新生研讨课平时报告1

基于时空大数据的智能决策

——听时宏伟教授《Intelligence Decision Based on Big Data》课的理解感受

什么是大数据

对于大一的新生来说，大数据是一个熟悉又陌生的概念。熟悉在于，日常生活中，大数据已经成为一个科技热词，我们对大数据也有一定的了解，知道智能推送什么的就是靠大数据分析来实现的。陌生在于，作为计算机系的学生，肯定不能这样浅尝辄止，得知道其如何实现的，但是，C语言课上一个小小的大数字计算就让我们焦头烂额了，又何谈知悉大数据的工作方式。所以，我去百度上找了找大数据的相关资料，如下。

对于“大数据”（Big data）研究机构Gartner给出了这样的定义。“大数据”是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力来适应海量、高增长率和多样化的信息资产。

麦肯锡全球研究所给出的定义是：一种规模大到在获取、存储、管理、分析方面大大超出了传统数据库软件工具能力范围的数据集合，具有海量的数据规模、快速的数据流转、多样的数据类型和价值密度低四大特征。

大数据技术的战略意义不在于掌握庞大的数据信息，而在于对这些含有意义的数据进行专业化处理。换而言之，如果把大数据比作一种产业，那么这种产业实现盈利的关键，在于提高对数据的“加工能力”，通过“加工”实现数据的“增值”。

从技术上看，大数据与云计算的关系就像一枚硬币的正反面一样密不可分。大数据必然无法用单台的计算机进行处理，必须采用分布式架构。它的特色在于对海量数据进行分布式数据挖掘。但它必须依托云计算的分布式处理、分布式数据库和云存储、虚拟化技术。

随着云时代的来临，大数据（Big data）也吸引了越来越多的关注。分析师团队认为，大数据（Big data）通常用来形容一个公司创造的大量非结构化数据和半结构化数据，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。

大数据需要特殊的技术，以有效地处理大量的容忍经过时间内的数据。适用于大数据的技术，包括大规模并行处理（MPP）数据库、数据挖掘、分布式文件系统、分布式数据库、云计算平台、互联网和可扩展的存储系统。

维基百科：指的是传统数据处理应用软件不足以处理它们的大或复杂的数据集的术语 百度百科：指无法在一定时间范围内用常规软件工具进行捕捉、管理和处理的数据集 合，是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、 高增长率和多样化的信息资产 互动百科：指需要新处理模式才能具有更强的决策力、洞察力和流程优化能力的海量、 高增长率和多样化的信息资产 舍恩伯格《大数据时代》定义 ：指不用随机分析法（抽样调查）这样的捷径，而 采用所有数据进行分析处理 Gartner ： “大数据”是需要新处理模式才能具有更强的决策力、洞察发现力和流程 优化能力的海量、高增长率和多样化的信息资产

由资料看来，大数据可以直接望文生义，就是大量的数据。但是，还是不太清楚。时讲授讲解了其中原因。理解大数据需要上升到认识论的高度。

世上万物都可数据化，形成一个与现实世界相对应-关联的数据世界（数字孪 生、元宇宙…），推动人类社会的不断发展进步。从宏观来看，的确，世间万物都可抽象为数据，人是众多生命指标和基因数据的结果，每个人基因不同，故人人不同。倘若拥有足够算力和数据，未来完全是可推演的，未来似乎已经是定局。这是大数据的应用之一——预测。

在科幻作品灵笼中就有超级计算机推演未来的情节，并阐述了通过极致算力计算未来的可能性。

但设想是美好的，现实却是骨感的。时空大数据动态，多元，泛在，海量的特点让其难以被发现，管理，建模。GNSS，时空基准，大地测量，重磁测量，遥感影像……多种异构时空数据类型已经让人感到大数据的“桀骜不驯”，对大数据的应用仍然任重道远

基于大数据的智能决策

决策是什么？——决策存在于人类一切实践活动当中：小到一台机器, 大到一个国家的治理， 都离不开决策；例如, 工业领域的操作优化与资源分配、 商业领域的个性 化推荐与供应商选择、交通领域的车流控制与路径导航、 医疗领域的疾病 诊断与治疗策略等等，均属决策范畴。

随着社会发展, 来自各领域行业的决策活动在频度、广度及复杂性上有本质性提高，决策问题的不确定性程度随着决策环境的开放程度以及决策资源的变化程度而越来越大，越来越复杂。

传统的基于人工经验、直觉及少量数据分析的决策方式已远不能满足日益个性化、 多样化、复杂化的决策需求；在当前信息开放与动态交互环境下,机遇与挑战并存，如何把握机遇，需要出色的决策能力。

大数据可望为人们提供全面的、精准的、实时的态势洞察和决策指导。

杨院士等指出：大数据的价值在于其“决策有用性”,通过分析、挖掘来 发现其中蕴藏的知识, 可以为各种实际应用提供其他资源难以提供的决策支持。

美国应用信息经济学家 Hubbard 认为“一切皆可量化”,并积极倡导数据化决策；纽约大学 Provost教授等认为数据科学的终极目标就是改善决策，从数据到知识,从知识到决策,是当前大数据智能的计算范式；

研究大数据的意义就是不断提高“从数据到决策的能力”。随着大数据技术的 发展, 人们传统的决策模式与思维方式正在发生着变革，基于大数据的决策方式正逐渐成为决策应用与研究领域的主旋律,大数据决策时代已经到来。

大数据的应用场景设想

当前，大数据广泛用于商业场景。被广泛宣传的就是智能推送：通过社交媒体、网站访问、呼叫记录以及其他来源收集数据，进而改善客户互动，为客户提供个性化产品，降低客户流失率。

但是，公司其实还利用大数据来预测客户需求。他们对过去和当前产品或服务的关键属性进行分类，并对那些属性和成功商业产品之间的关系进行建模，从而为新产品和服务构建预测模型，把握市场动向，侧重对热门商品的投入。大数据有助于研究人、组织、实体以及流程之间的相互关系，进而基于深度洞察，以全新方式推动创新。在大数据的帮助下，领导层可以有效改善财务和企业计划决策，验证趋势和客户需求，更好地为客户提供新产品和新服务，还可以实施动态定价，从而充分实现收益。简而言之，大数据将打开创新世界的大门，带来无穷的可能性

在工业场景中也可以应用。比如，各种结构化数据（例如设备年份、品牌、型号等信息）以及非结构化数据（包括数以百万计的日志条目、传感器数据、错误消息和引擎温度）中往往深藏着可供预测机械故障的信息，通过分析这些数据，企业可以在事故发生前识别潜在问题，从而更加经济高效地安排维护活动，充分延长零部件和设备的正常运行时间。

网络安全方面应用设想：如今，系统面临的威胁远不止几个心怀不轨的黑客，还有人员配置完善的专家团队。同时，安全形势与合规要求也在不断变化，带来了重重挑战。借助大数据，我们可以通过识别数据模式发现欺诈迹象，汇总海量信息，加速生成监管报告。控制欺诈，保证网络合规性。

小结和一些感想

大数据的好处是巨大的，但弊端也是显而易见的。数据存在哪里，其中的价值只能由专业人员来解读，但未来并不只是数据科学家和程序开发员的天下。如何发掘数据将是未来的重点。发掘数据是一种能力，如果我们大学四年不学好只是，我们将错过这种能力。大学时期，大数据还未发育到盛期，当四年之后，大数据将会是新一轮机遇，大数据时代将引发新一轮信息化投资与建设热潮。

因而，我辈应该踔厉奋发，学好数学，计算机知识，培养创新精神，迎合时代热潮，在大数据应用领域开拓荒原，锐意进取，做出自己的贡献！