



机理与数据的融合计算

张平文

北京大学

2022年9月

目录

- 一. **应用数学的演进**
- 二. 融合计算的模式
- 三. AI for Social Science
- 四. 新的增长点
- 五. 体制机制

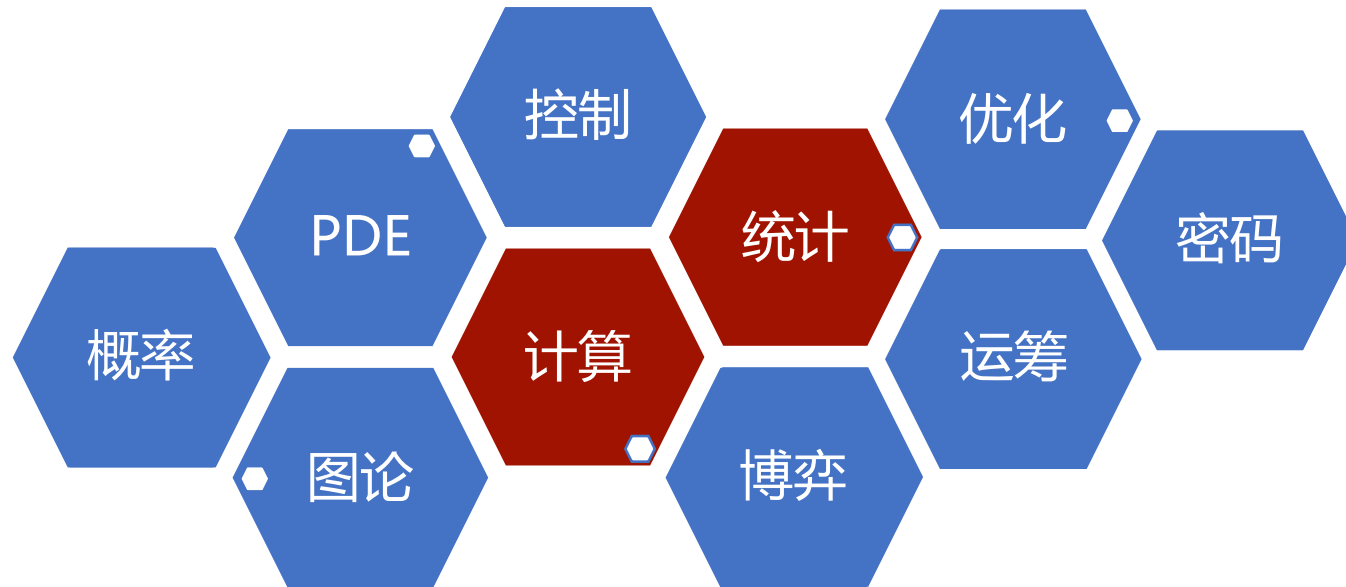
科技
创新
在北大



一. 应用数学的演进



应用数学研究方向



- 方向碎片化
- 发展并行化

一. 应用数学的演进



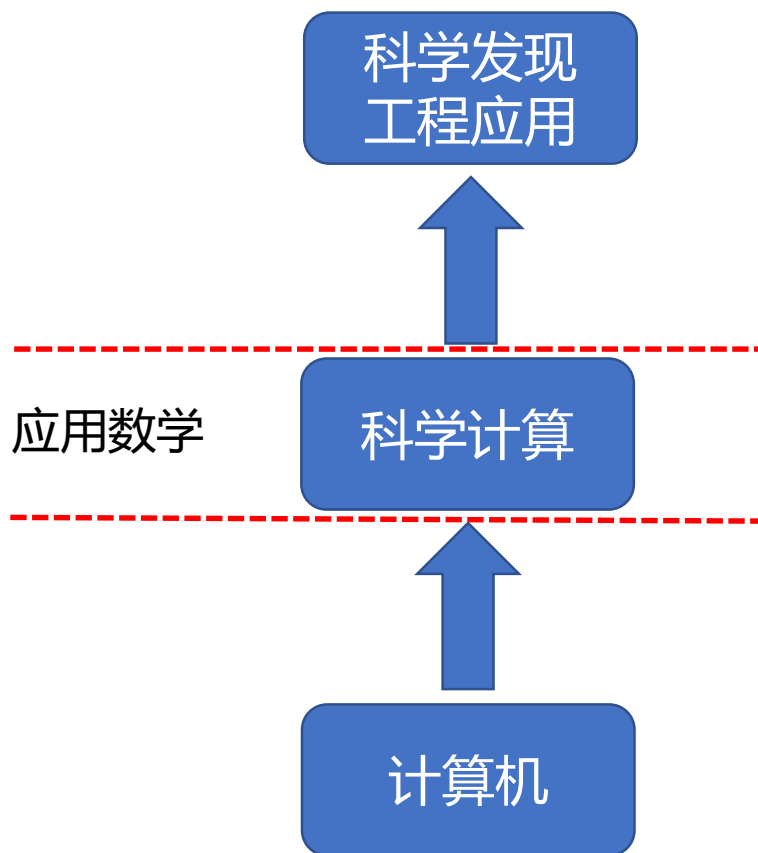
当代应用数学发展的三个阶段-1

早期

以机理为中心



科朗所

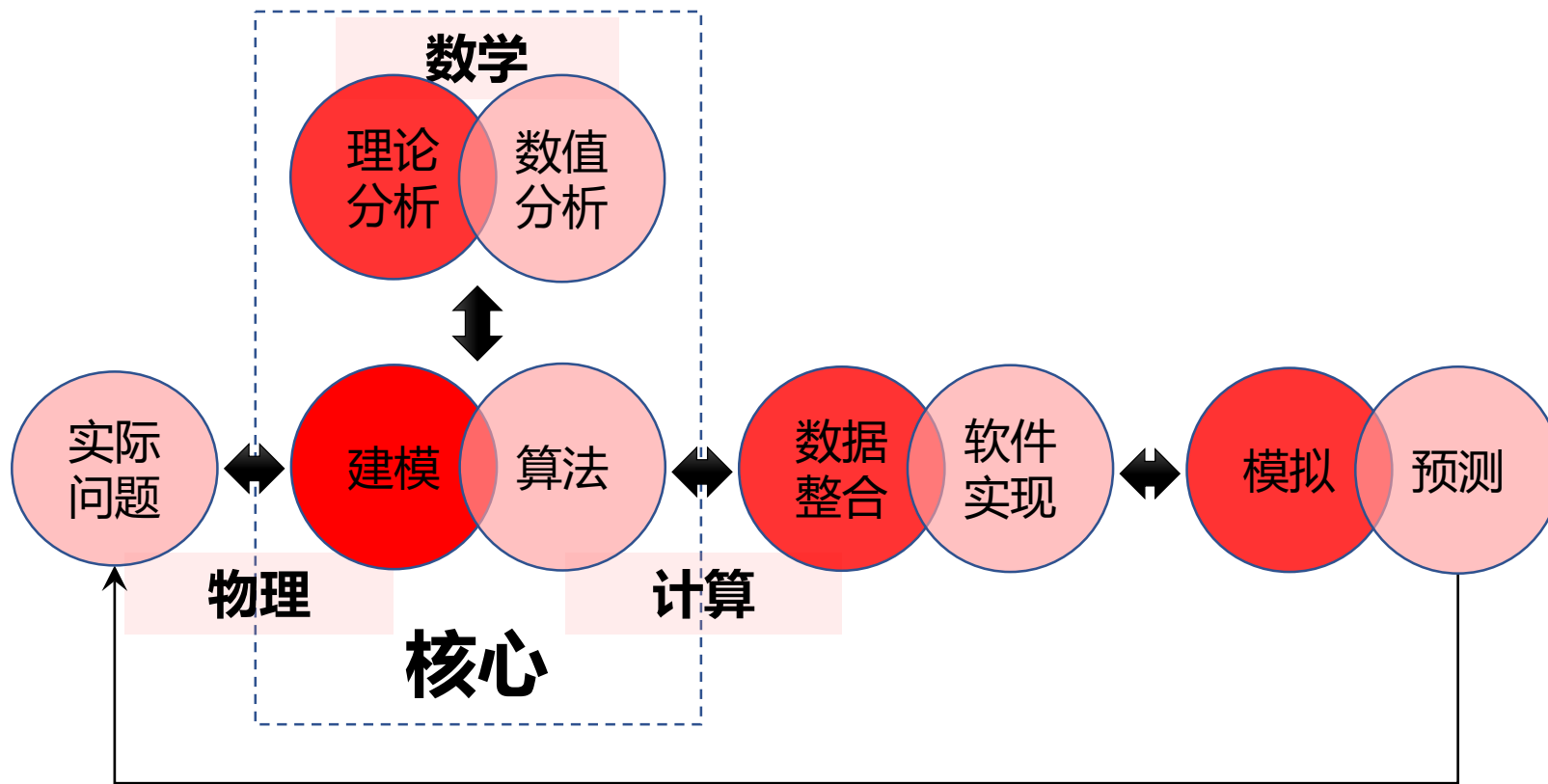


- 研究基于**第一性原理**构建的偏微分方程及数值算法
- 在**流体力学**、**结构力学**等领域取得巨大成功

一. 应用数学的演进

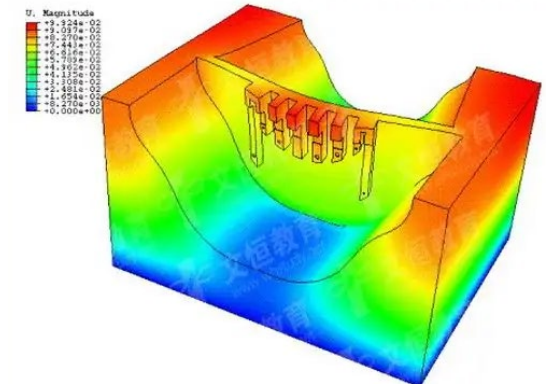


基于机理的计算



特点:

- 问题明确, 有一套系统的方法
- 重点在于建立有效的模型和高效的算法
- 不需要数据, 或只用较少数据



大坝桥梁的应力分析

一. 应用数学的演进



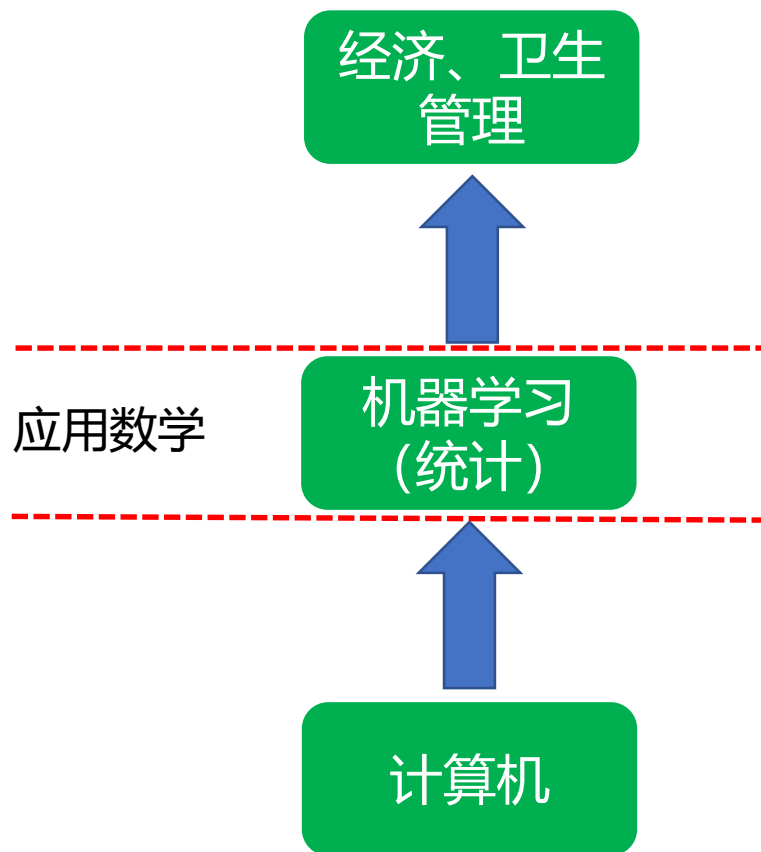
当代应用数学发展的三个阶段-2

近20年

以数据为中心



斯坦福大学

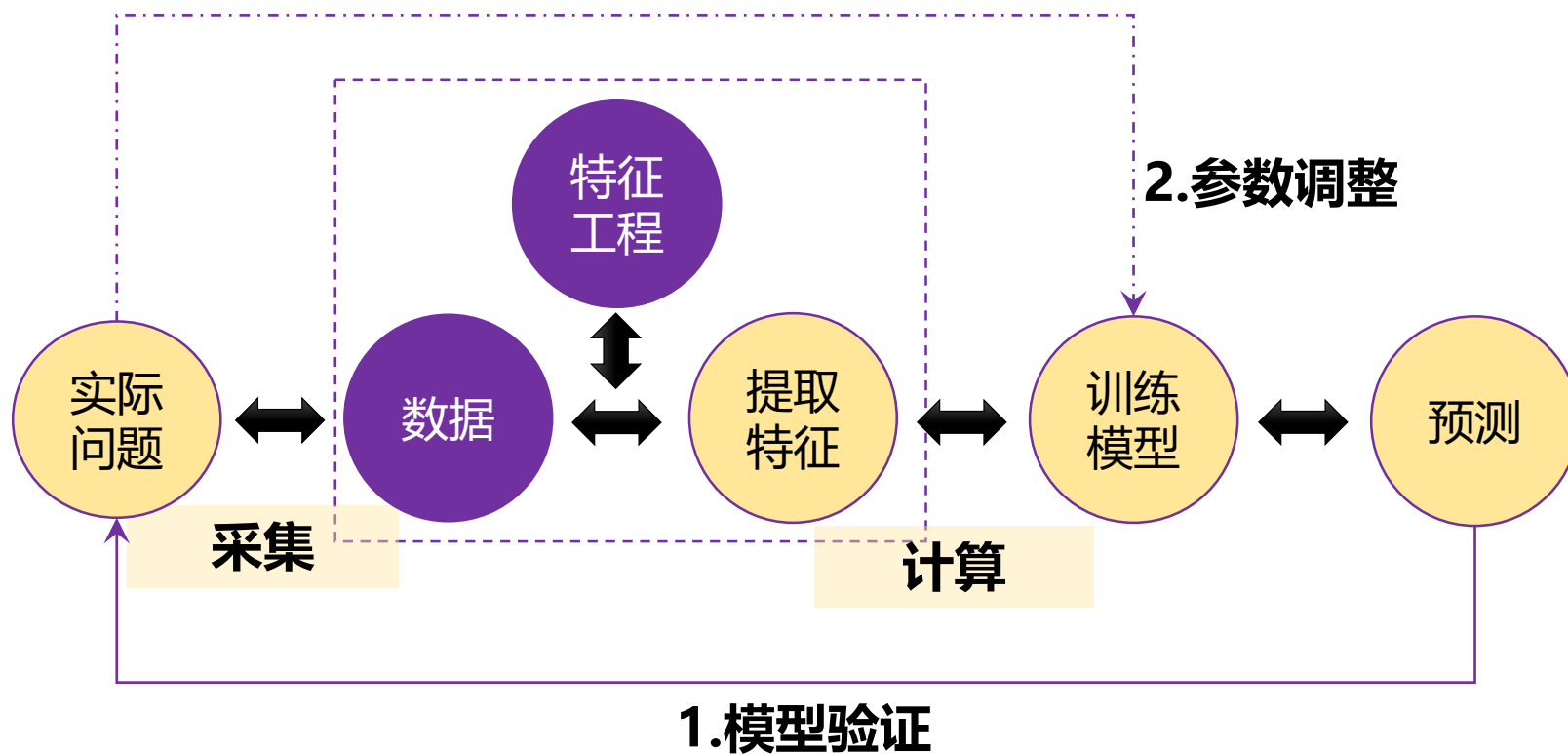


- 利用**数据**直接构建数学模型，规避传统模型中复杂的计算问题
- 在**经济、卫生**等传统统计领域中效果显著

一. 应用数学的演进



基于数据的计算



特点：

- 模型是否简洁不再重要
- 重要的是需要足够多的数据和有效的特征



基于用户数据的广告推荐

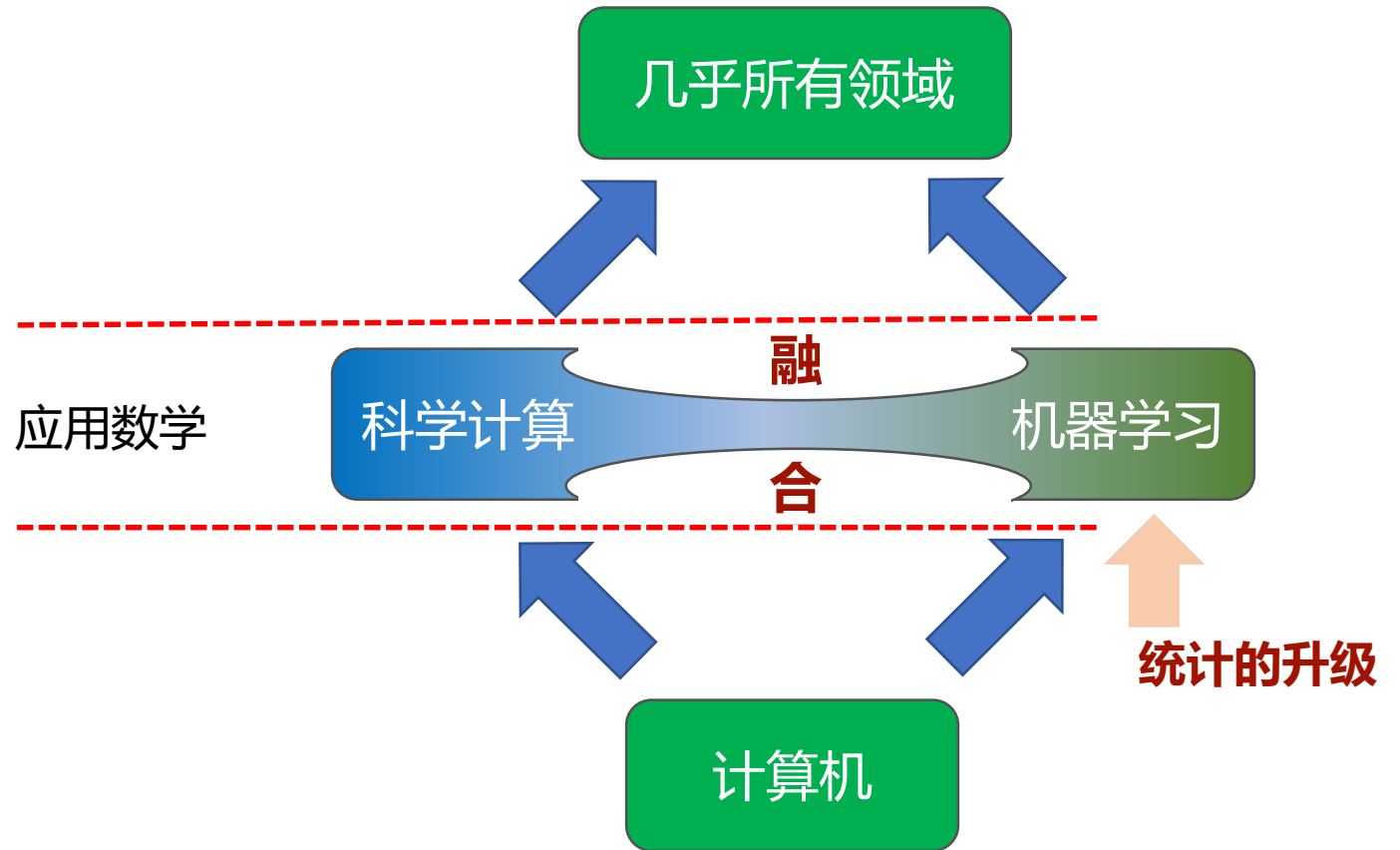
一. 应用数学的演进



当代应用数学发展的三个阶段-3

现在

机理与数据的
融合计算



一. 应用数学的演进



机理与数据融合的原因

机理

- 特点：追求简洁与美
- 优势：时效性长
- 不足：部分问题过于复杂，难以处理

数据

- 特点：追求实用性
- 优势：适用范围广
- 不足：可解释性差，时效较短

取长 ↓ 补短

机理与数据的融合计算

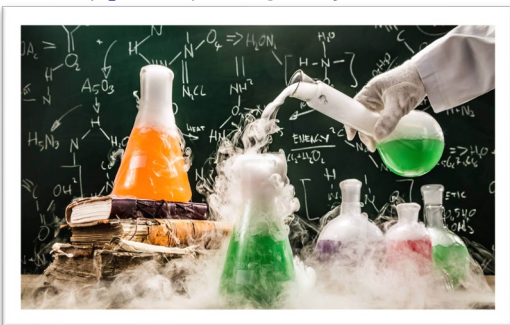
- 兼具二者优势
- 为工程问题的最后一公里提供了实现途径

一. 应用数学的演进

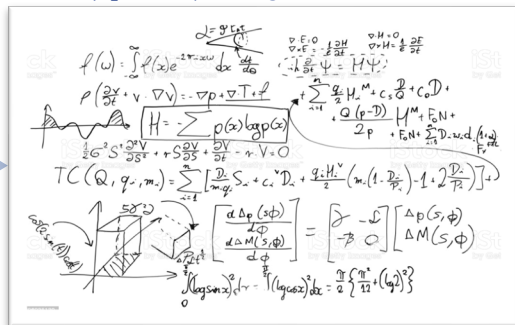


计算已成为**所有**科研和创新的重要范式

第一范式：实验



第二范式：理论



科学计算

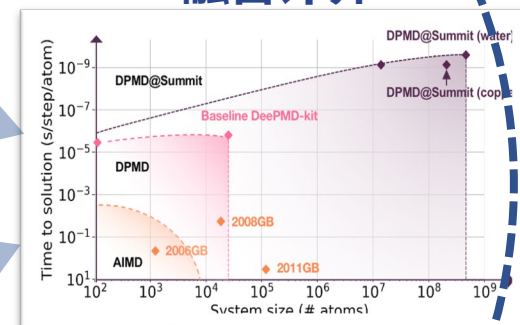


数据



计算

机理与数据的融合计算



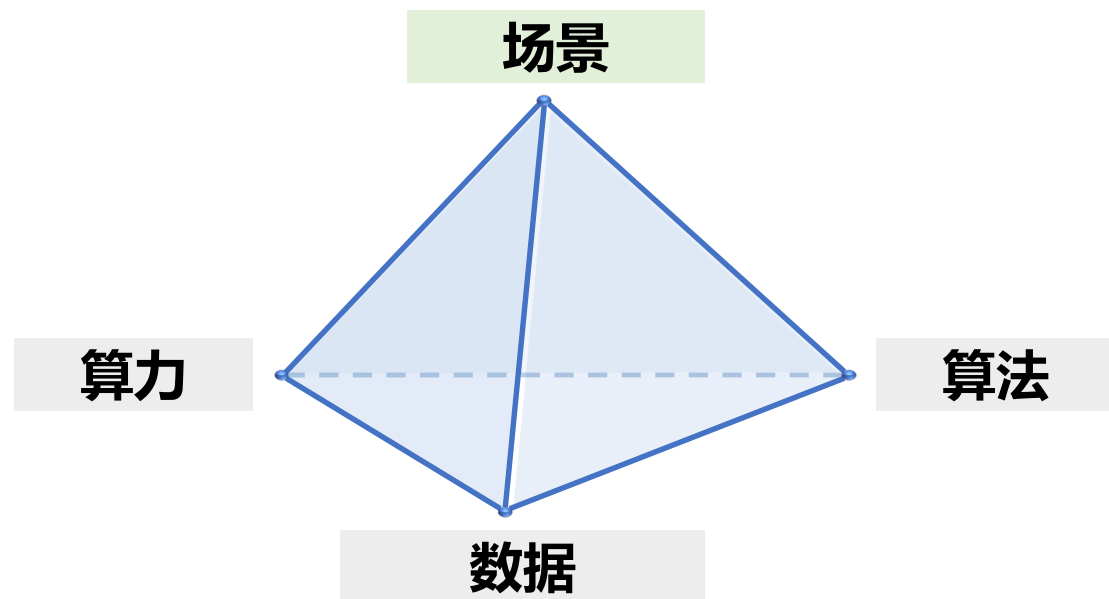
一. 应用数学的演进



计算的4个要素

计算的要素

- 计算与数据、算力、算法、场景密切相关，成为计算的4个要素



计算4要素
形象称之为
计算四面体
或计算金字塔

一. 应用数学的演进



原始数据类型多样

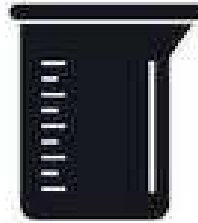
实验数据



观测数据



记录数据



调查数据



模拟数据



- 4个要素中变化最大的是**数据**

目录

- 一. 应用数学的演进
- 二. 融合计算的模式**
- 三. AI for Social Science
- 四. 新的增长点
- 五. 体制机制

科技
创新
在北大



二. 融合计算的模式



从AlphaGo到AlphaZero



- 基于数据的计算
- 使用**记录**数据



- 基于数据的计算
- 使用**模拟**数据

- ◆ 临界点（模拟数据）
- ◆ 极限（迭代升级）

二. 融合计算的模式

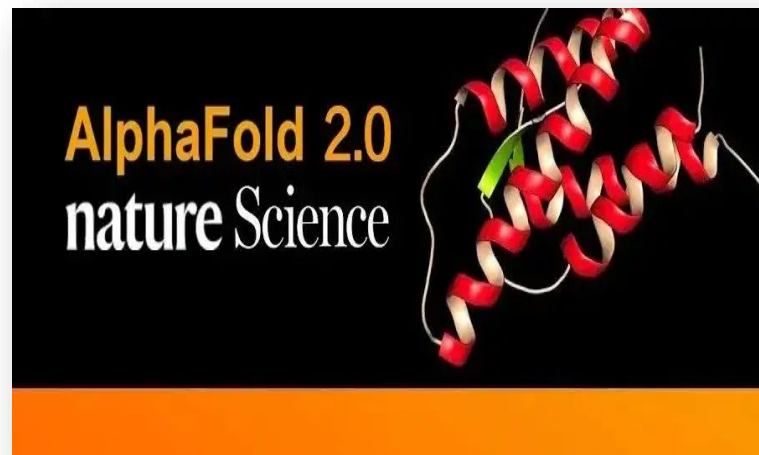


AlphaFold2

蛋白质折叠
研究的路径

基于机理的计算

基于数据的计算



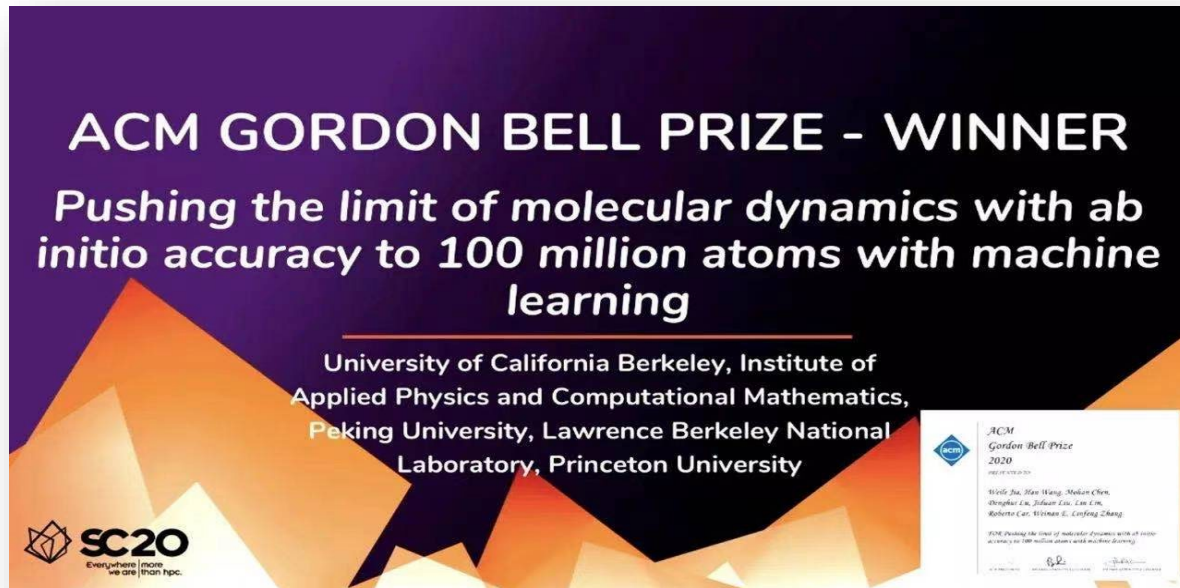
- AlphaFold2采用了基于数据的计算, 并使用**实验**数据
- **融合计算**是该领域研究的未来方向

二. 融合计算的模式

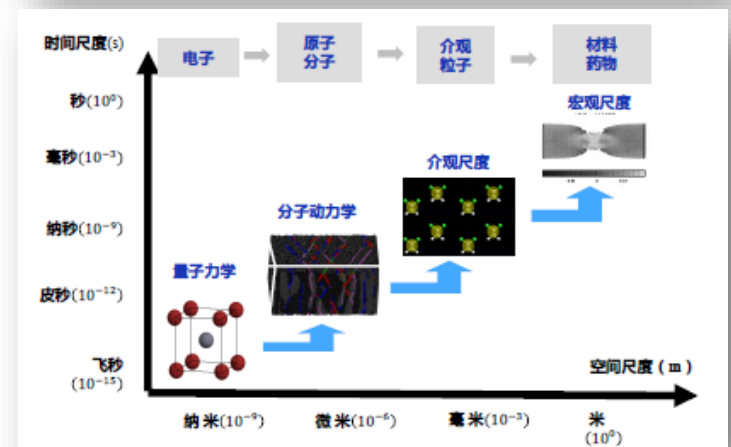
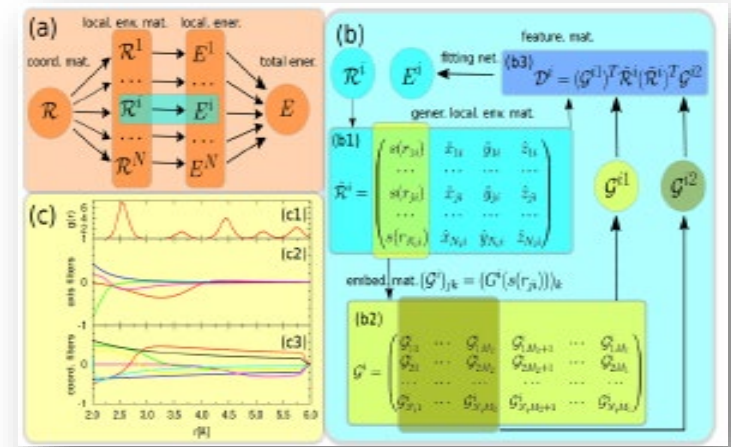


深度势能分子动力学

- 以机理为主的融合计算



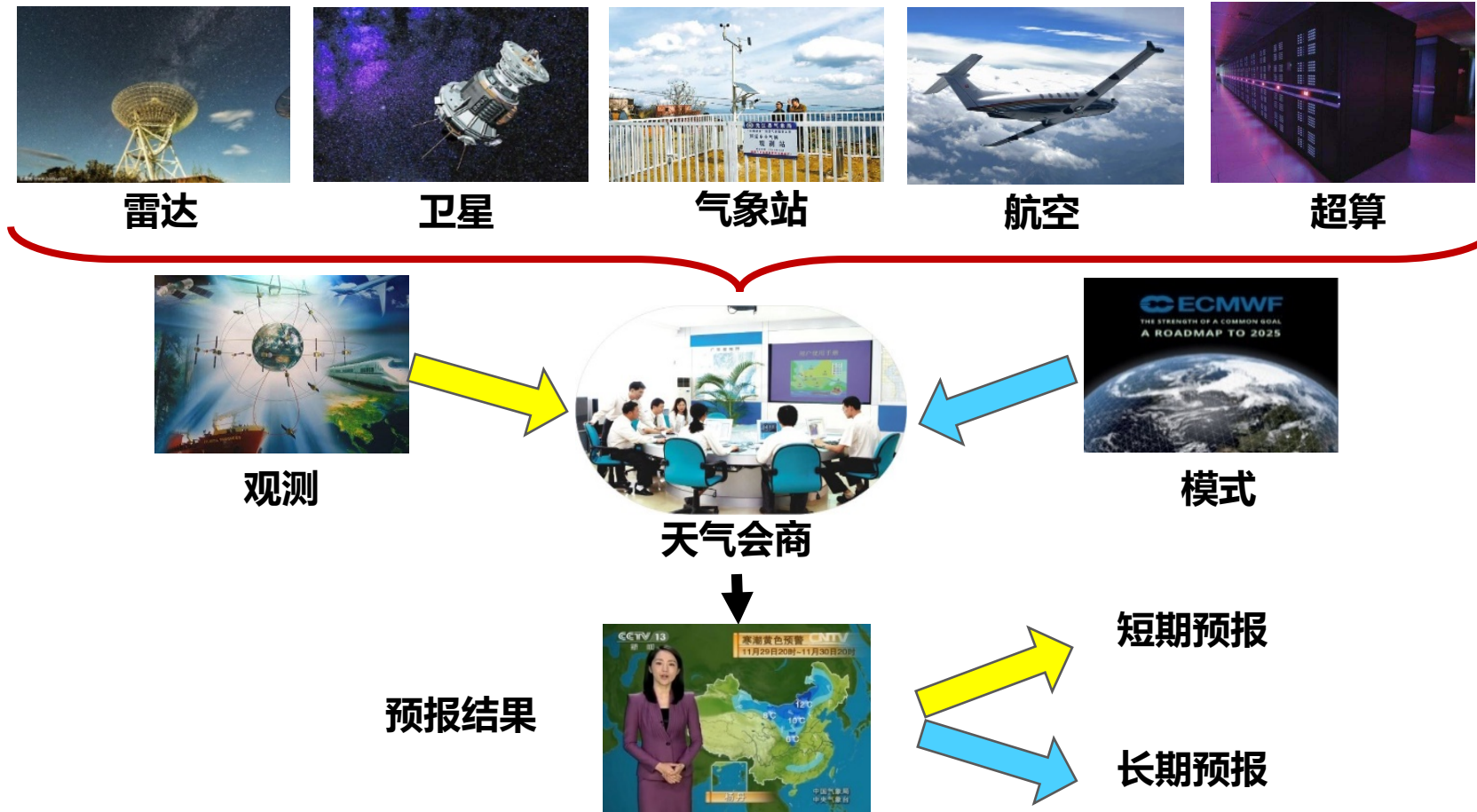
2020年获戈登贝尔奖



二. 融合计算的模式



气象预报：一个典型的融合计算场景



二. 融合计算的模式



资料同化：气象预报中融合计算的成功案例

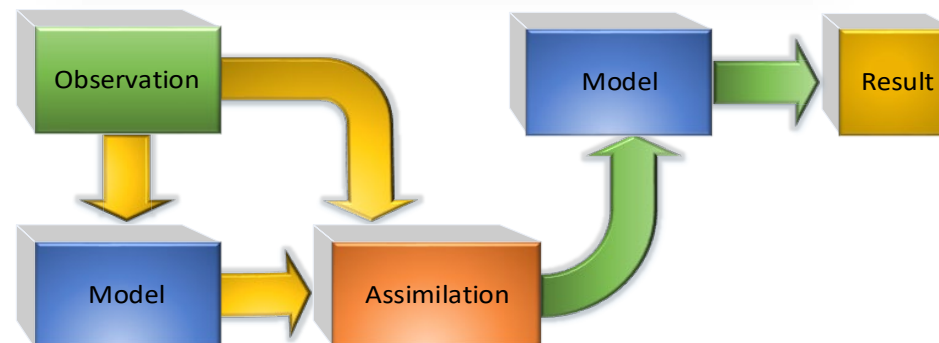
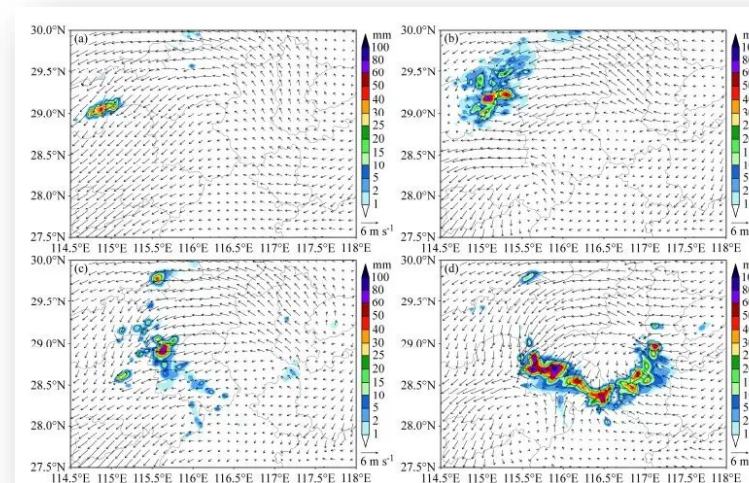
资料同化

含义：将动力系统的 **观测** 数据并入该系统数值模式的 **初值** 的过程

目的：为数值模式提供一个好的**初值**

资料：不同精度、不同时段、不同类型的各类观测数据

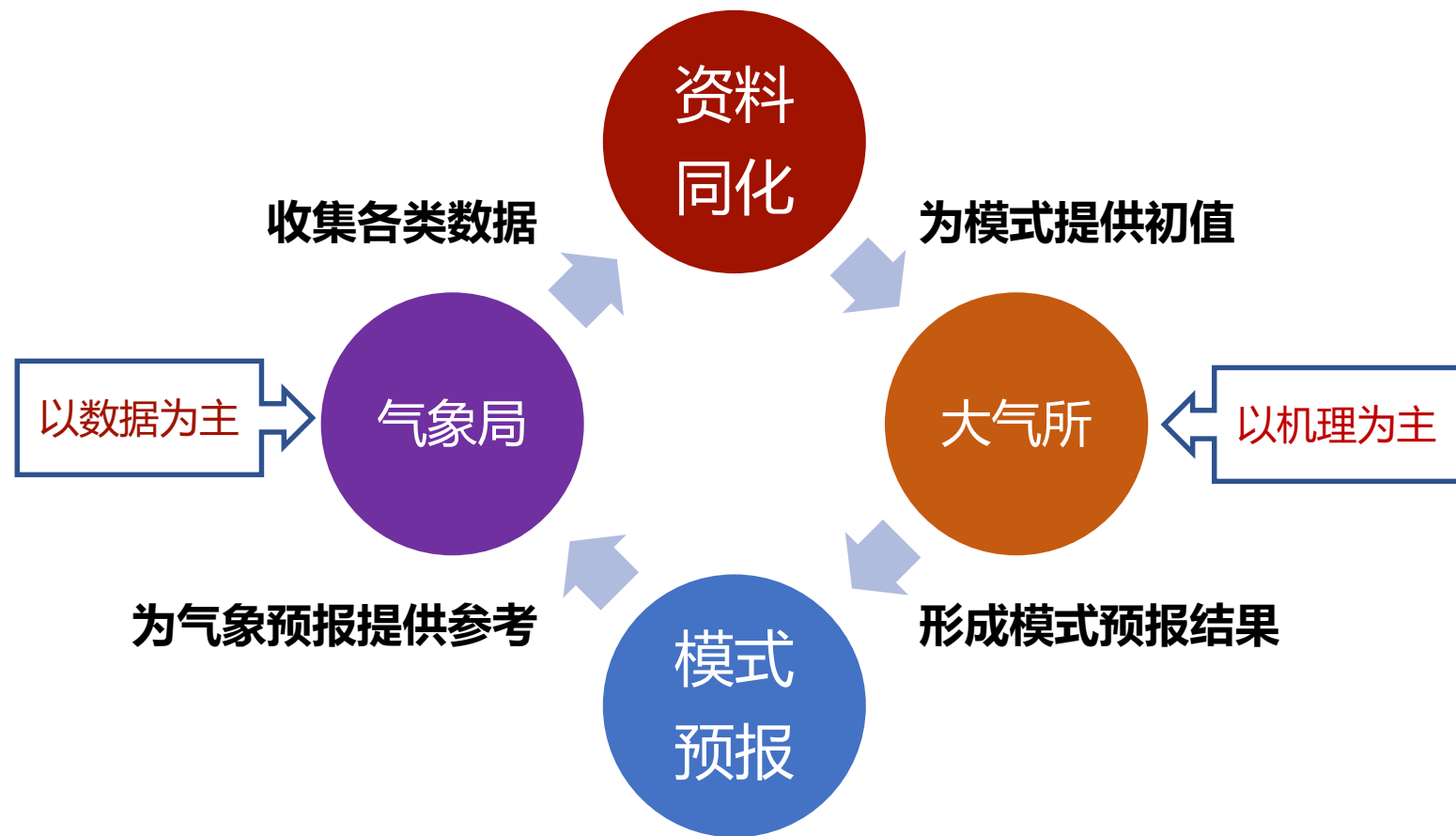
方法：逐步订正法、最优插值法、卡尔曼滤波法、变分法等



二. 融合计算的模式



资料同化：气象预报中融合计算的成功案例



- 资料同化有效提升了气象预报的水平
- 资料同化是机理与数据融合计算的典型案例
- 融合计算在气象预报中可以有更广泛应用

- 现阶段主要用**优化**方法
- 未来**机器学习**非常重要

二. 融合计算的模式



观测和模式如何更深度的融合？

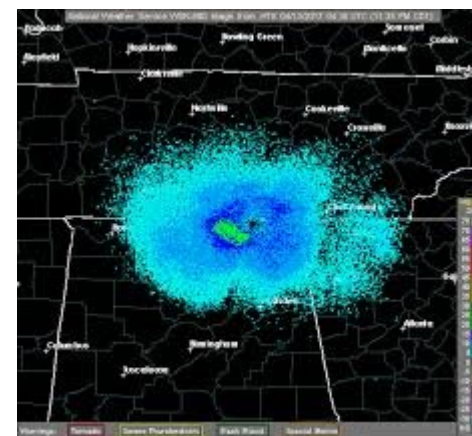
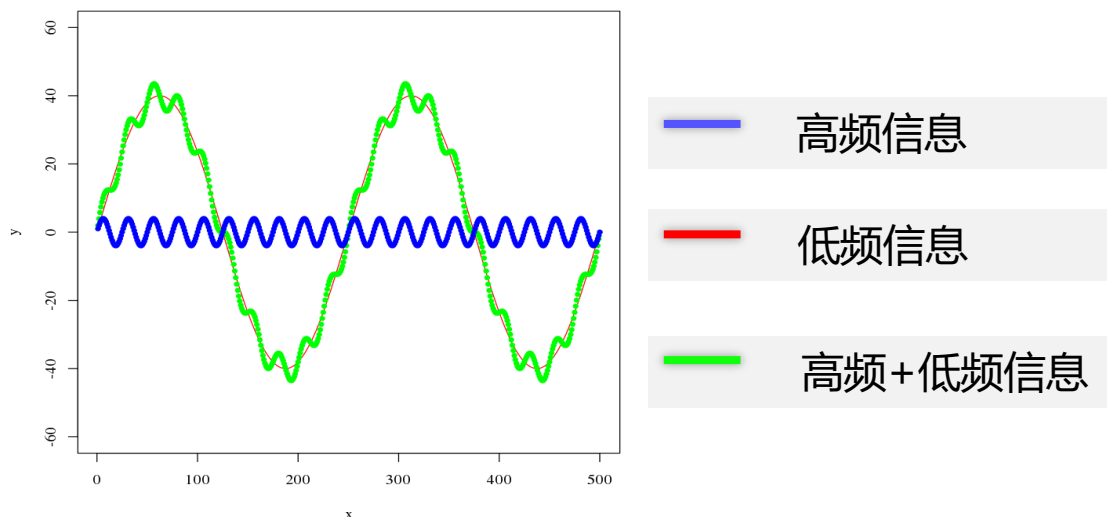
● 观测的优点和不足

优点：

- 真实数据
- 同时包含**高频**和**低频**信息

不足：

- 局部信息，不同时空尺度，或非直接信息
- 仅利用观测数据进行预测的时效较短



局地天气示意图

- 局部信息
- 只有当前和历史信息
- 只能提供短期预测
- (≤ 4 hours)

二. 融合计算的模式

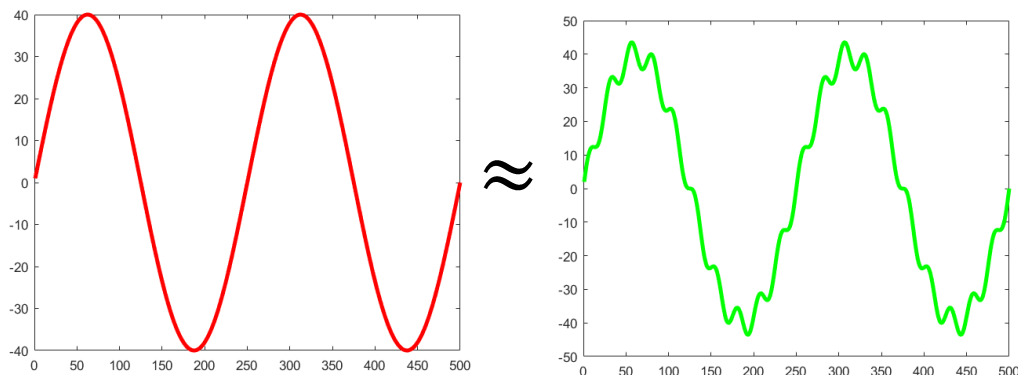


观测和模式如何更深度的融合？

● 模式的优点和不足

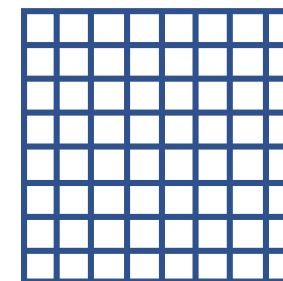
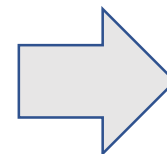
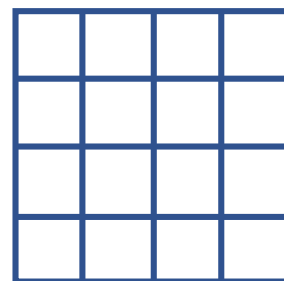
优点：

- 大尺度数据
- 能够很好地捕捉到较长时间(15天)预报的低频信息



不足：

- 极端天气、湍流等仍然无法解决
- 加密网格可能不会显著提高预报结果
- 时效性、精细化要求不易满足



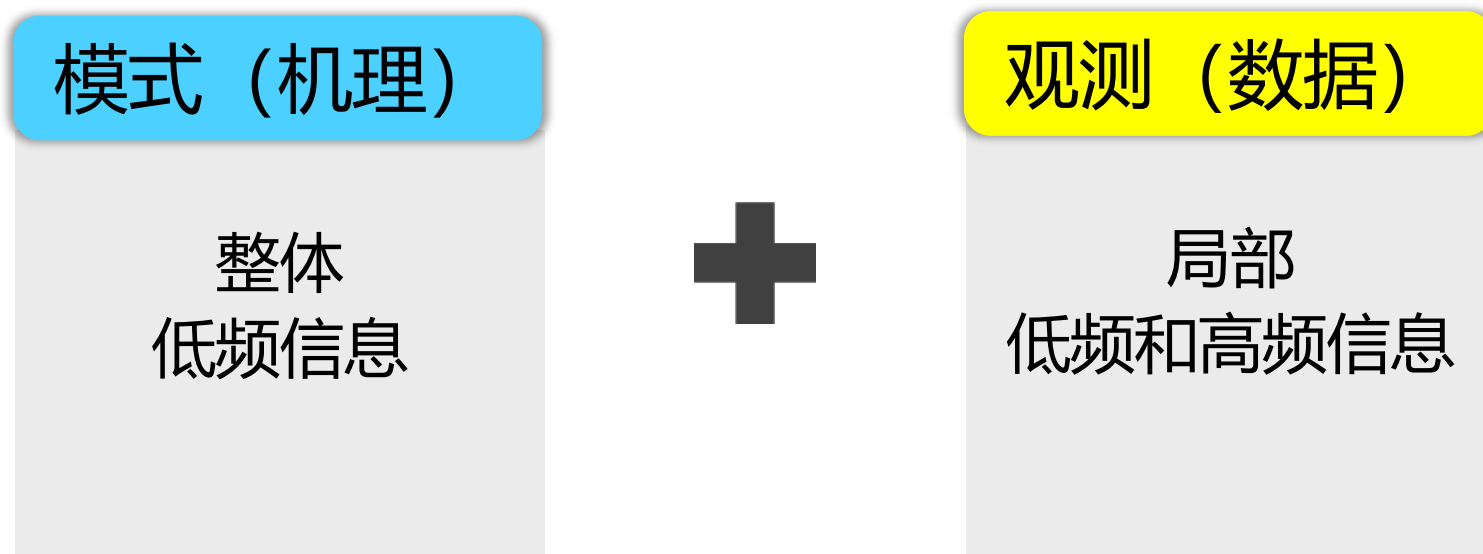
直接加密网格并不能得到更加精准的预报结果

二. 融合计算的模式



模式与观测的深度融合

- 充分发挥**不同机构**各自的优势
- 机理和数据融合的机器学习算法



二. 融合计算的模式



案例：面向冬奥的精准天气预报

- 冬奥赛会对天气预报要求极高



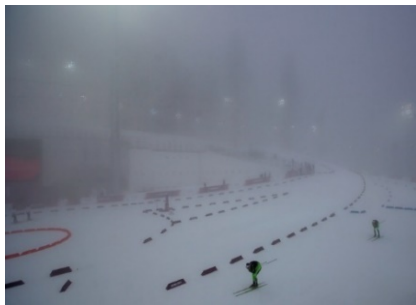
造雪



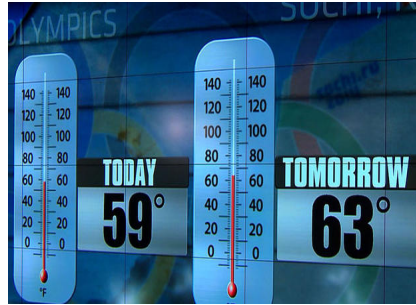
风速



交通



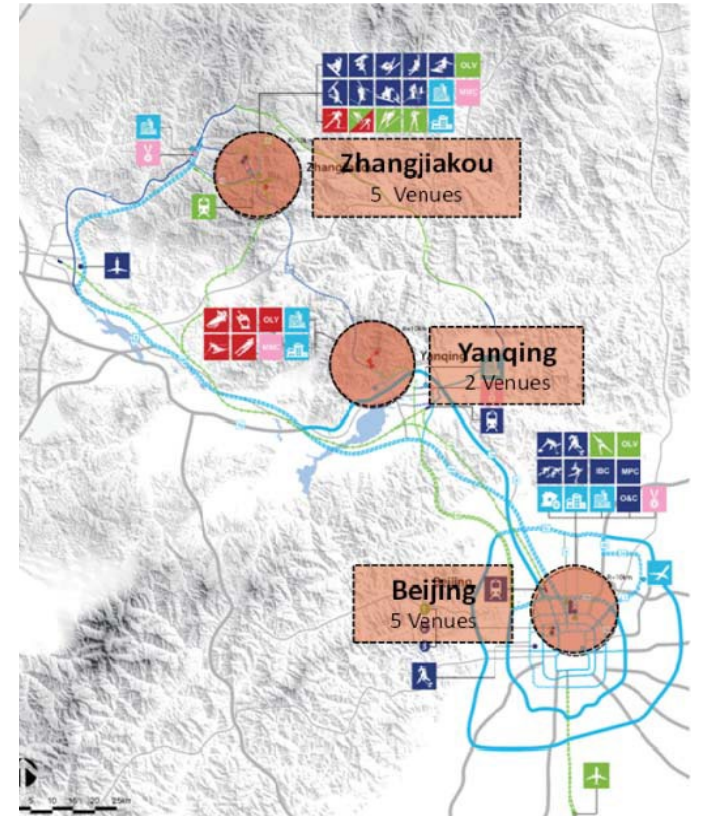
能见度



气温



降雪



赛区分布图

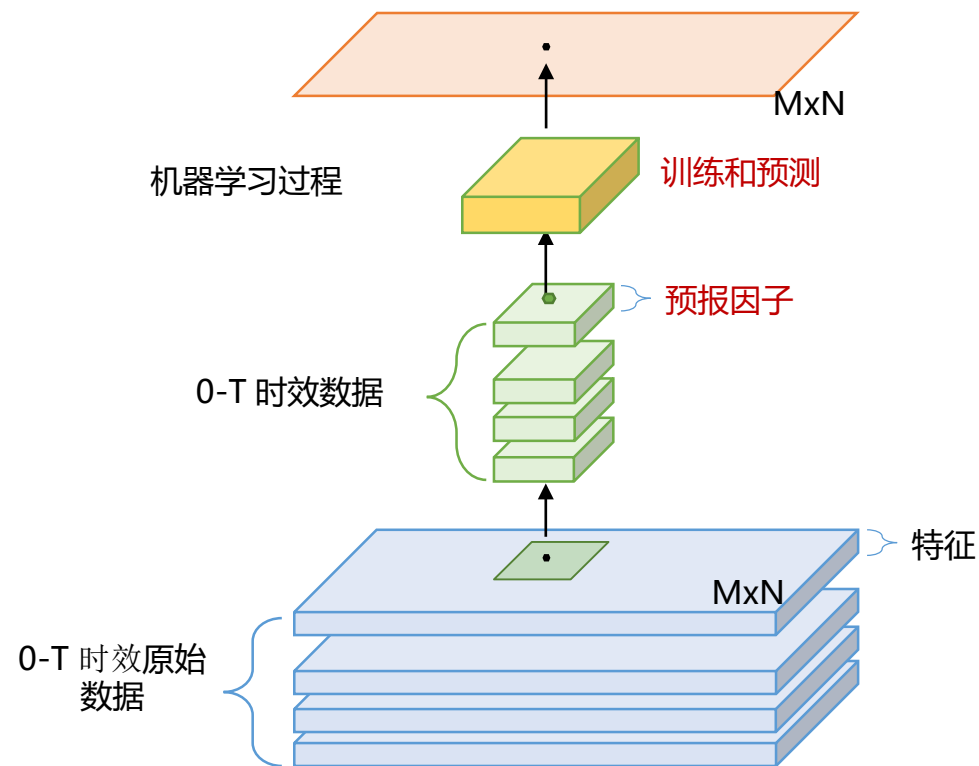
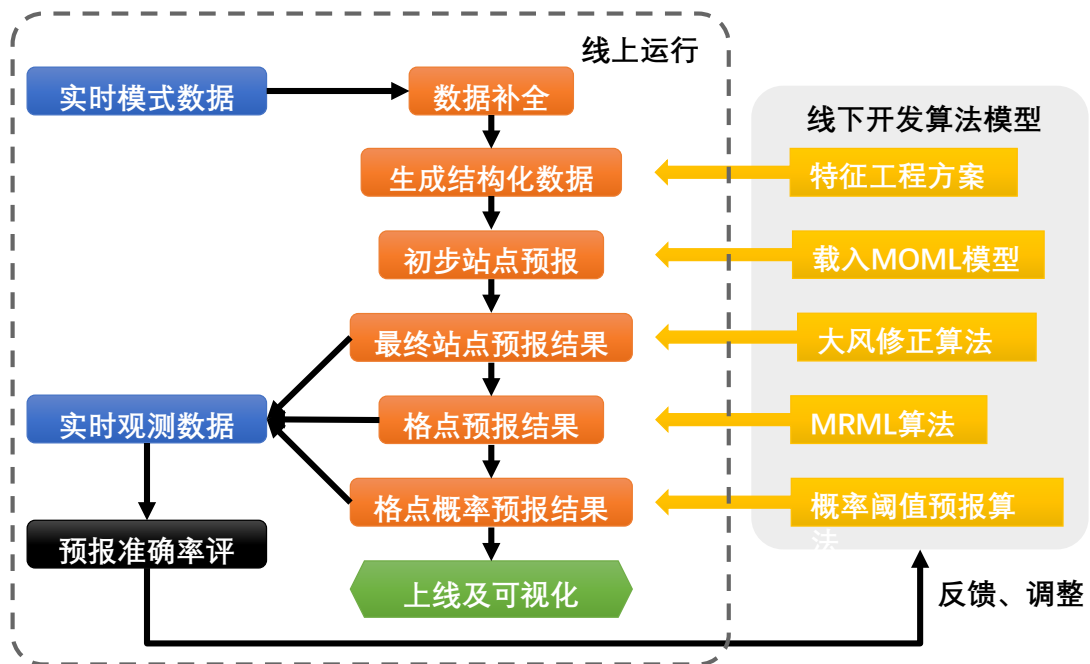
二. 融合计算的模式



案例：面向冬奥的精准天气预报

- 天气预报机器学习MOML算法

- 模式数据：大尺度、低频、可预测
- 观测数据：小尺度、高/低频，不可预测



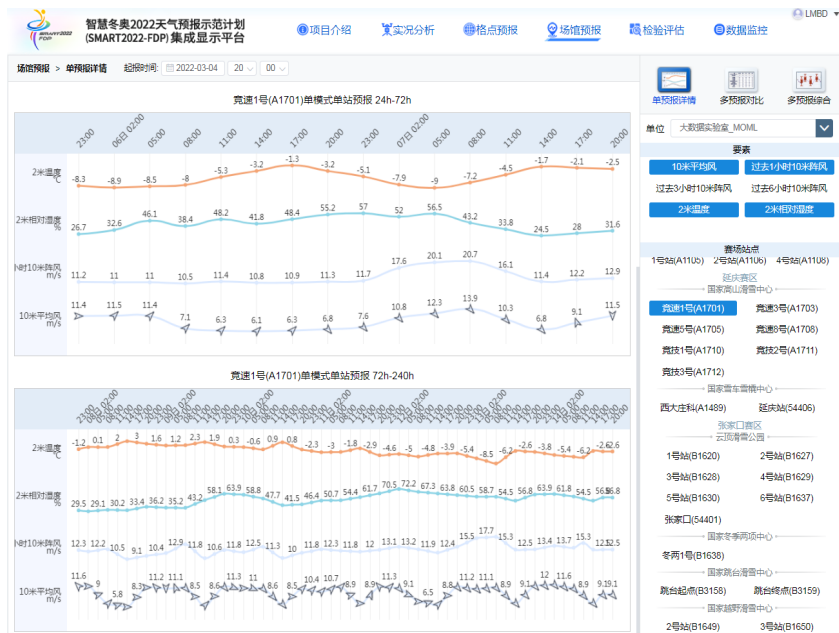
天气预报机器学习MOML算法流程图

二. 融合计算的模式

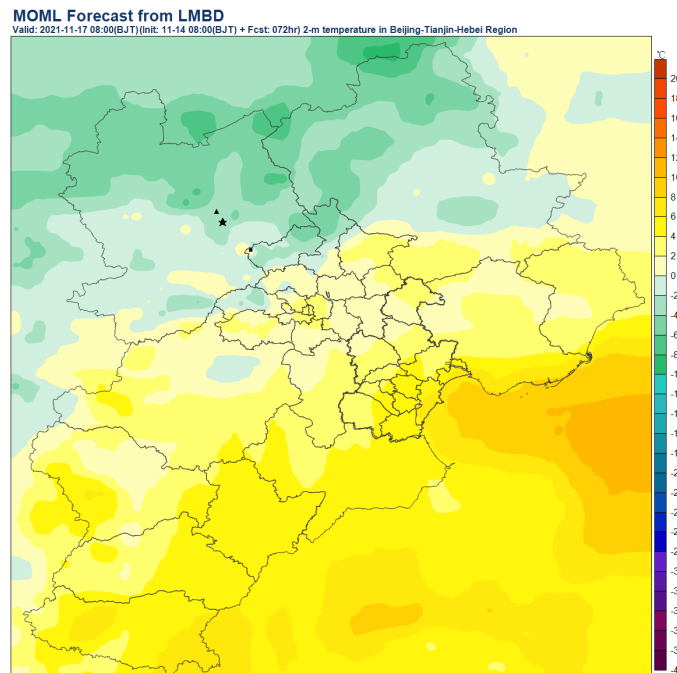


案例：面向冬奥的精准天气预报

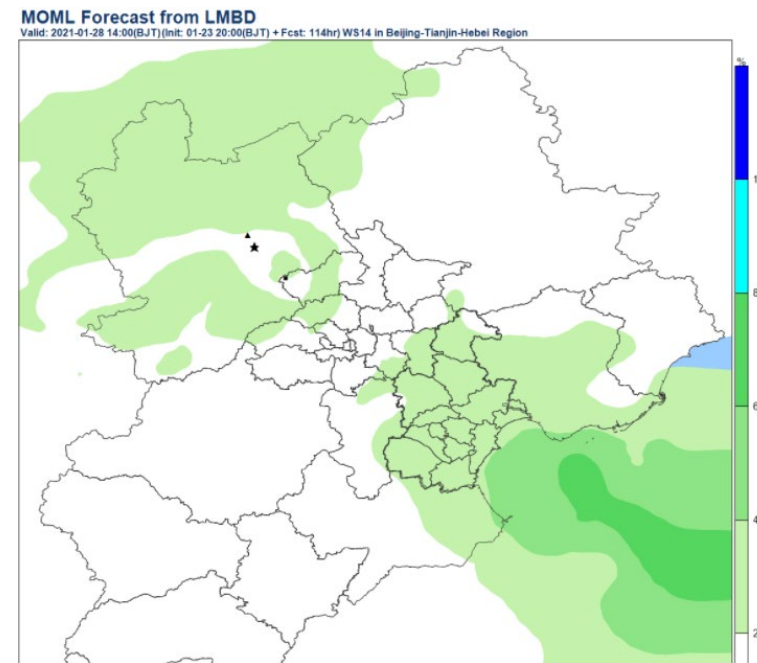
- 模式输出机器学习MOML算法冬奥产品展示



冬奥站点预报



格点预报(温度)



概率阈值预报(风速大于14m/s概率)

二. 融合计算的模式

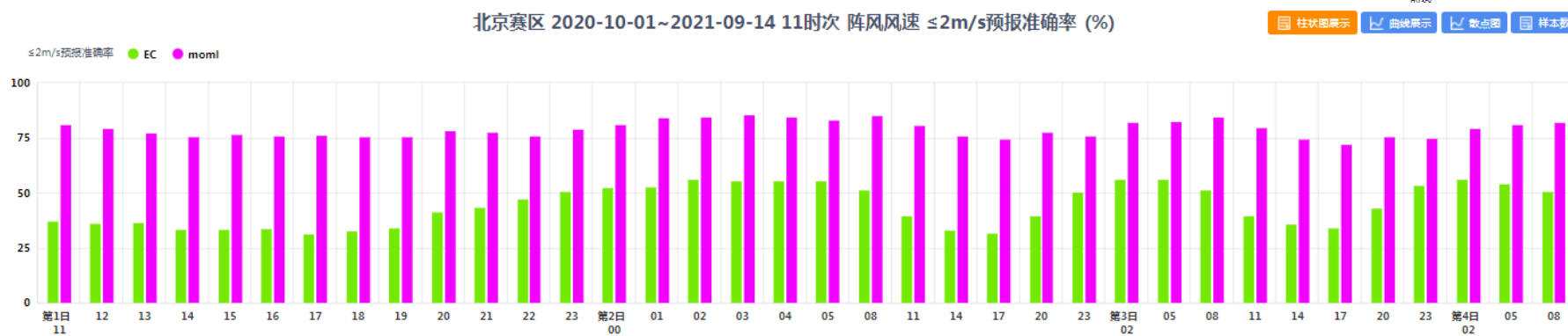
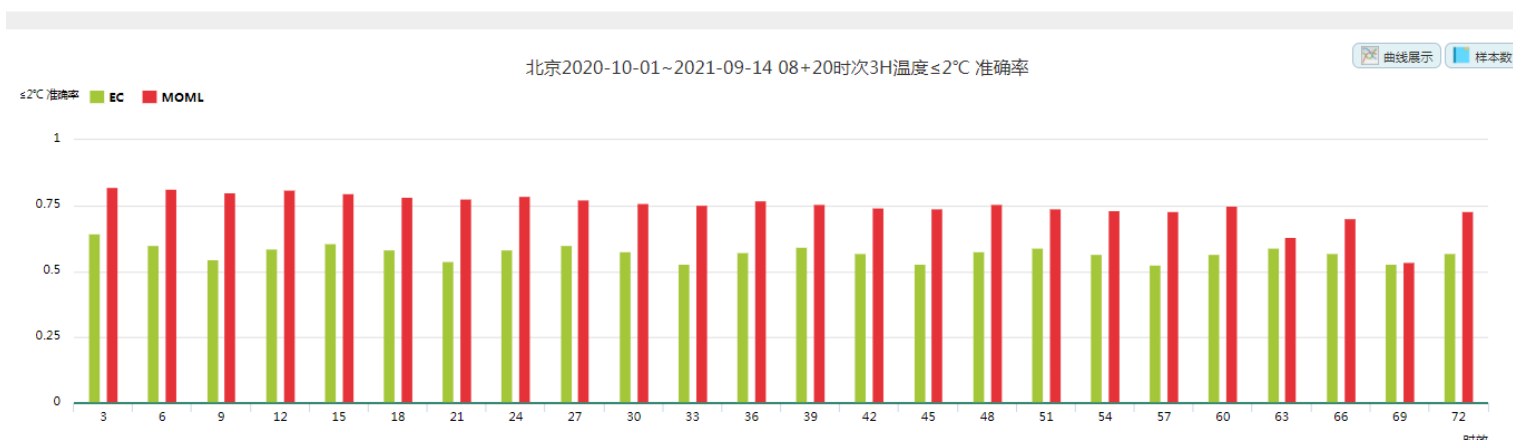


案例：面向冬奥的精准天气预报

- 气象台业务平台MOML算法过去一年结果展示

类: 图表展示 常用 起始时间: 2020-10-01 确定 区域选择: 北京 时次选择: 08 20 08+20
型: 地区分布 精确 结束时间: 2021-09-14 模式选择: 模式选择 站点类型: 城市检验 乡镇检验 考核站 单站检验

综合评分 逐日评分



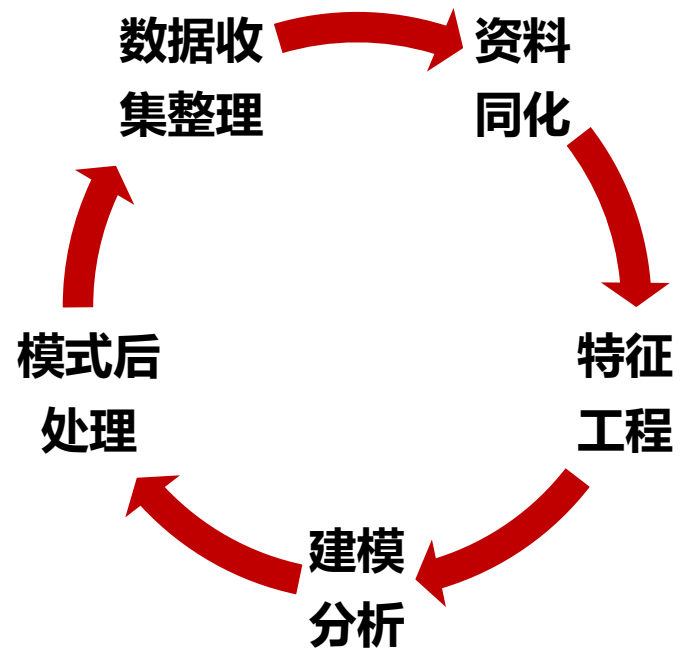
二. 融合计算的模式



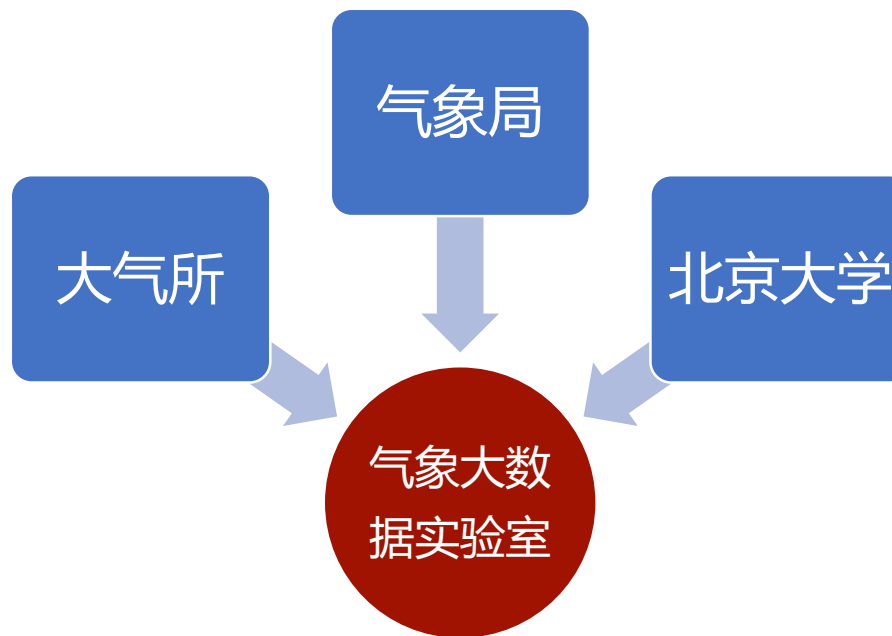
案例：面向冬奥的精准天气预报

- 小结

融合计算已渗入冬奥天气预报的各个环节



融合计算需要多方配合，共同参与



二. 融合计算的模式



融合计算的未来发展

- 可解释、可通用的下一代人工智能方法
 - 数据驱动与知识驱动融合的人工智能

关于发布可解释、可通用的下一代人工智能方法重大研究计划2022年度项目指南的通告

日期 2022-05-17 来源: 作者: 【大 中 小】 【打印】 【关闭】

国科金发计〔2022〕18号

国家自然科学基金委员会现发布可解释、可通用的下一代人工智能方法重大研究计划2022年度项目指南, 请申请人及依托单位按项目指南中所述的要求和注意事项申请。

国家自然科学基金委员会

2022年5月16日

国家基金委发布重大研究计划

科学目标

- 面向以深度学习为代表的人工智能方法鲁棒性差、可解释性差、对数据的强依赖等基础科学问题, 挖掘机器学习的基本原理, 发展可解释、可通用的下一代人工智能方法, 并推动人工智能方法在科学领域的创新应用

目录

- 一. 应用数学的演进
- 二. 融合计算的模式
- 三. AI for Social Science**
- 四. 新的增长点
- 五. 体制机制

科技
创新
在北大



三. AI for Social Science



Social Science: 人文与社科

- 人文科学
 - **以人为本**的学科，包括文学、历史学、哲学、艺术学等
- 社会科学
 - 研究各种**社会现象**的科学，包括政治学、经济学、社会学、法学等



三. AI for Social Science



- 人文社科的**科研范式**
- 算法(人工智能)的**伦理**
- **科技智库**

人文社科的科研范式

- 统计软件
- 结构模型
- 基于数据的计算
- **未来**: 机理与数据的融合计算

主要数据类型

- 记录数据
- 调查数据
- 观测数据

结构模型

- 经济学

宏观经济学

结构经济学

微观经济学



- 社会学

非常重视结构

◆ 社会科学的很多方法
都来自社会学研究

三. AI for Social Science



基于数据的计算

- **搜索算法**颠覆了图书情报学



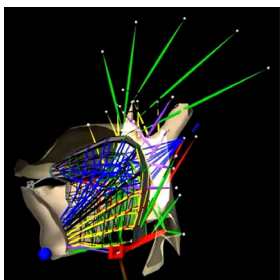
- **推荐算法**颠覆了新闻传播学



头条、抖音等新媒体

机理与数据的融合计算

- 示例：人类语言的进化研究



- 肌肉模型

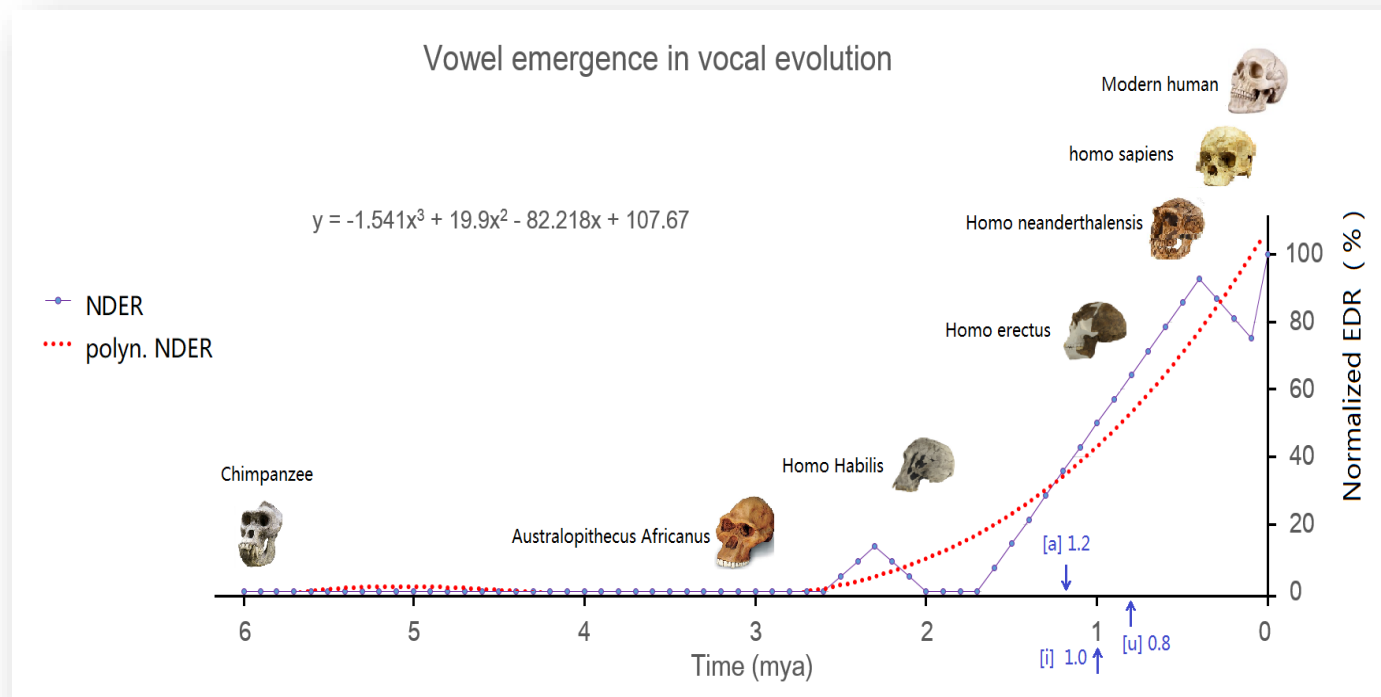


- 口腔模型

- 其他模型



- 语音数据
- 考古数据
- 其他数据



三. AI for Social Science

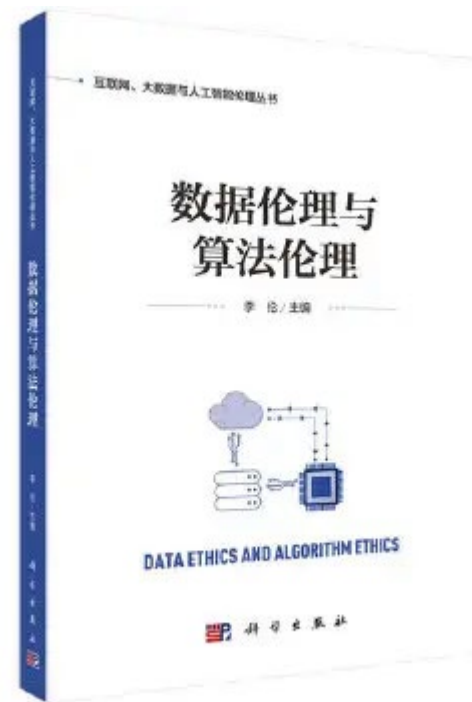


算法(人工智能)的伦理

- 算法(人工智能)应**以人为本**



算法的困扰



科技智库

- **智库**研究是人文社科的重要方向

数字化具有极强的穿透力



数字战争



数字教育

激发数据要素活力



科技智库研究变得
愈发重要

数字化发展状况度量

- 生态的视角：**数字生态**

北京大学团队2019年开始数字生态研究
并连续2年发布数字生态指数

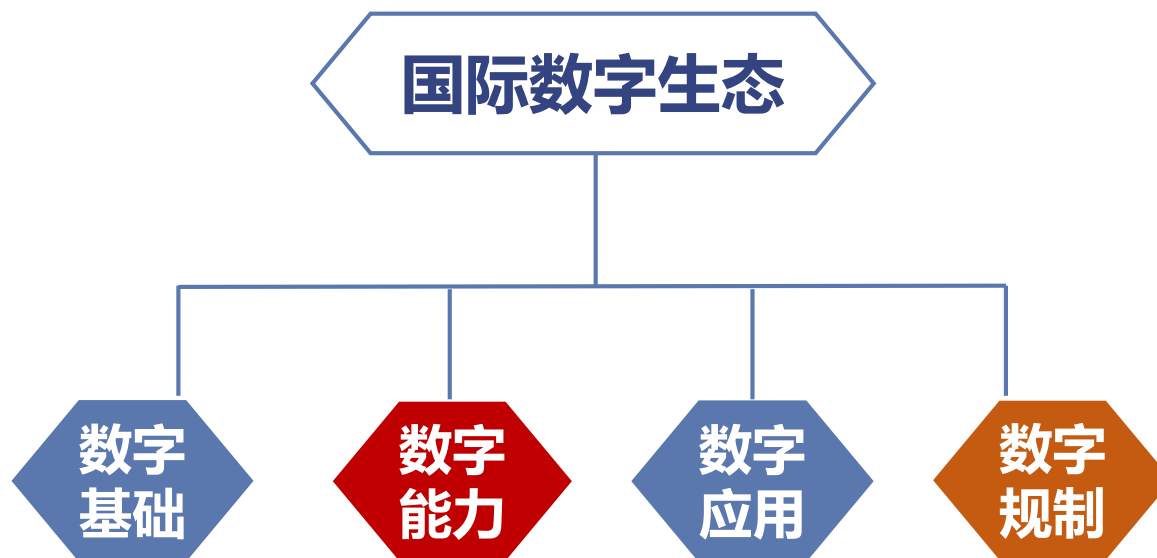
国际数字生态研究已取得阶段性成果



数字生态指数2020



数字生态指数2021

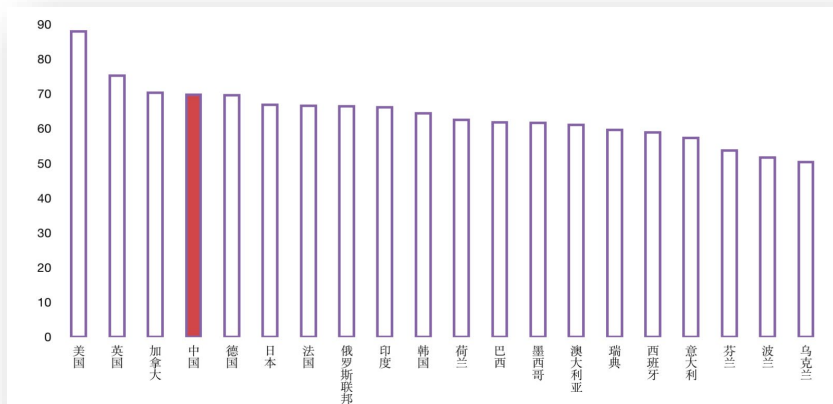


通过四个维度刻画数字化发展状况

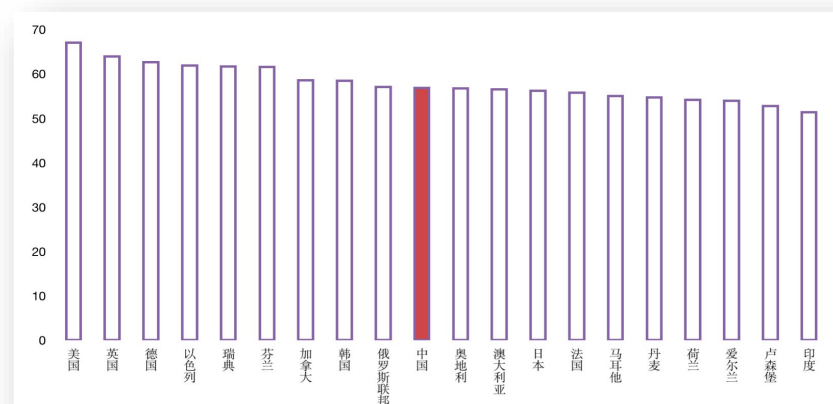
三. AI for Social Science



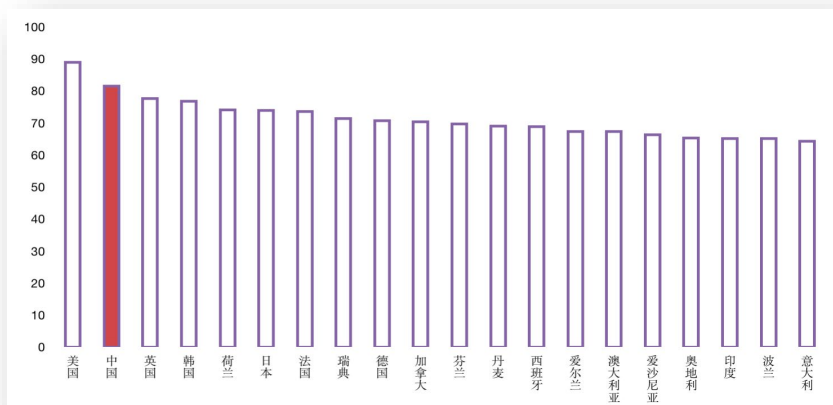
分维度排名中的中国位置



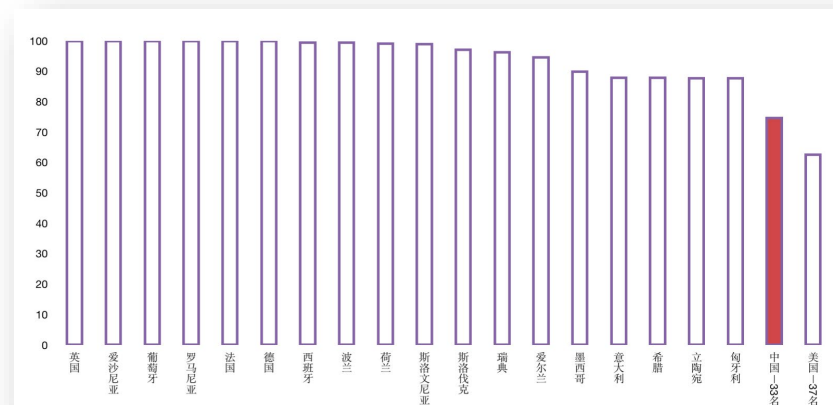
数字基础有较强优势



数字能力赶超空间大



数字应用仅次于美国



数字规制完备性较低

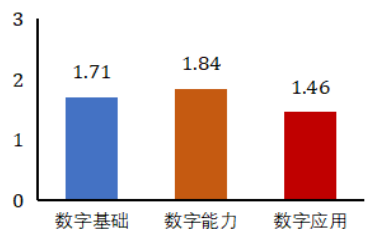
三. AI for Social Science



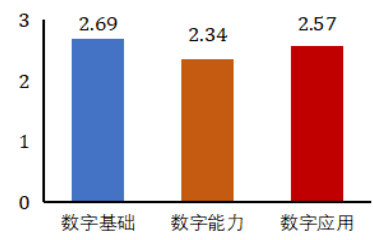
中美欧数字生态驱动模式

全面发展

英国

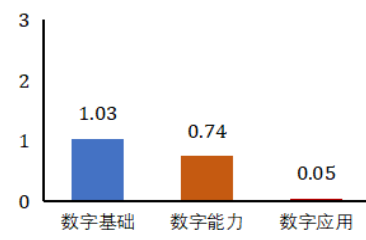


美国



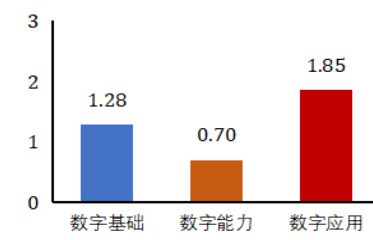
基础引领

俄罗斯



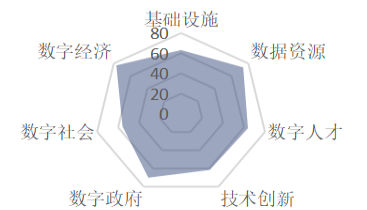
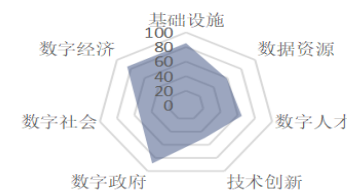
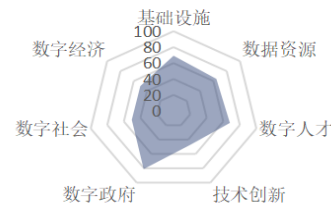
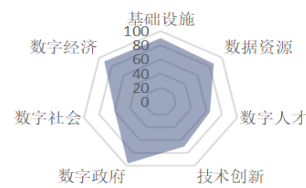
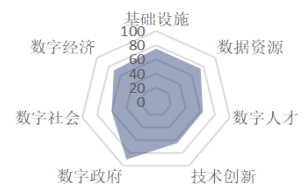
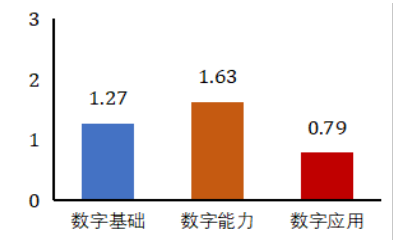
应用引领

中国



能力引领

德国



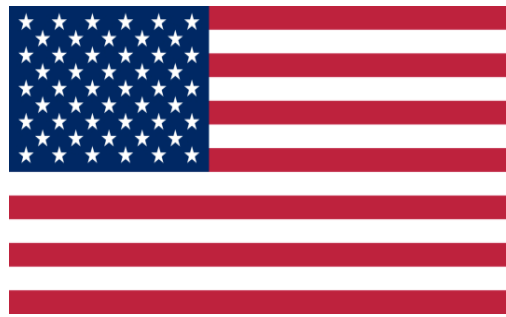
三. AI for Social Science



数字规制与融合计算



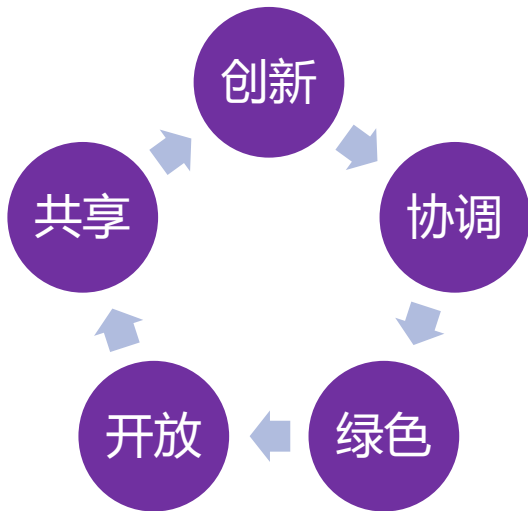
中国规制目标集中
突出**安全有序**



美国规制目标趋于多样
强调**市场创新**



欧盟规制目标相对多元
侧重**市场秩序**



科技智库
特点

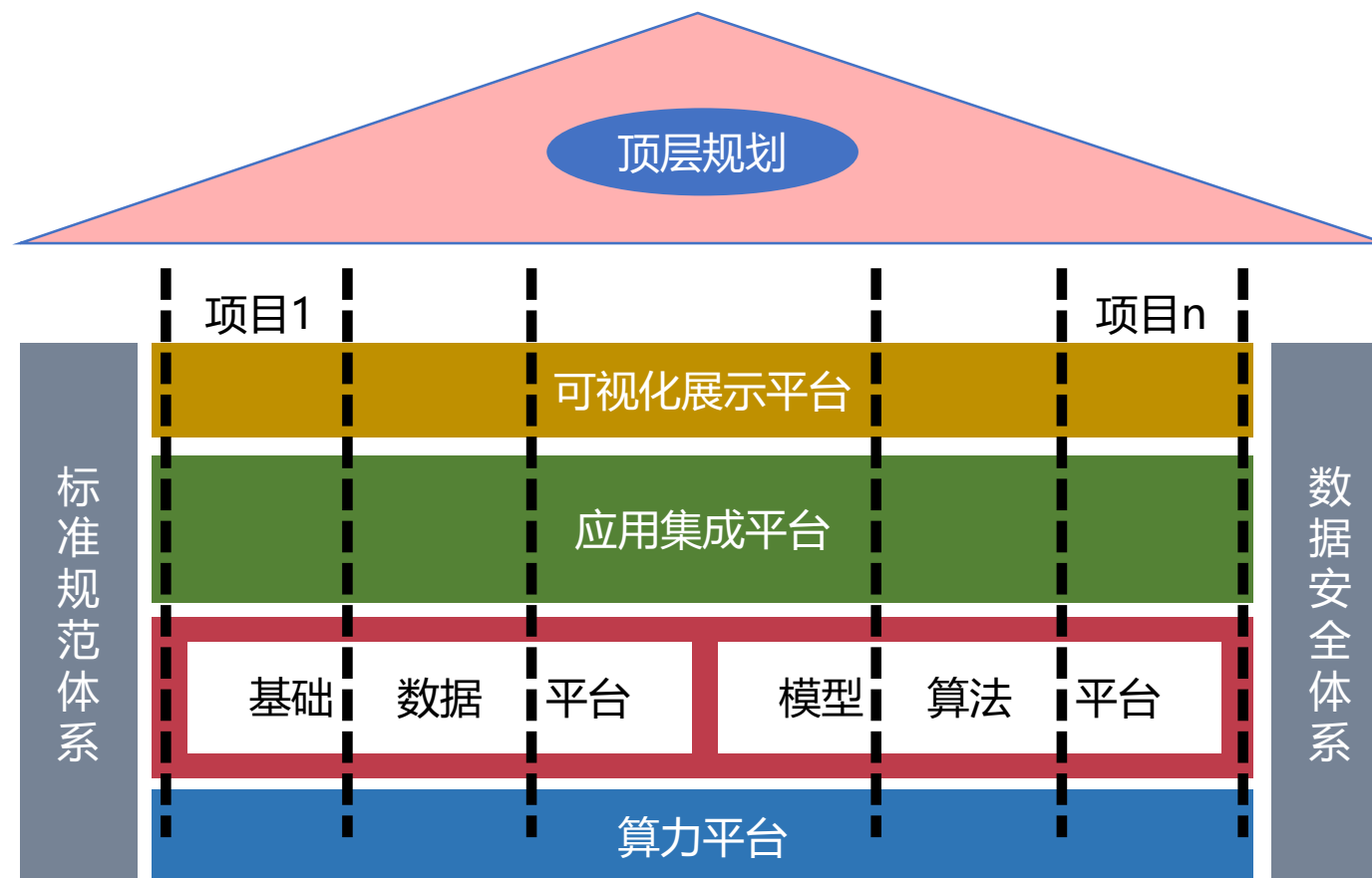
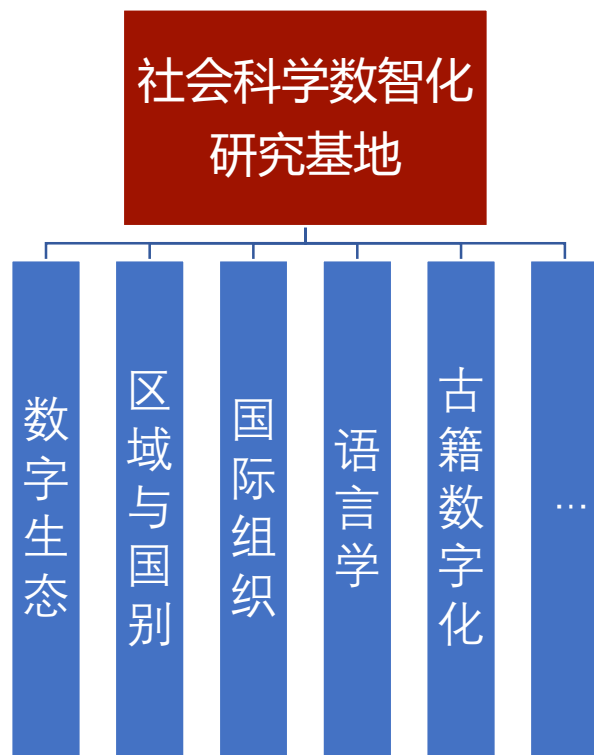
- 场景复杂
- 涉及面广
- 海量数据
- 机理缺乏

需要引入融合计算

北京大学社科数智化建设

- ◆ **北京大学大数据分析与应用技术**
国家工程实验室**牵头**

- 2022: 北京大学 **“数字与人文”** 年



目录

- 一. 应用数学的演进
- 二. 融合计算的模式
- 三. AI for Social Science
- 四. 新的增长点**
- 五. 体制机制

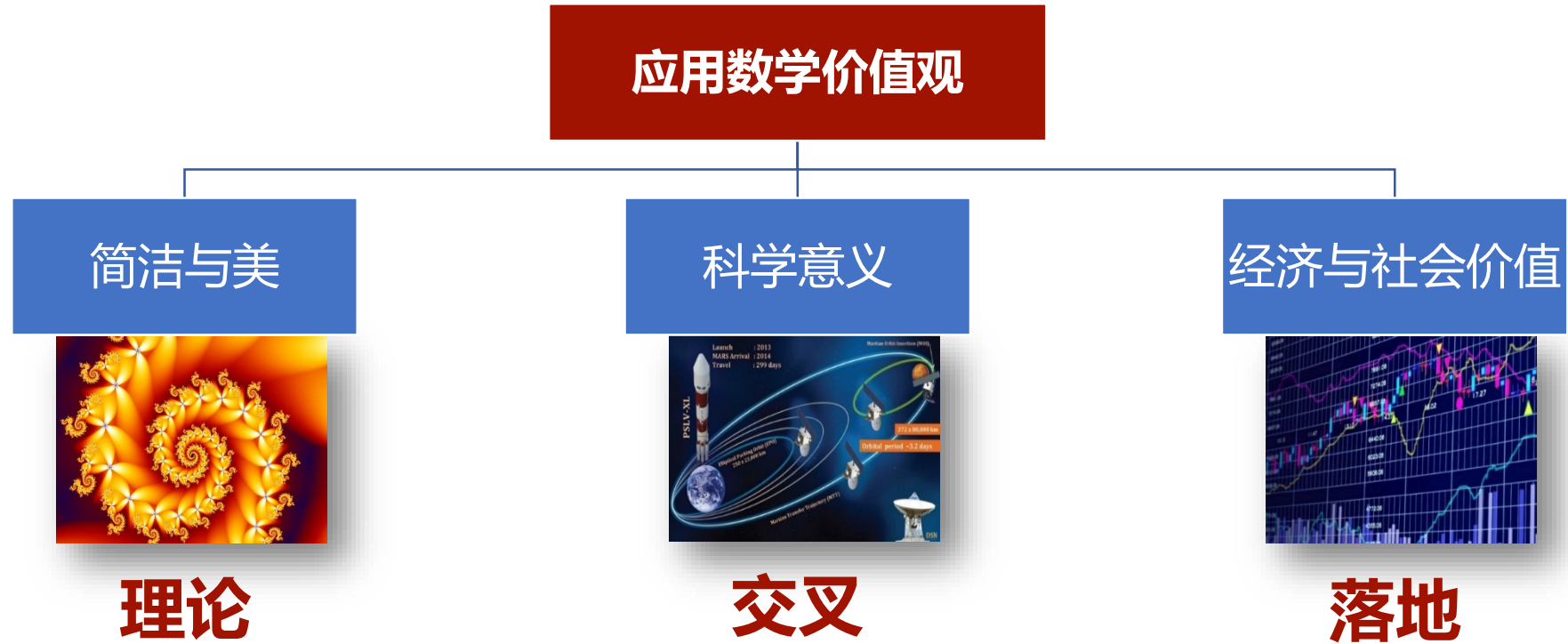
科技
创新
在北大



四. 新的增长点



应用数学价值观



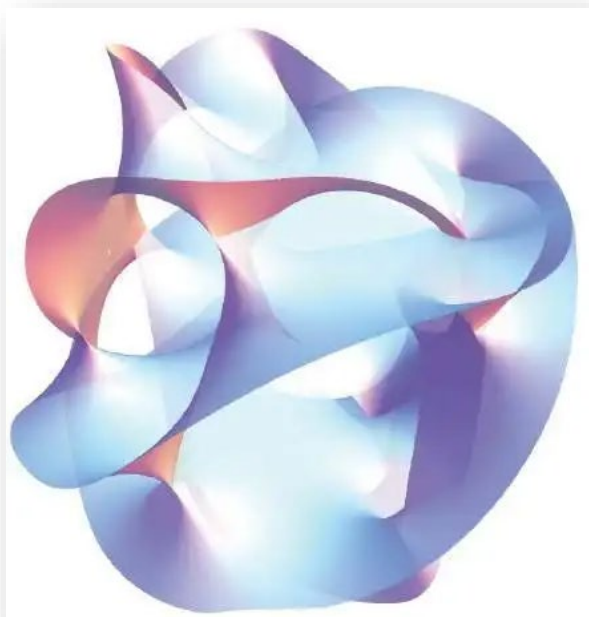
- 融合计算为应用数学带来什么新的增长点?

四. 新的增长点



理论方面

- 新的**研究领域**



- **高维空间中的数值分析**



- **融合计算的模式**

四. 新的增长点

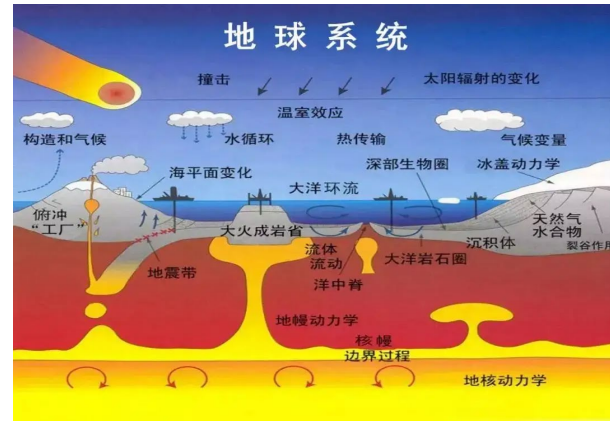


交叉方面

- 极大**拓展**了应用场景



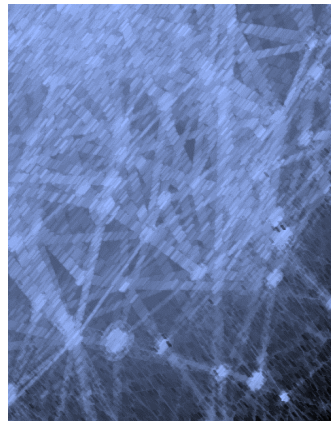
社会科学



地球系统



复杂系统



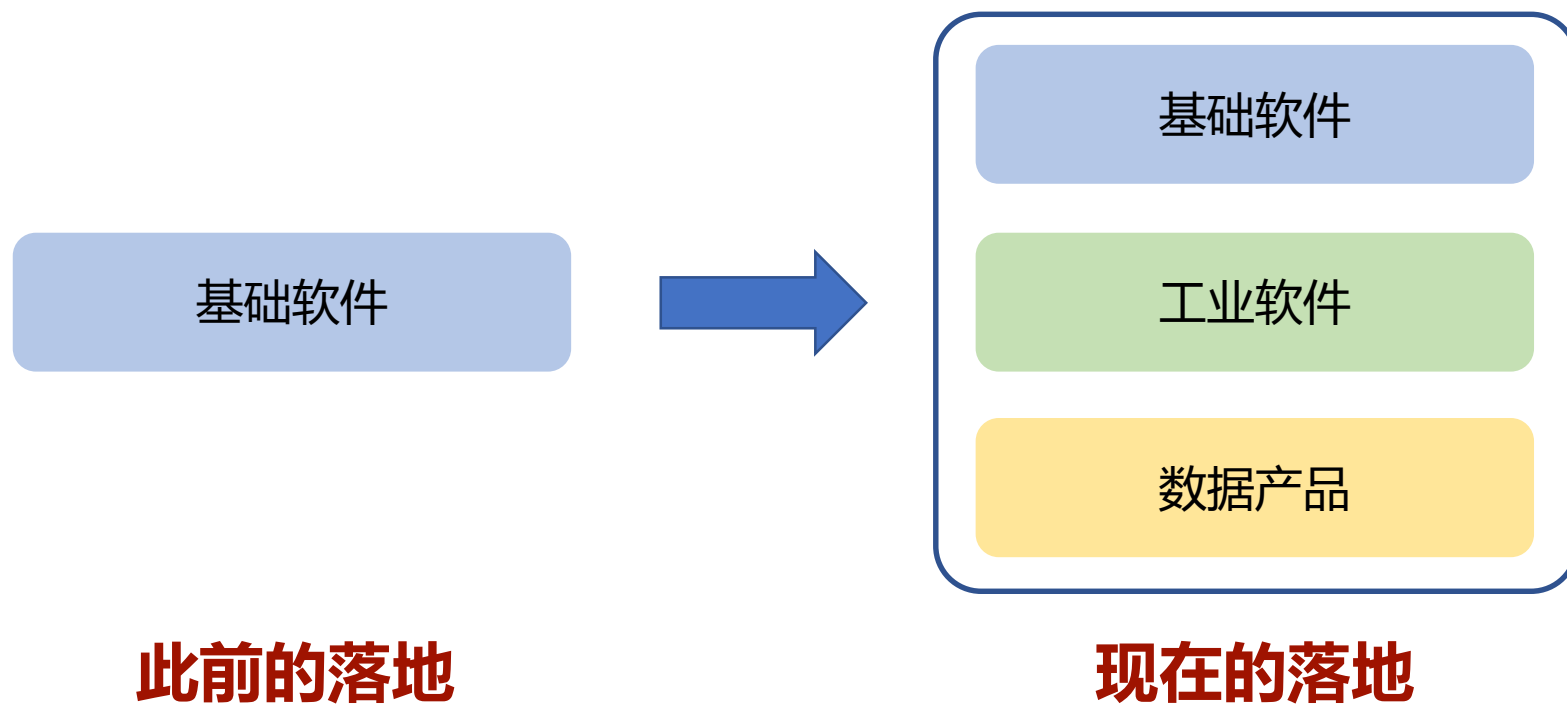
...

四. 新的增长点



落地方面

- 扩充了应用数学**落地**方式



目录

- 一. 应用数学的演进
- 二. 融合计算的模式
- 三. AI for Social Science
- 四. 新的增长点
- 五. 体制机制**

科技
创新
在北大

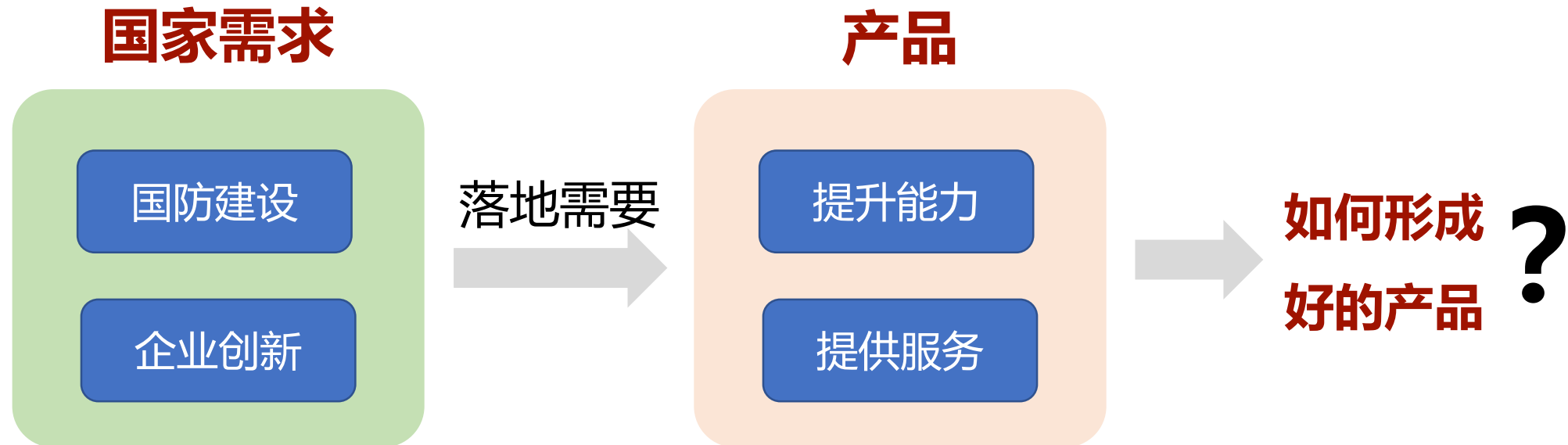


五. 体制机制



什么是应用数学的落地?

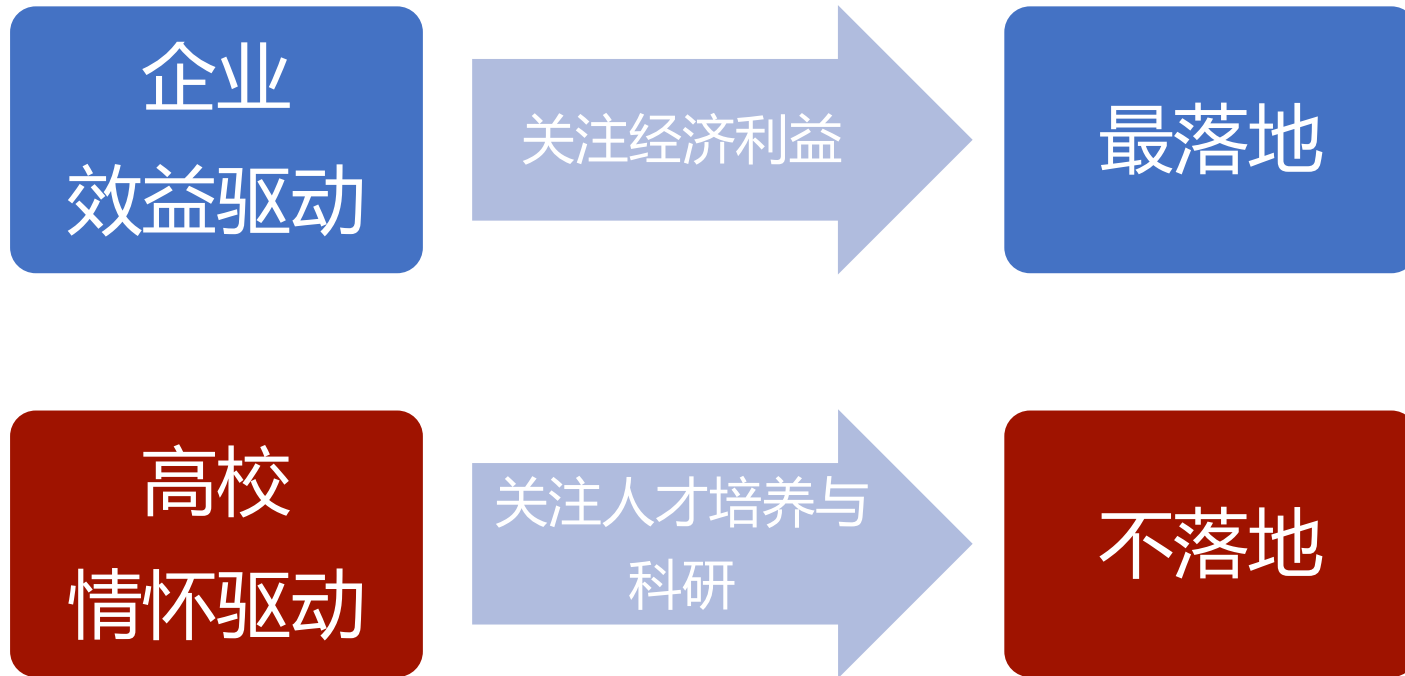
- **国家需求**驱动应用数学发展



五. 体制机制



不同机构的关注点



二者的联接很难

五. 体制机制



北大的探索

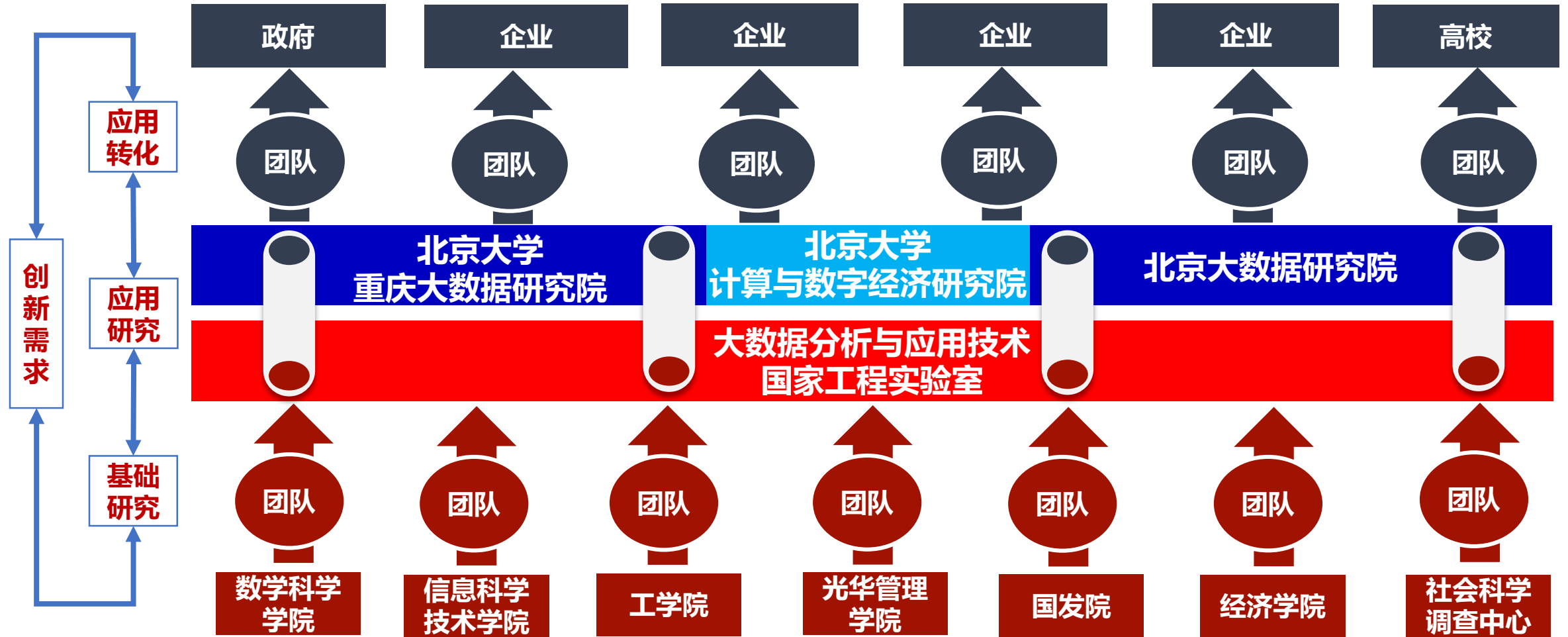
学术创新链：从基础研究到“四个面向”的高效连接和有机闭环



五. 体制机制



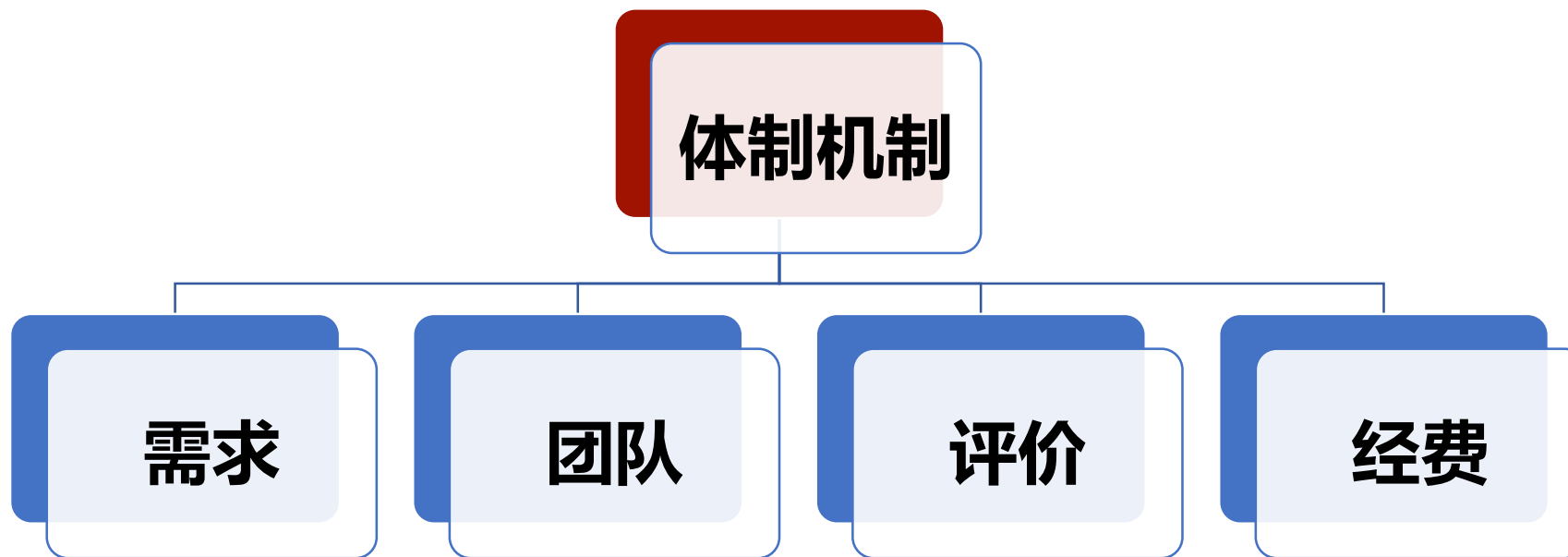
北大的探索



五. 体制机制



体制机制的4个主要问题

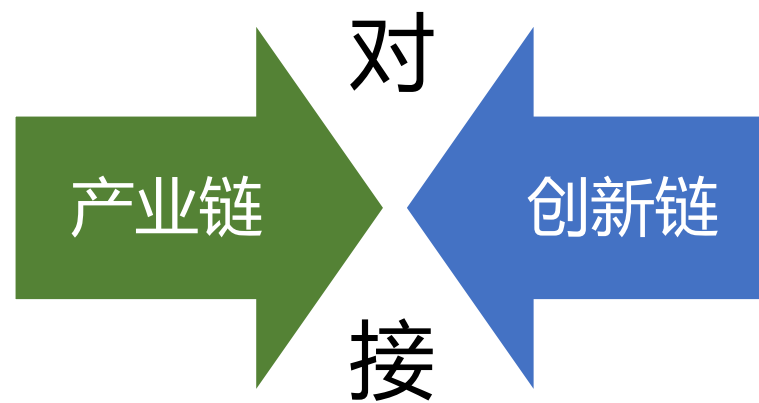
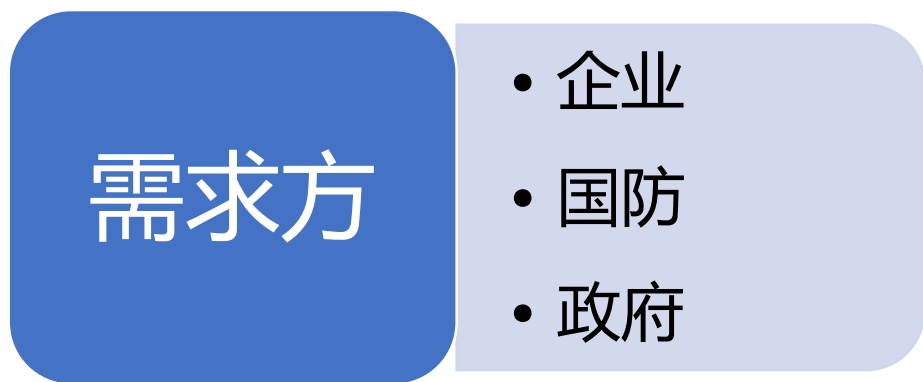


五. 体制机制



需求

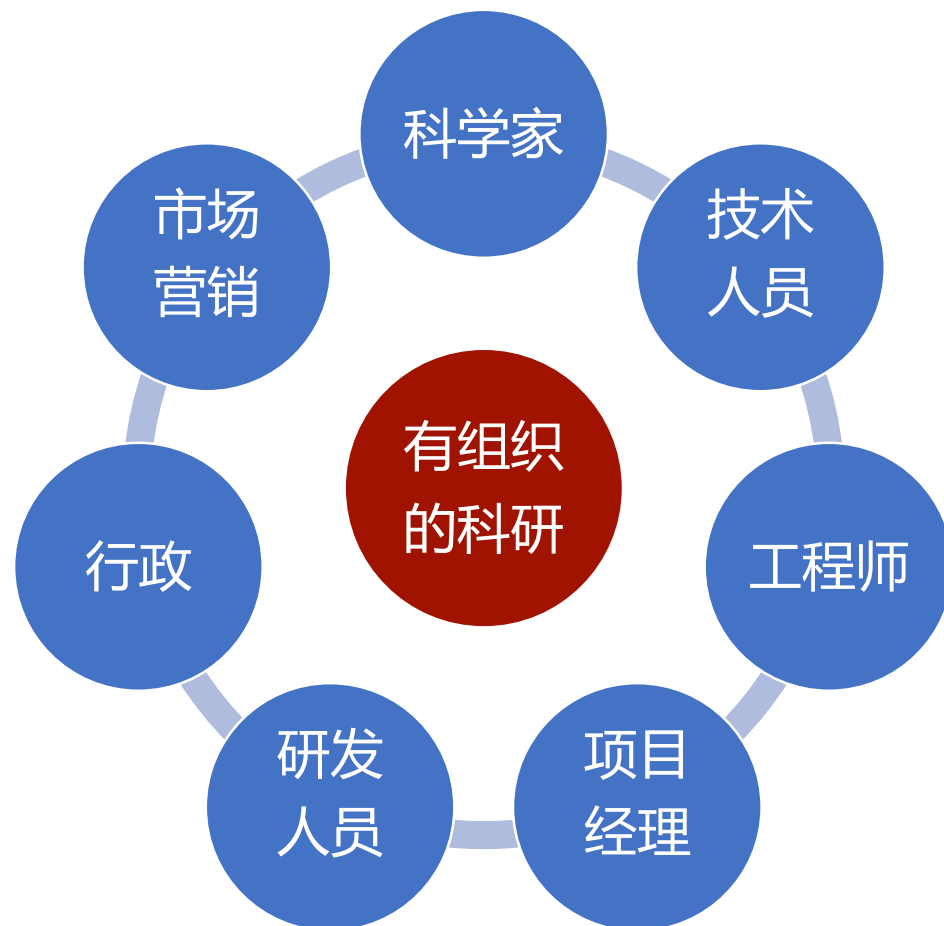
- 需求即**场景**
- 需求在哪里？



团队

- 开展有组织的科研

1. 人员构成



2. 机构组成

五. 体制机制



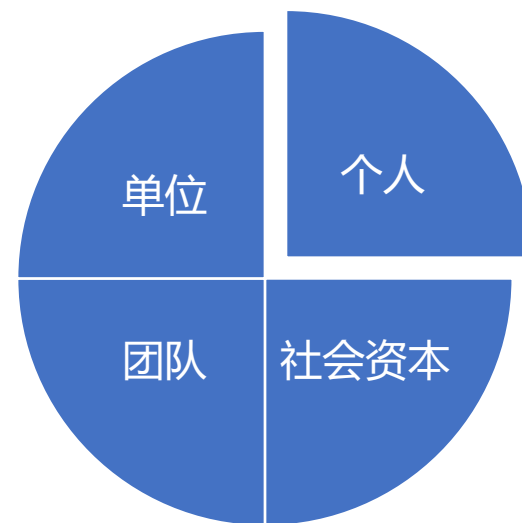
评价

- 不同人员的评价标准不同
- 根本上是解决**知识产权与利益分配**的问题

人员类型

- 科学研究人员
- 科技开发人员
- 落地应用人员

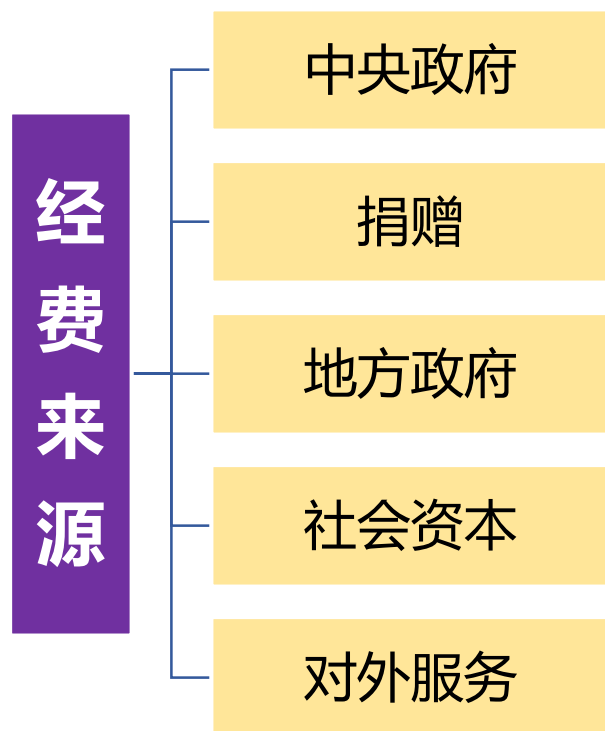
落地人员关注其落地能力
主要靠**知识产权**激励



落地阶段的利益分配

经费

- 经费关系到机构的可持续发展

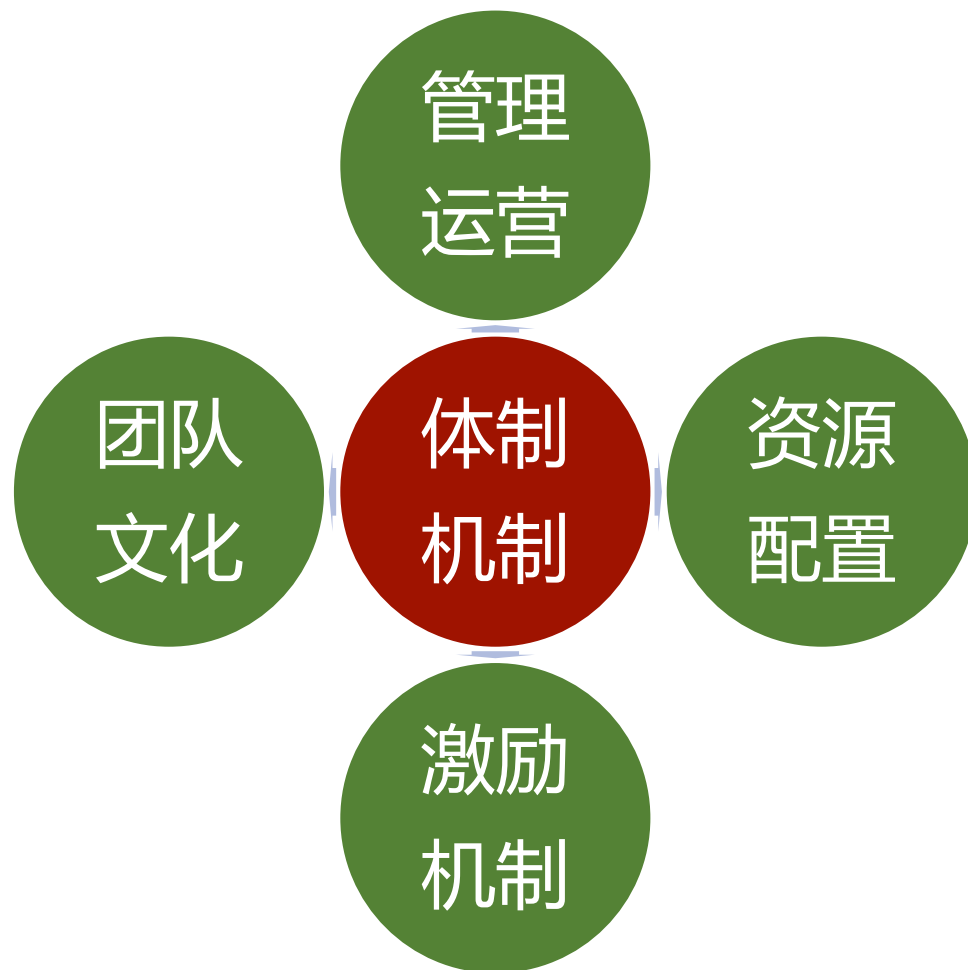


不同机构的**定位**与其经费来源？

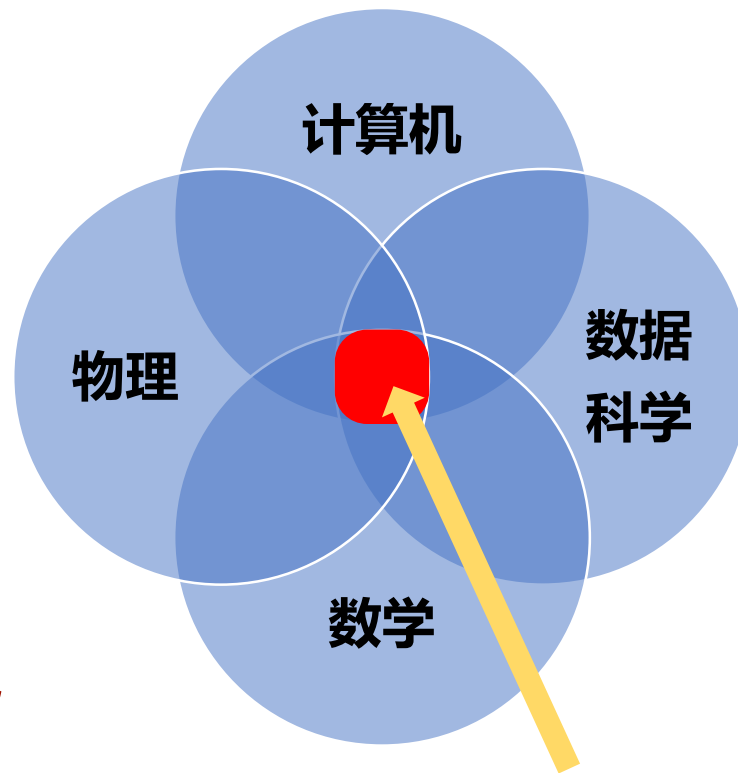
五. 体制机制



小结



- 应用数学学科正在**走向成熟**
- **融合计算**方兴未艾
- AI for **Social Science**
- 融合计算极大促进了应用数学**落地**
- **体制机制**是应用数学落地的重要保障



- ◆ 应用数学学科的核心部分
- ◆ 也是交叉学科的基础



谢谢各位!