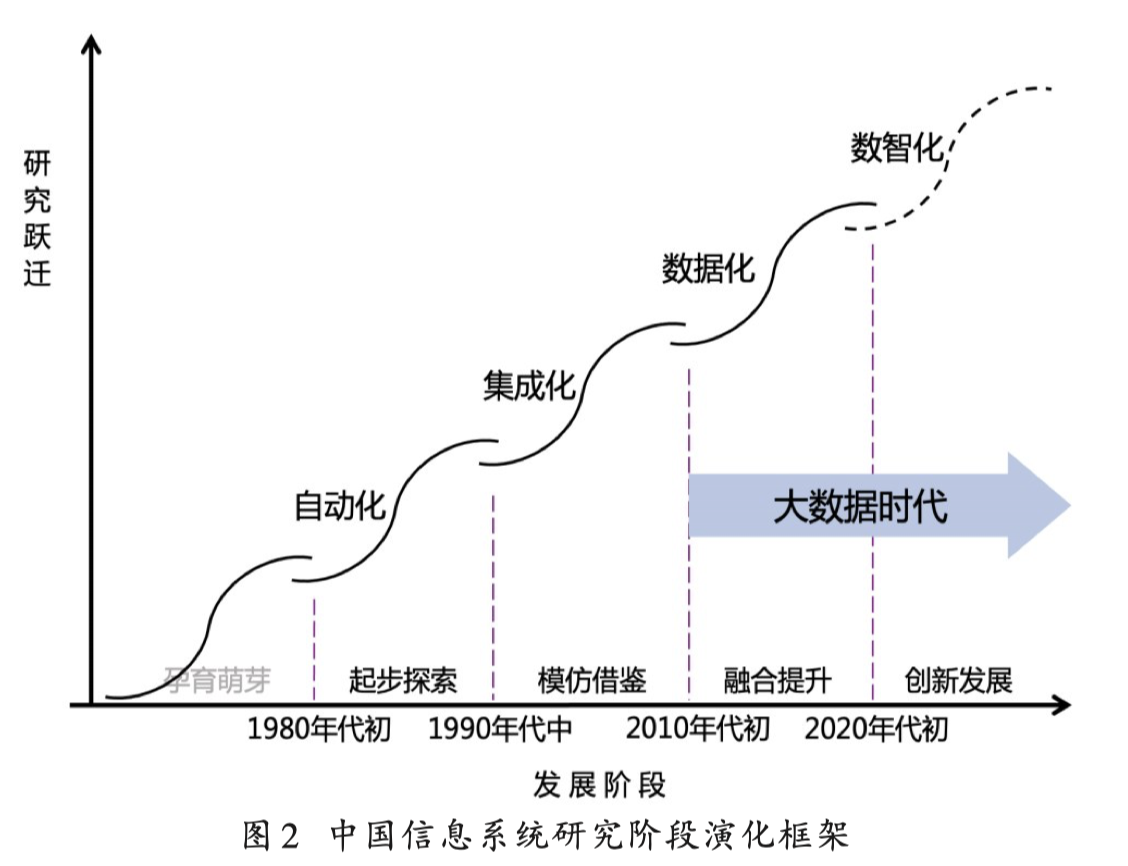
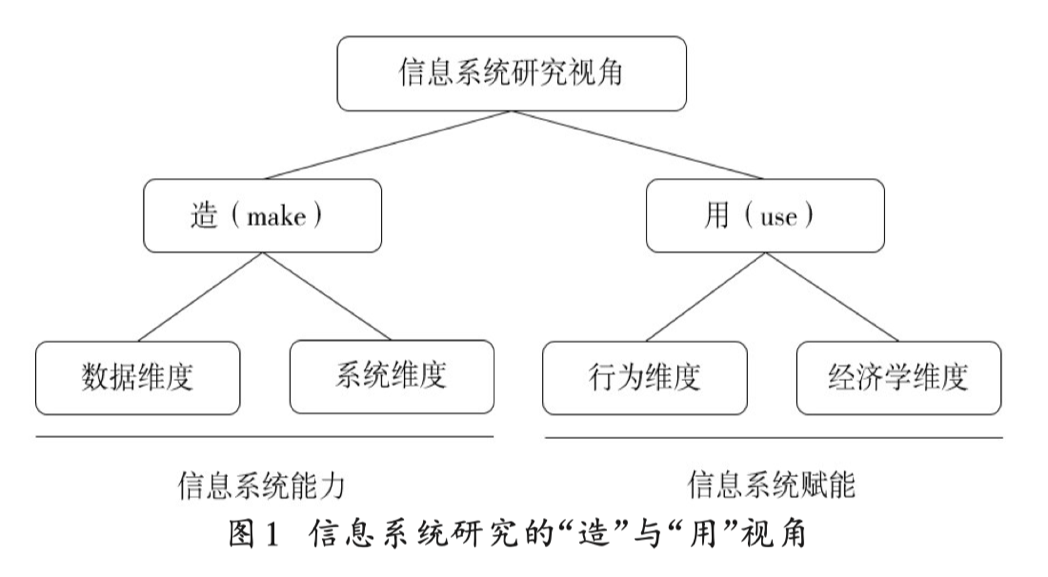
1. 中国信息系统研究的发展历程

数据化->数智化



2. 信息系统研究的“造”与“用”



3. 大数据的特征

4V：volume, velocity, variety, value

体量大——规模大

流数据——数据连续不断产生

多样性——数据形式多样，富媒体

低价值密度——数据中的信息密度低

4. 大数据问题的特征

粒度缩放：问题要素的数据化，并且在不同层级数据化

跨界关联：打破业务或行业边界，对要素约束和领域视角加以扩展，实现外部视角与传统视角的联动和协同，比如将内部数据与社会媒体内容等外部数据相关联，发掘创造价值的机会。

全局视图：指问题定义与求解具有全局性。 这需要基于数据分析和平台集成的全景式“成像”能力，不再局限于单一或局部的角度，而是从更广泛的业务视角 生成问题的整体画像及其动态演化。

举例：大数据医疗

5. 大数据决策

关键属性“关联+因果”。从管理决策问题来看，管理信息系统可以分为三类——事务性、分析型和预测型，分别着重“发生了什么”，“为什么发生”和“将发生什么”。大数据决策意味着：

* 面对“发生了什么”，业务层要清楚业务状态，数据层要有相关数据粒度，决策层要有全景决策视图；
* 面对“为什么发生”，业务层要具备业务联系，数据层要具备数据链接，决策层要有因果对应；
* 面对“将发生什么“，业务层要有业务动态，数据层要有数据演化，决策层要形成前瞻预测；

6. 大数据决策范式的转变

* 跨域转变：许多管理决策问题从领域内部扩展至跨域环境，公众以及其他决策相关者的信息纳入考量。 这些跨域信息的补充使决策要素的测量更完善可靠，进而提升管理决策的准确性。

例子：财务管理决策。不仅使用企业内部的信息，还可以利用企业外部社交媒体、新闻报道、用户口碑等信息。

* 主体转变：决策主体的转变，一些决策受众可能成为决策主体（如：众包），而大量计算机智能系统的应用，使得决策主体转变为人或组织与智能系统的结合。

例子：人力资源领域的智能招聘

* 假设转变：一些通过假设来概括和简化不可测不可获现象的情形也受到冲击， 进而影响着基于这些假设而研发出来的许多经典决策模型。

例子：传统上假设顾客到达 模式（如泊松分布）以及站点服务模式（如指数分布）的一类决策模型可能需要重新审视，因为现在根据数字轨 迹感测到的到达和服务模式（如网络购物等）不一定满足模型的假设，进而决策模型的表达和求解都可能发生变化。

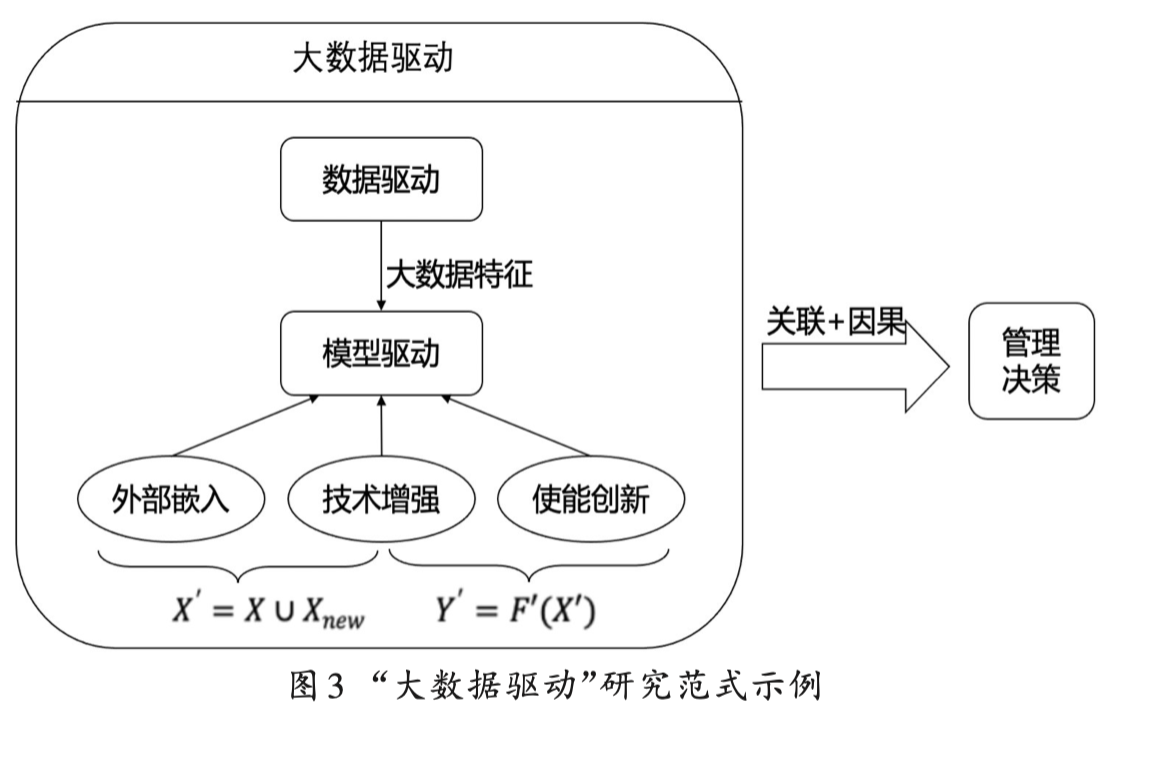
* 流程转变：传统的决策理论中的决策过程通常是近线性的、分阶段的，但是现在的很多决策场 景呈现出决策阶段/环节之间的复杂交错互动的非线性模式。 （信息的实时捕捉与反馈）

例子：在营销管理决策中，传统的营销漏斗理论 将用户转化路径分为“意识—考虑—购买—忠诚—宣传”，并采用“吸引—转化—销售—保留—联系”的营销策略。 然而，在网络购物场景中，可以发现消费者在营销漏斗的各个阶段间的转换率和转换方向具有高度的概 率随机性，而非线性步骤。、

7. “大数据驱动“研究范式

信息系统研究的传统主流范式是“模型驱动”和“数据驱动”。 “大数据驱动”范式，其是一类“数据驱动+模型驱动”的融合范式，面向大数据 和管理决策情境，体现“关联+因果”的诉求。三个要素特征：

* 外部嵌入：外部嵌入是指通过引入外部视角，将大数据提供的、传统模型视角之外的一些新的重要变量（包括构念、因素等）引入到模型的变量集合中。
* 技术增强：指通过机器学习等新型技术方法和工具将引入的新变量（如多模态富媒体变量）融入变量集合，并支 持新型变量关系的构建。
* 使能创新是指通过构建新型变量关系，形成大数据驱动的价值创造。



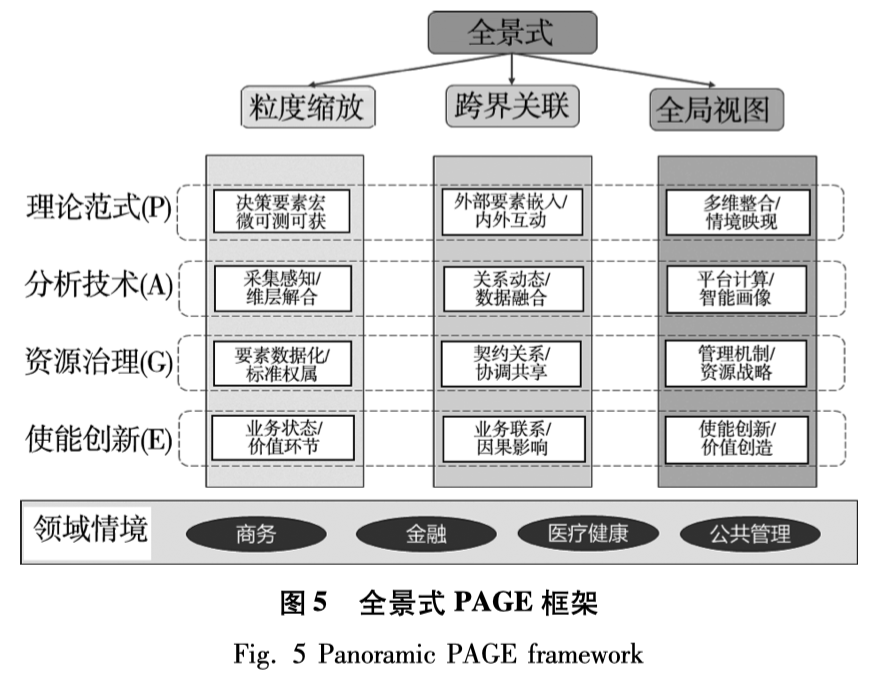
8. 智能方法分析与人机融合行为

9. 数智化跃迁

* 数据层面：尽管“像素”扩张提升将继续 到发展，但是，随着学界、业界、社会大众对于数据安全、隐私保护、数据权属、共享机制、数据要素市场等问题 的日益关切，数据治理将成为一个主基调。“数智化”新跃迁将在数据层面通过数据治理更追求平衡发展。
* 算法层面：在“数智化”新阶段，随着“像素”扩张提升到新的高度，“成像”将成为焦点，高阶智能将成为算法层面的主基调。“数智化”新跃迁在算法层面技术方法创新更追求智能发展。
* 赋能层面：“数智化”在“数据化”的基础上，更凸显出数据层面的治理和算法层面的智能，进而深刻影响着赋能及其价值创造的过程。如：商业模式创新与重塑，新的行为理论与机理探究等。综上所述，“数智化”赋能层面将在“治—智”新形势下更追求重塑与拓新。

8. PAGE全景式框架

全景式PAGE框架是融合大数据特征和重要研究方向的要素矩阵，旨在刻画大数据驱动的“全景式”管理决策框架。



9. 大数据小数据问题

出于数据的可获性及 其成本、时间、乃至人们的认知能力、阅读心理等相关因素影响，人们面对或者能够直接处理的数据往往是有 限的、部分的。

基于小数据的决策与基于大数据的决策在效果上取决于小数据与大数据之间的信息不对称程度。 在此，我们将该问题称为“大数据—小数据”问题。

语义一致性（显式/隐式）

3种典型类型：

代表性语义：小数据集能够代表大数据中的内容

一致性语义：小数据集的特征分布一致

多样性语义：小数据集中数据结构和大数据一致

10. BS问题vs Top-K问题

Top-k 问题寻求的子集具有高偏序特征（根据排序策略涵义），而 “大数据—小数据”问题寻求的子集具有高语义相近性（根据语义反映涵义）。