



中国科协学会学术部 编

新

观点新学说学术沙龙文集

82

大数据时代 对建模仿真的挑战与思考



中国科学技术出版社
CHINA SCIENCE AND TECHNOLOGY PRESS

新观点新学说学术沙龙文集 ⑧2

大数据时代对建模仿真的挑战与思考

中国科协学会学术部 编

中国科学技术出版社

·北京·

图书在版编目 (CIP) 数据

大数据时代对建模仿真的挑战与思考 / 中国科协学会学术部编. -- 北京: 中国科学技术出版社, 2014.7

(新观点新学说学术沙龙文集; 82)

ISBN 978-7-5046-6667-3

I. ①大… II. ①中… III. ①系统建模—研究②系统仿真—研究 IV. ①N945.12 ②TP391.9

中国版本图书馆 CIP 数据核字 (2014) 第 164953 号

选题策划 赵 晖
责任编辑 赵 晖 夏凤金
封面设计 照 心
责任校对 何士如
责任印制 张建农

出 版 中国科学技术出版社
发 行 科学普及出版社发行部
地 址 北京市海淀区中关村南大街 16 号
邮 编 100081
发行电话 010-62173865
传 真 010-62179148
投稿电话 010-62103182
网 址 www.cspbooks.com.cn

开 本 787mm × 1092mm 1/16
字 数 150 千字
印 张 9.5
印 数 1-2000 册
印 次 2014 年 7 月第 1 版
版 次 2014 年 7 月第 1 次印刷
印 刷 北京长宁印刷有限公司

书 号 ISBN 978-7-5046-6667-3/N · 190
定 价 18.00 元

如有缺页、倒页、脱页, 请与本社发行部联系调换。

倡导自由探究

鼓励学术争鸣

活跃学术氛围

促进原始创新

序

大数据作为人类探索世界的一种新方法，在改变我们思维方式的同时，也对仿真科学的创新与发展带来新的思考，提出新的挑战。在与之相关的一系列问题中，既有认识论的问题，也有方法论的问题；既有概念、理论问题，也有具体技术、方法问题；既有仿真学科的问题，又有横跨系统科学、信息科学等多个学科的问题。但目前对这些问题的理解认识都还很模糊，甚至有些混乱。因此，对这些问题的研究将会对大数据及大数据时代的仿真科学与技术的发展，起到积极的推进作用。

为此，中国科协组织举办了第82期新观点新学说学术沙龙活动，邀请了25位相关领域的专家学者，围绕“大数据时代对建模仿真的挑战与思考”这一专题展开研讨、辩论和交流。本次沙龙活动共划分为三个主题单元：

第一单元，以大数据为基础的第四范式是否成立？重点讨论数据密集型的科研形式是否可以单独成为一种新的科研范式，大数据是否动了建模仿真的“奶酪”等问题。

第二单元，大数据方法对仿真建模带来了什么挑战？如，对仿真相似性理论、建模仿真方法、仿真平台等产生什么重大影响？如，无需数学模型的仿真是否存在？

第三单元，大数据方法对仿真建模带来了什么机遇？是否能为复杂系统特性（如涌现、智能等）仿真提供更好的方法？

在一天半的研讨中，本期沙龙很好地弘扬了“敢于质疑、勇于创新、宽容失败”的精神。针对上述三个主题，25位老中青专家在自由、活跃、严谨的研讨气氛中畅所欲言，学术观点鲜明，甚至针锋相对。在思想的不断碰撞与交锋过程中，取得了丰硕的成果，形成了具有原创性的学术观点，归纳起来主要表现在以下三个方面：

第一，通过正反辩论“以大数据为基础的第四范式是否成立”这一主题，深入探讨了大数据来临对于科研范式革命的影响。科研范式，是指那些在一段时间内为科学家集团或者“科学家研究共同体”所共同接受的科学信念，是用于指导现实科学研究的一组假说、理论、准则和方法的总和。一旦范式无法指导新的研究就会发生危机，产生出新的科学成就，即科学革命，而科学革命的结果就将产生“范式转换”。大数据时代，人类不仅面对着数据爆炸的新问题，也呼唤着一种认识改造世界的新范式诞生。

第二，围绕第二个主题，探讨了大数据方法对现有仿真建模理论与技术所提出的挑战。大数据时代，现有建模仿真科学技术在思维方式和科研方式、方法和手段等方面都存在问题，需要变革仿真的思维方式、科研方式及其方法手段才能应对大数据时代的挑战。

第三，围绕第三个主题，探讨了大数据方法给仿真技术与应用的发展所带来的机遇。一致认为，大数据方法与仿真建模方法融合将为仿真技术与应用的发展带来崭新的机遇。主要表现在，将变革现有仿真的思维方式和科研模式，现有的建模方法学，仿真支撑技术系统与仿真系统，以及仿真应用工程技术等方面。为了记录本期沙龙的成果，同时进一步促进更为深入和广泛的探讨，本文集将25位专家的发言进行了汇编与整理，且未做大的改动，一些在沙龙上尚存争议的观点也被收录文中，希望能对读者有所启发。同时，也期望本期沙龙的成果能够促进我国仿真科学技术在大数据时代的发展，为我国建模与仿真理论与技术的发展及人才培养做出新的贡献。

最后，谨代表主办方感谢中国科协的大力指导和支持，感谢各位与会专家和新闻界朋友的热情参与和奉献，感谢沙龙组织委员会的同志们为本期沙龙成功举行所作的辛勤而有效的努力。

李伯虎 刘继中

2013年9月15日

目 录

大数据时代对建模仿真的挑战与思考 (引导发言)	(3)
基于大数据的第四范式问题思考	(17)
以大数据为基础的第四范式现在不能证实成立	(23)
大数据时代 SBE&S 面临的挑战	(45)
大数据技术及其对复杂系统仿真建模的影响	(56)
复杂系统仿真建模的一种新方法——大数据	(92)
大数据的智能处理与仿真	(97)
总结	(122)
专家简介	(126)
部分媒体报道	(138)

会议时间

2013年9月14-15日

会议地点

吉林省延吉市延边白山大厦

主持人

李伯虎

李伯虎：

这是中国系统仿真学会承办的第四次沙龙，主题为“大数据时代对建模仿真的挑战与思考”。为了更好地弘扬新观点新学说学术沙龙“敢于质疑、勇于创新、宽容失败”的精神，我就本次沙龙活动提几点希望：

第一，在解放思想中创新发展。大数据时代已经降临。“大数据”是继云计算、物联网之后 IT 领域又一次颠覆性的技术变革。人们纷纷预测，它对国家、社会、经济、科学的发展和人民生活将产生巨大的价值，并将积极推进人类从“后工业时代”向“智慧经济”时代发展。与此同时，大数据时代中大数据的采集、存储、传输、处理、应用等问题对各行各业的文化、组织、管理、方法、技术等提出了崭新的挑战。不容置疑，它对作为科学研究第三种范式“计算科学”中重要组成部分的“仿真科学”也带来了新的机遇、挑战与变革。这正是我们这次沙龙召开的背景。从各位专家提交的观点摘要中提炼总结、精心设置了建模仿真在大数据时代最具代表性的三个研讨议题，即“以大数据为基础的第四范式是否成立，大数据方法对仿真建模带来了什么挑战，大数据方法对仿真建模带来了什么机遇”。我们殷切期望大家以创新为驱动，大胆发表自己的学术观点和学术灵感，力争取得与“大数据、大价值、大变革”相吻合的成果。

第二，在理性质疑中寻求突破。“科学的本质是批判，交流的本质是质疑”。

大数据时代对建模仿真的挑战与思考

我们的沙龙是一个开放、自由、宽松的平台，鼓励大家敢于质疑。“敢于质疑”是科学研究者应该具备的基本素质。著名教育家陶行知先生曾说过：“发明千千万，起点是一问。”只有敢于质疑，才有可能冲破传统理论的羁绊去创新，才有可能进步，得到完善。但是敢于质疑并不是随意否定，其前提是对相关领域有较为深入的研究，对问题有理性的思考。

第三，学术民主，积极探索。科研创新是一种有风险的探索，可能成功，更可能失败。只有不惧失败，才有可能成功。我们鼓励创新，更宽容失败，目的就是要尽可能多地积累经验，把失败作为财富、把失败作为动力，在持续深入的研究中积蓄力量，探寻真正的硕果。同时，保护科研创新的积极性，尽可能地为创新营造有利的氛围，推动科研领域创新人才的成长。

相信在本次沙龙讨论中，各位专家独到精湛的认知将相互碰撞，激发出智慧之光，为我们仿真科学的进一步发展做出积极的贡献。

大数据时代对建模仿真的挑战与思考（引导发言）

◎胡晓峰

本次沙龙的题目是“大数据时代对建模仿真的挑战与思考”，主题不完全是“大数据”，而是要对“大数据”背景下的建模仿真影响进行研讨。

近两年来，“大数据”这个词已经被广泛提及，甚至有被用滥了的嫌疑。但大数据究竟会对科学研究带来哪些影响，还需要更深入的研究。对受到直接影响的建模仿真领域来说，同样面临这个问题。大数据对建模仿真带来哪些困惑、挑战和机遇？这就是中国科协召开这次沙龙要讨论的主要问题。

一、问题的相关背景——大数据及其特性

什么是大数据？大家听着很熟悉，但至今仍没有一致认可的准确定义。一般用4个V进行定义，也有用6个V的。即：第一个V（Volume），即规模性，指的是体量大，一般在TB、PB乃至EB以上；第二个V（Variety），即多样性，也就是信息的种类多，并以各种信息载体形式存在；第三个V（Velocity），即高速性，要求处理速度在合理时间之内；第四个V（Value），即价值性，或者叫真实性（Veracity），是说大数据一般是高价值低密度，但也指对真实数据的处理。

也有其他的定义，比如说维基百科的定义是，大数据是指利用常规软件工具去捕获、管理和处理数据所耗时间超过可容忍时间限度的数据集。这个定义是从处理方法角度来说的。当然，还有专家认为，大数据是在赛博空间（Cyberspace），“人、机、物”三元交互融合产生出来的结果。

其实不管它的定义是什么，大家知道这个意思就差不多了。但是，我想提出一个问题，即为什么现在才有了大数据？我觉得主要有以下几个原因：

第一，信息技术的发展创造了数据产生和处理条件。像云计算及网络、存储设施、数据库等技术的发展，以及现在开始冒出苗头来的广泛应用，像物联网、

RFID 的使用以及视频监控等技术的普及应用等，都是如此。

第二，特别是因为互联网运用的广泛普及。互联网带来了大量数据，例如社交网络、博客、微信、基于位置服务、搜索服务等，已经遍地开花。有统计说，“近两年产生的数据等于 2010 年前人类产生数据的总和”，每两年数据翻一番，可见数据爆炸性的发展。

第三，各类大数据应用产生了很好的效果并提出了更高的要求。对数据的深度挖掘获得了出人意料的效果，已远超早期数据挖掘“啤酒与尿布”等经典案例的水平，出现了诸如“纸牌屋”、“点球成金”这样的新传奇，并且得到各界的广泛关注。

此外，各国政府对此十分关注。美国奥巴马政府 2012 年 3 月发布了“大数据研究与发展倡议”，已将其作为美国未来发展的重要战略，并作为国家战略启动了“大数据发展计划”。如果我们跳出技术看国家战略问题，我认为奥巴马的想法就是要通过“大数据发展计划”，试图再次重复“信息高速公路计划”给美国人带来的互联网霸权所获得的广泛优势。从这次“斯诺登事件”我们可以发现，美国的互联网霸权已经置其他国家于非常危险的境地，而大多数人还浑然不觉。现在，美国人已经把目光瞄准到大数据等未来新的领域，我觉得这就是为创造未来的大数据霸权奠定基础。

经济界也很关心。达沃斯经济论坛发表了关于大数据的研究报告，在论坛上成为各国首脑和企业家关注的主题，这是很引人注目的。比如说，论坛的主题就叫做：“大数据、大影响，国际开发新可能”。我国各有关部门和企业也十分重视，召开了很多会议，许多企业家也发表了一些重要的言论。

学术界也十分关注。2008 年，《自然》杂志推出大数据专刊，研究“PB 时代的科学”，探讨科研形态变化，认为“以数据为准绳的理念指导，以及强大的计算能力支撑，正在驱动一次科学研究方法论的革命”。《科学》杂志 2011 年也推出专刊“Dealing With Data”，围绕“数据洪流”展开讨论，将大数据深度分析看成未来的研究重点与突破点。同时，各类学术机构也纷纷组织各种研究和探讨，发表了大量研究报告，召开了各种会议，并成立了许多大数据学术组织等。这些都已经如火如荼地展开了。

但是，也有很多疑问同时在我们每个人头脑中会产生出来。大数据到底是什

么？值得我们这么兴师动众吗？主要疑问有以下三个方面：

第一个疑问：大数据与以前一些数据概念有哪些不同？它与我们早期提出的海量数据 (Massive Data)、超大规模数据 (Very Large Data) 有何不同？是不是一个意思？如果是一个意思，对它来说这些就是大数据；但如果不是，又应该是什么意思？

第二个疑问：大数据方法与过去的的数据方法有什么差异呢？比如说，我们早就提出了数据挖掘的概念，以及数据分析等理论，但现在又来说大数据的数据挖掘和数据分析，有什么不同吗？

第三个疑问：大数据应用与过去基于数据分析的应用又有什么不同呢？最典型的的就是商业智能 BI (Business Intelligence)。商业智能也是用数据分析得到一些对未来的洞见。现在大数据出来到底又有什么不同呢？

这些名词都似曾相识，一点都不陌生，但是为什么它又掀起了一个如此值得大家关注的热门话题呢？我个人认为最根本的是，大数据带来了全新的研究思维和方式，主要表现在四个革命性的特征上：

第一，从局部到全体：将网络化的大数据作为分析对象。

这不同于以往的方法。直接面向全体的数据，这里“全体”主要指的是某一研究的所有数据，或者说是网络化的数据。

(1) 网络化的数据分析，“数据大”是条件。这不是过去只对少量样本数据进行分析的数据分析，而是直接面向整体数据，或者叫作所有数据，甚至说是全部数据的数据分析。没有一定规模的量，没有这个“大”字是不成立的。而网络化则是核心。网络化最重要的一点是，它终结了还原论的分解式分析方法，而从整体关系进行考虑。也有专家认为，如果没有网络，大数据也不能成立。

(2) 对数据的处理完全不同于传统数据库的分析与处理方法。在大数据方法中，一是要将“局部的和明确的数据”转化为“所有几乎全部且不明确”的数据。有位专家打了一个很好的比喻，他说：我们过去的数据库处理是在池塘里抓鱼，池塘里养了多少鲤鱼、鲢鱼、草鱼都是心中有数，投放多少就收获多少，这就是“池塘里抓鱼”，是我们原来的处理方式；而现在大数据方法是要在“大海里捞鱼”，有“鱼”与否并不知道，也就是有什么问题也不知道，需要我们想各种方法把“鱼”捞出来。我觉得这个比喻非常恰当。二是要变“脱机”处理为

“联网”处理。处理的同时，数据还可能发生变化，因为它与网络的关系极为密切！因而大数据的处理规模、类型、模式、工具、对象都会有所不同。

第二，从单纯到繁杂：接收数据的繁杂和不精确。

我们常用的数据库一般要求数据是非常干净的，对不干净的还要加以清洗。而大数据则要以非结构化、种类繁多的数据为主，因此就要抛弃对有条理和纯净数据的偏爱，容忍凌乱数据的存在。在操作上，不以“匹配性查找、增删改管理”为数据库应用目标，而是更深层次的应用。“海量”、“超大规模”指的都是数据量，而没有涉及数据的本质问题。另一方面，大数据方法的不确定性和涌现性的特点比较突出。在不确定性方面，它的数据来源不确定、处理模型不确定、模型参数学习也不确定等。在涌现性方面，包括演化模式的涌现、群体行为的涌现、网络智慧的涌现等。这些都可以找到大数据的不确定性和涌现性的影子。

第三，从因果到关联：更强调相关性而非因果性。

大数据最重要的思想是放弃对事情原委的追求，取而代之的是对相关性的接纳，因此它更适合于回答“是什么”，而不是回答“为什么”。这就为“知其然而不知其所以然”找到了依据：直接获取答案，不去问为什么。也就是说，它可以告诉你这样不会错，但是为什么？不知道。

之所以这样做，是因为许多事物的因果关系是难以明确的，有的可能是找不到，而也有的是根本就不存在这样的因果关系。这就完全颠覆了我们过去关于牛顿、爱因斯坦科学体系下因果关系明确的还原论思想，当然就这个问题大家还可以进一步探讨。因此，大数据方法认为，根据海量数据的相互关系，已经足以产生新的发现。也就是因为如此，美国和欧盟都展开了一些相关的研究，初步统计大概有 20 多个研究计划，比如说“大脑扫描计划”、“星球皮肤计划”、“太空追踪计划”等，或多或少都采用了这个思想。

第四，从简单到深入：更强调深度和间接分析。

将简单分析方法发展为大数据的深度分析方法。我们过去的智能分析实际上还是强调以因果关系为主的简单分析，主要针对已有数据的分析，如商业智能 BI 的因果分析。但是大数据就具有了自己明显的特色，更关注深度、间接、外推分析等。有许多数据分析结果的质量依赖于数据量的多少，比如说苹果公司手机的 Siri 语音识别，就是根据联网数据分析的。还有的分析属于外延分析，也就

是分析的结果并不是分析的初衷，而是通过这个结果得到另外的结论。比如说基于搜索词的流感趋势分析，就是典型的范例。分析的是搜索词，但结果却是流行病的预警。还有就是按需分析，有意地产生所需数据，再进行分析，比如说“数据客”。

这些革命性的特征变化，在很多有关大数据的会议上都在反复探讨，大家都有自己的看法。但是，我们这期沙龙的主题是要研究它对建模仿真带来的影响，以及它会带来哪些挑战和机遇。

二、讨论的主要问题：大数据带来的挑战和机遇

第一个问题：以大数据为基础的第四范式是否存在？

首先说一下什么是科学研究的范式（Paradigm）。这是托马斯·库恩在1959年《科学革命的结构》一书中首先提出的概念。所谓范式，是指那些在一段时间内为科学家集团或者称为“科学家研究共同体”所共同接受的科学信念，是用于指导现实科学研究的一组假说、理论、准则和方法的总和，也是我们大家都公认的一些方法。一旦范式无法指导新的研究就会发生危机，产生出新的科学成就，这就叫科学革命，而科学革命的结果就是一种新范式的诞生，这个过程就叫“范式转换”。因此，从一个范式到另一个范式不是渐进的，只会是跳跃式的。这就是托马斯·库恩提出的范式的概念。

现已存在的科学范式，总的来说主要有以下几种，这是大家已经公认的：

第一范式称为“科学实验”，主要是通过实验的方法来观测、记录和验证得出发现。比如说伽利略的斜塔落球实验，以及天文观测实验等等，都属于这类。

第二范式是理论推导，也就是通过逻辑推导、数学证明得到发现。爱因斯坦的相对论是典型代表。

第三范式是科学计算，其中最重要的是建模仿真，所以也有人将其简单地称为建模仿真范式。我们学会组织召开第58期主题是“复杂系统的困惑和思考”，讨论建模仿真是不是第三范式这个问题。通过科学计算、建模仿真的方法来得到科学发现，大家有不同的看法，但大多数人还是认可科学计算是第三范式的。

那么，第四范式是什么呢？是不是大数据范式呢？第四范式是微软公司的吉姆·格雷提出来的。他说，数据探索性的研究方式，也就是基于数据密集型的科

学发现，是未来一个非常重要的趋势。这些科学研究从以数学模型计算为中心的方式，将要转为对海量数据处理为中心的方式。在数据达到一定规模之后，科学研究的模式就会发生根本的转变，“量变”转换为“质变”，这就是一种新的范式的诞生。

因此，大数据可以独立于基于数学模型的科研形式，单独成为一种新的科研范式。针对他提出的这个观点，现在就有不同的争论。

第一种观点认为第四范式成立。只要数据足够大，只靠数据可以完成科学发现，因此不再需要数学模型。传统上我们认为，数据只是模型仿真运行试验的基础，模型是“引擎”，数据是“汽油”，是模型最重要的组成部分。而现在数据可以成为发现的主体，数据的来源可以多种多样，可以通过仪器采集，通过网络收集，也可以通过仿真系统生成，但在之后数据就可以脱离模型成为科学发现的主体了。这就是所谓的“数据优先”模式。

《连线》杂志主编克里斯·安德森有一个惊人的断言：“数据的洪流使传统科学方法变得过时”，“相互关系已经足够，没有了具有一致性的模型、统一的理论和任何机械式的说明，科学也可以进步”。我理解他的意思，即建模方法对于科学而言，并不是必须的。其实，模型的形式是很多样的，一个沙盘可以是模型，一个公式也是一个模型，各种各样的科学方法都是建立在各种不同种类的模型基础上的。但是，他认为未来模型可以不要了，大数据方法就是一种新的科研范式，特别是将搜索方法可以发展成为一种科学研究的方式。正如我们每次研究前都喜欢“Google一下”，或者“百度一下”，查查看看有什么东西可以参考。可见这种方法已经成为了科学研究的一种常见手段。

因此，有人认为，科学研究将会产生三个重大的变化：第一，将一般科研活动中“精心设计并提出问题”的研究环节，变为关键词的选择，因而就不再需要假设，也不需要理论的指导；第二，摆脱试验的束缚，在海量观测数据或现实镜像世界中去寻找关联，即使有个别模型，也不影响其整体理论的存在，也就是所谓的相关性理论；第三，不再对研究结果进行解释，因而出现了能够预测但不解释的科学，这在以前我们会觉得匪夷所思，但现在却非常正常。有人说，有五种科研方法：比如亚里士多德提出的逻辑方法，培根提出的试验方法，牛顿提出的数学方法，以及费米提出的模拟方法等。现在又多了一个谷歌提出

的关联方法。

第二种观点认为第四范式不成立。这一观点认为，单独将大数据的分析独立出来，并不能形成独立的科学研究模式，更不能单独进行科学发现。大数据需要获得大量的数据，这些数据要么从现实中采集，要么利用仿真获得。它们都是科研过程中一个不可缺少的组成部分，不能割裂，大数据方法只是使分析方法更丰富了一些而已。任何分析都需要模型，没有数学模型是不可能进行分析的。即使谷歌搜索，也用到了各种搜索算法和匹配排序模型，使用数据都需要初筛，初筛就要用到模型和假设。正是由于模型和计算机的辅助，才能理解和驯服大量的数据，因此，脱离模型的数据范式，也就是第四范式是不能成立的。

这种观点认为，要说第四范式是一种新的科研范式，是不是显得这种范式太粗糙了？新的发现依靠的居然是毫无技术含量的重复性计算，而不是科学家的敏锐思想。对科学结果只预测、不解释，又有什么意义呢？数据量的增加可以提高研究的质量，但不会引起科学研究方法的本质变化。“不解释”只能说明认识活动还没有完成。“大数据”是否预示着科技“去人类化”？这会改变科学家在科研活动中的角色，也将改变科研活动的基本规律。我们知道，科学的神圣任务是“揭示隐藏在混沌世界中的有序结构”，难道可以只靠网络和计算机的蛮力就可以完成吗？科学家的思考就不再重要了吗？这就是第二种观点的主要思想。

在这里我们还要回答一个问题，就是大数据的出现，是否动了仿真的“奶酪”？因为我们知道，以大数据为基础的第四范式是从第三范式（仿真建模、科学计算）中独立出去的，因而我们需要回答两个方面的问题：一方面，第四范式是否成立？它发生的实质性变化是确实需要“范式转换”，还是仅仅是对第三范式的一种扩充，或者是其一种特殊形式？在第58期学术沙龙中讨论了第三范式的必要性，其中说到，如果从一个坐标轴来看，以现实观测为主要形式的科学实验范式占据着一端，完全抽象的理论研究占据着坐标轴的另一端，而建模仿真则以“虚拟现实”的方式，正好居于这个轴的中间，非常的合适和匹配。但现在大数据来了，那它应该摆在什么位置呢？另一方面，如果第四范式成立，它又应该包括哪些内容？它对第三范式产生什么影响？原来的第三范式又会发生什么变化？因为它与其直接发生冲突了。

第二个问题，大数据对建模仿真带来什么挑战？

大数据理论的出现，对传统建模仿真学科带来了挑战，很多问题需要我们认真地研究和讨论，因为这些挑战有可能会动摇或变革原有仿真实论的基础。

谷歌研究部主任彼得·诺维格有一句名言：“所有的模型都是错误的，进一步说，没有模型也可以成功。”怎么去理解这句话？《复杂》杂志也说过：“量子力学和混沌摧垮了精确预测的希望，哥德尔和图灵的结果摧垮了数学和计算无所不在的希望”，大数据会不会也是这样，摧毁我们原来的一些观点呢？会带来哪些挑战呢？

第一，对仿真基本理论的挑战。

众所周知，仿真是基于相似性理论的，但它可否基于相关性理论，这是第一个观点。我们知道，系统仿真是建立在相似性理论基础之上的，通过对实际系统的建模，使两者之间具有相似性，因而推论，如果输入相似，则认为输出也应该是相似的，从而从仿真系统中得到实际系统的结果，达成仿真目的，得到实际系统难以得到的结果。用相似性中的类比方法来获取结果，是仿真科学最基础的观点。

仿真的目的就是发现问题和预测未来，时间轴前推就可以是预测。但在某些情况下，大数据方法可能做得更好，比如它可以在预测方面做得更好。例如，已有研究证实，通过分析谷歌中相关搜索词语的统计结果，如“流感症状”等，可比国家疾控中心早一周预测流感的爆发。它依据的是相关性，通过外延效应进行的间接预测，而且这种结果与官方结果相关性高达 97%，当然这是美国人得出的结果。如果我们真的去建立“流感预测”仿真模型，可能费了半天工夫，还建不出来这个模型，或者建出来的模型也很粗糙，根本不能用，因为很难找到两者之间直接的因果关系。所以，我们的问题是，基于相关性理论是不是可以部分取代基于相似理论的建模仿真？

第二，对建模方法的挑战。

有没有不要数学模型的仿真？某些复杂的事物未必有可行的数学模型，比如说它的复杂度非常高、计算量非常大、在可行的时间内做不到，等等。但我们可以建立起认识问题的“数据模型”，比如说谷歌关联研究等。也就是说，大数据提供了利用“数据模型”的新途径。

有人举过一个例子：你们知不知道自己的第一好友是谁？在没有数学模型的