产因素。人们对于海量数据的挖掘和运用,预示着新一波生产率增长和消费者盈余浪潮的到来。”[2]“大数据”在物理学、生物学、环境生态学等领域以及军事、金融、通讯等行业存在已有时日，却因为近年来互联网和信息行业的发展而引起人们关注。大数据作为云计算、物联网之后又IT行业又一大颠覆性的技术革命。云计算主要为数据资产提供了保管、访问的场所和渠道，而数据才是真正有价值的资产。企业内部的经营交易信息、物联网世界中的商品物流信息，互联网世界中的人与人交互信息、位置信息等，其数量将远远超越现有企业IT架构和基础 设施的承载能力,实时性要求也将大大超越现有的计算能力。如何盘活这些数据资产，使其为国家治理、企业决策乃至个人生活服务，是大数据的核心议题，也是云计算内在的灵魂和必然的升级方向。其次，进入2012年，大数据(bigdata)一词越来越多地被提及，人们用它来描述和定义信息爆炸时代产生的海量数据，并命名与之相关的技术发展与创新。它已经上过《纽约时报》《华尔街日报》的专栏封面，进入美国白宫官网的新闻，现身在国内一些互联网主题的讲座沙龙中，甚至被嗅觉灵敏的国金证券、国泰君安、银河证券等写进了投资推荐报告。数据正在迅速膨胀并变大，它决定着企业的未来发展，虽然很多企业可能并没有意识到数据爆炸性增长带来问题的隐患，但是随着时间的推移，人们将越来越多的意识到数据对企业的重要性。正如《纽约时报》2012年2月的一篇专栏中所称，“大数据”时代已经降临，在商业、经济及其他领域中，决策将日益基于数据和分析而作出，而并非基于经验和直觉。[3]哈佛大学社会学教授加里·金说:“这是一场革命，庞大的数据资源使得各个领域开始了量化进程，无论学术界、商界还是政府，所有领域都将开始这种进程。”[4]最后，随着云时代的来临，大数据(Bigdata)也吸引了越来越多的关注。著云台的分析师团队认为，大数据(Bigdata)通常用来形容一个公司创造的大量非结构化和半结构化数据，这些数据在下载到关系型数据库用于分析时会花费过多时间和金钱。大数据分析常和云计算联系到一起,因为实时的大型数据集分析需要像MapReduce一样的框架来向数十、数百或甚至数千的电脑分配工作。“大数据”在互联网行业指的是这样一种现象:互联网公司在日常运营中生成、累积的用户网络行为数据。这些数据的规模是如此庞大，以至于不能用G或T来衡量。大数据到底有多大?一组名为“互联网上一天”的数据告诉我们，一天之中，互联网产生的全部内容可以刻满1.68亿张DVD;发出的邮件有2940亿封之多(相当于美国两年的纸质信件数量);发出的社区帖子达200万个(相当于《时代》杂志770年的文字量);卖出的手机为37.8万台，高于全球每天出生的婴儿数量37.1万……，截止到2012年，数据量已经从TB(1024GB=1TB)级别跃升到PB(1024TB=1PB)、EB(1024PB=1EB)乃至ZB(1024EB=1ZB)级别。国际数据公司(IDC)的研究结果表明，2008年全球产生的数据量为0.49ZB，2009年的数据量为0.8ZB，2010年增长为1.2ZB，2011年的数量更是高达1.82ZB，相当于全球每人产生200GB 以上的数据。而到2012年为止，人类生产的所有印刷材料的数据量是200PB，全人类历史上说过的所有话的数据量大约是5EB。IBM 的研究称，整个人类文明所获得的全部数据中，有90%是过去两年内产生的。而到了2020年，全世界所产生的数据规模将达到今天的44倍。

“大数据”在互联网行业指的是这样一种现象:互联网公司在日常运营中生成、累积的用户网络行为数据。这些数据的规模是如此庞大，以至于不能用G或T来衡量。大数据到底有多大?一组名为“互联网上一天”[5]的数据告诉我们，一天之中，互联网产生的全部内容可以刻满1.68亿张DVD;发出的邮件有2940亿封之多(相当于美国两年的纸质信件数量);发出的社区帖子达200万个(相当于《时代》杂志770年的文字量);卖出的手机为37.8万台，高于全球每天出生的婴儿数量37.1万……，截止到2012年，数据量已经从TB(1024GB=1TB)级别跃升到PB(1024TB=1PB)、EB(1024PB=1EB)乃至ZB(1024EB=1ZB)级别。国际数据公司(IDC)的研究结果表明，2008年全球产生的数据量为0.49ZB，2009年的数据量为0.8ZB，2010年增长为1.2ZB，2011年的数量更是高达1.82ZB，相当于全球每人产生200GB 以上的数据。而到2012年为止，人类生产的所有印刷材料的数据量是200PB，全人类历史上说过的所有话的数据量大约是5EB。IBM 的研究称，整个人类文明所获得的全部数据中，有90%是过去两年内产生的。而到了2020年，全世界所产生的数据规模将达到今天的44倍。大数据已经是信息产业发展的必然趋势，可以说，大数据现在已经开始慢慢渗透入我们的生活，如:现在流行的打车软件、三维立体化社区的建立、某些从事生产销售的行业利用大数据来优化规模和实现利益最大化。而我们很多人对大数据还很陌生，只是被动的适应着大数据给生活带来的改变。大数据时代是以云计算为基础的，所以，要实现大数据，相关的很多的硬件设备都要更新换代，信息处理系统、信息传输系统、信息反馈系统、信息决策系统都将面临新的挑战，相关产业都要重新调整产业结构，在那时，可以夸张的说，信息就是黄金，信息就是石油。大数据时代的到来会解放更多的劳动生产力，势必将会更加加剧生产力过剩的现状，社会两极分化现象会更加明显，掌握不了信息资源，很难再翻身，要防止信息垄断带来的可怕局面。大数据时代的到来会使人们的生活节奏急速加快，信息的时效性决定了它的流通速率，人们的生活节奏要跟上信息流通的速率，就不得不加快自己的节奏，人们会越来越忙，到那时，就像现在的日本，可能想找个人听你说说话，真的是一件很难的事。第二，关于数据管理的看法。大数据时代，数据管理是一件很重要的工作，如何才能避免自己的数据被非法窃取、丢失和被盗?我的看法是，人防、技防、物防一体化。人防，即我们要从思想上牢固树立信息安全防范的意识，不主动泄露信息，要管理好自己身边的信息设备;技防，就是要运用软件来管理和处理数据，经常检查更新数据库，定时查杀电脑病毒，确保电脑状况安全;物防，就是重要的数据一定要备份保留，而且应当做到备份与原始文件是物理隔离，无关的信息应当及时删除，减轻硬盘的压力。三、怎么保护自己的隐私。隐私，顾名思义，就是不愿意让别人看到的东西，所以，在大数据时代，更要管理好自己的隐私，以免对自己和家人造成麻烦和损失。越是隐私的信息，越要远离网络，不要再公开的社交等信息，免得被非法人士采用和窃取。建议还是用纸质的日记代替电脑日记，避免信息传播范围太大，管理好自己的日记本。研发一种新的硬件连接器，总是以随机码来保护自己真实IP地址，提高网络安全的可靠性，加强对联网信息的管理和保护网络储存和展示个人图片、资料。

**三、参考文献**

[1]. 《人民日报》，2012年2月。

[2]. 麦肯锡,《Big data:The next frontier for innovation,competition and productivity》, 2011年5月。

[3].《纽约时报》，2012年2月。

[4]. 哈佛大学社会学教授加里·金,2004年8月。

[5].《互联网上的一天》，2012年03月06日。