

“全国科技工作者日” 徐宗本院士网络科普报告 “数据科学：它的内涵、方法与意义”

人们普遍认为：大数据发展催生了数据科学，而数据科学承载着大数据发展的未来。然而，数据科学到底是什么？它对于科学技术发展、社会进步有什么特别的意义？它有没有独特的内涵与研究方法论？它的发展规律、发展趋势、与其它学科的关系，乃至人才培养规律是什么？澄清和认识这些问题对于准确把握数据科学发展方向、促进以数据为基础的科学技术发展、有着极为重要而现实的意义。

2022年的5月30日是第六个“全国科技工作者日”。今年全国科技工作者日的主题是“创新争先、自立自强”。在这个专属于科技工作者的日子里，中国数学会联合中国工业与应用数学学会和中国运筹学会特别邀请西安交通大学徐宗本院士，为广大科技工作者献上了精彩的网络科普讲座“数据科学：它的内涵、方法与意义”。中国数学会副理事长周爱辉研究员主持了讲座，一起出席的还有中国工业与应用数学学会副理事长王兆军教授、中国运筹学会科普工作委员会主任刘歆研究员、中国数学会副理事长兼秘书长巩馥洲研究员。





5月30日15点，在网友热切地期盼中，讲座正式开始。徐院士的报告从四个部分展开：一、数据科学的产生背景：大数据；二、数据科学的相关概念与方法；三、数据科学的内涵、定义与方法论；四、数据科学的未来发展与重大科技问题。

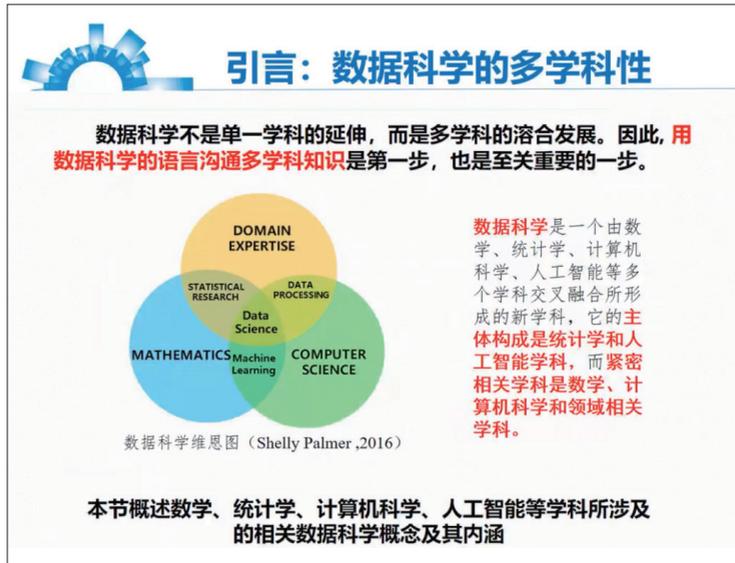
一、数据科学的产生背景：大数据

徐院士首先详细阐述了数据科学产生的背景、大数据存在的大价值、大数据的原理、发展大数据具有迫切性与极端重要性；接着徐院士强调要实施国家大数据战略就应该准确把握大数据在管理创新、科学发现、产业发展、学科发展方面的机遇，因为大数据正在且必将引领未来生活新变化、孕育社会发展新思路、开辟国家治理新途径、重塑国际战略新格局。尽管大数据为国家创新发展带来了大机遇，并已上升为国家战略，但要真正实现大数据的大价值，特别是将大数据转化为现实生产力，仍然面临诸如传统方法失效、认识论上的困惑、方法论上的冲击等科学基础挑战，针对这些挑战徐院士解释了如何从宏观战略方面科学应对。

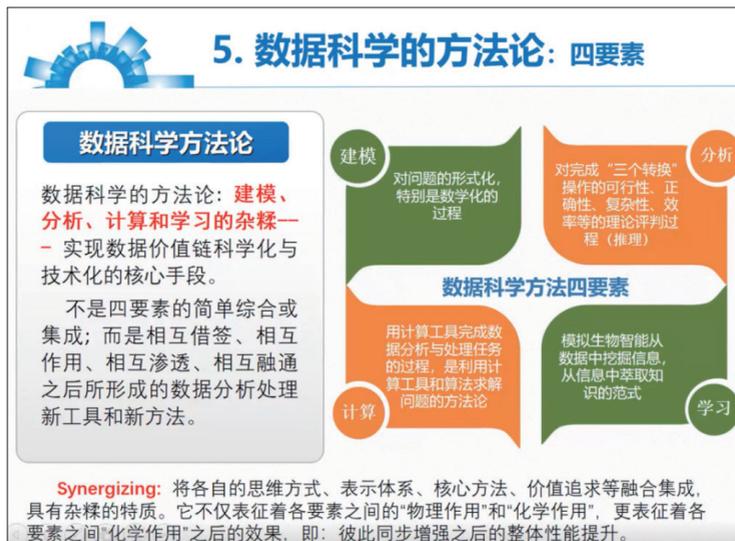
徐院士还阐述了数据科学从人类社会、物理世界、信息空间三元世界理论出发，依据三元世界之间的相互关系来认识数据，具有很强的科学性、完全性与实用性；数据科学融通创新的方法论能够非常完美地将多学科方法进行“杂糅”，从而形成高效地大数据分析与管理技术；数据科学能够支撑大数据原理的实现，从而赋能大数据，使其转化为现实生产力，产生大价值。最后徐院士总结：大数据催生了数据科学，数据科学承载了大数据的未来！

二、数据科学的相关概念与方法

徐院士从数据科学的多样性引入，全面阐述了与数据相关的概念与方法、与计算机相关的概念与方法、与统计相关的概念与方法、与人工智能相关的概念与方法。

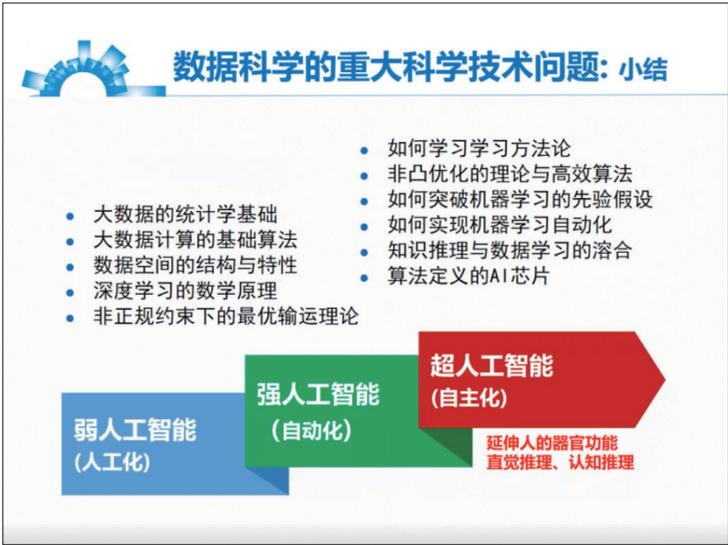


三、数据科学的内涵、定义与方法论



徐院士首先从四个代表性观点出发给出了数据科学的内涵；用新观点给出了数据科学的定义：数据科学是有关数据价值链实现过程的基础理论与方法学。它运用建模、分析、计算和学习杂糅的方法研究从数据到信息、从信息到知识、从知识到决策的转换，并实现对现实世界的认知与操控。其次徐院士从研究对象、科学目标、方法论三个角度解释了数据科学定义的合理性。接着徐院士重点概括了数据科学在计算机科学领域、人工智能领域、统计学领域的重大进展，简单介绍了数据科学的历史演进。最后徐院士从四要素（建模、分析、计算、学习）、与数学的关联与区别、与统计的关联与区别、与计算机科学的关联与区别这四个模块详细介绍了数据科学的方法论。

四、数据科学的未来发展与重大科技问题



从趋势分析数据科学的未来发展，徐院士强调数据科学要在多学科中立身、在大数据中立信、在强基础中立魂。

接着徐院士给出了数据科学十个重大科学技术问题：大数据统计学基础、大数据计算的基础算法、数据空间的结构与特性、深度学习的数学原理、非正规约束下的最优运输问题、如何学习学习方法论、非凸优化的理论与高效算法、如何突破机器学习的先验假设、如何实现机器学习自动化、知识推理与数据学习的融合、算法定义的 AI 芯片。



报告结束后，周爱辉研究员主持了提问环节。嘉宾们分别代表网友提出了三个问题，分别是：大数据未来在算法理论方面会如何发展？做统计理论方向如果想向数据科学发展，或者说与数据科学相结合，该如何规划自己的学习和科研？数学做的人工智能交叉研究和工科的人工智能相比，会有哪些突出的优势？这三个问题是通过中国数学会微信公众号收集遴选的。徐宗本院士对这些问题做了详细的回答。