如何提高文献研究的质量？

洪朝辉博士

美国福坦莫大学经济史教授

2020年12月4日

一、导论

1、学术创新vs. 学术规范

- 学术“八股”？学术“镣铐”？

2、学术质量与学术态度

- 工匠精神、农民心态

3、文献研究的目的与要求

- 一是发现已有
- 二是批判现有
- 三是填补没有

4、文献研究定义(Literature Review：文献综述、文献回顾？）

- 叙、评、补

二、发现已有文献

- 文献的**发现**：收集、阅读、选择、分类、总结

1、为何要发现？

1）避免抄袭、尽弃前功

2）提高起点、增加通透

3）厘清误区、寻根溯源

- 例一：亚当·斯密“看不见的手”
- 例二：中庸与适度哲学的鼻祖
- 例三：Affirmative Action

2、如何收集和阅读？

1）粗读教科书
- 代表性、权威性、前卫性

2）三层次搜索
- EBSCO、JSTOR

3）泛读
- 略读、概读、详读

4）精读
- 书评、多遍、不同版本
- 例四：亞當·斯密《道德情操論》六版

5）穷尽所有文献
- 多次、持续、最后搜索
- 交流vs保护

3、如何选择、分类和总结？

1）时间分类法
- 例五：欧洲中世纪黑死病研究
- 例六：社会权利贫困理论
- Historiography；History of Historical Writing

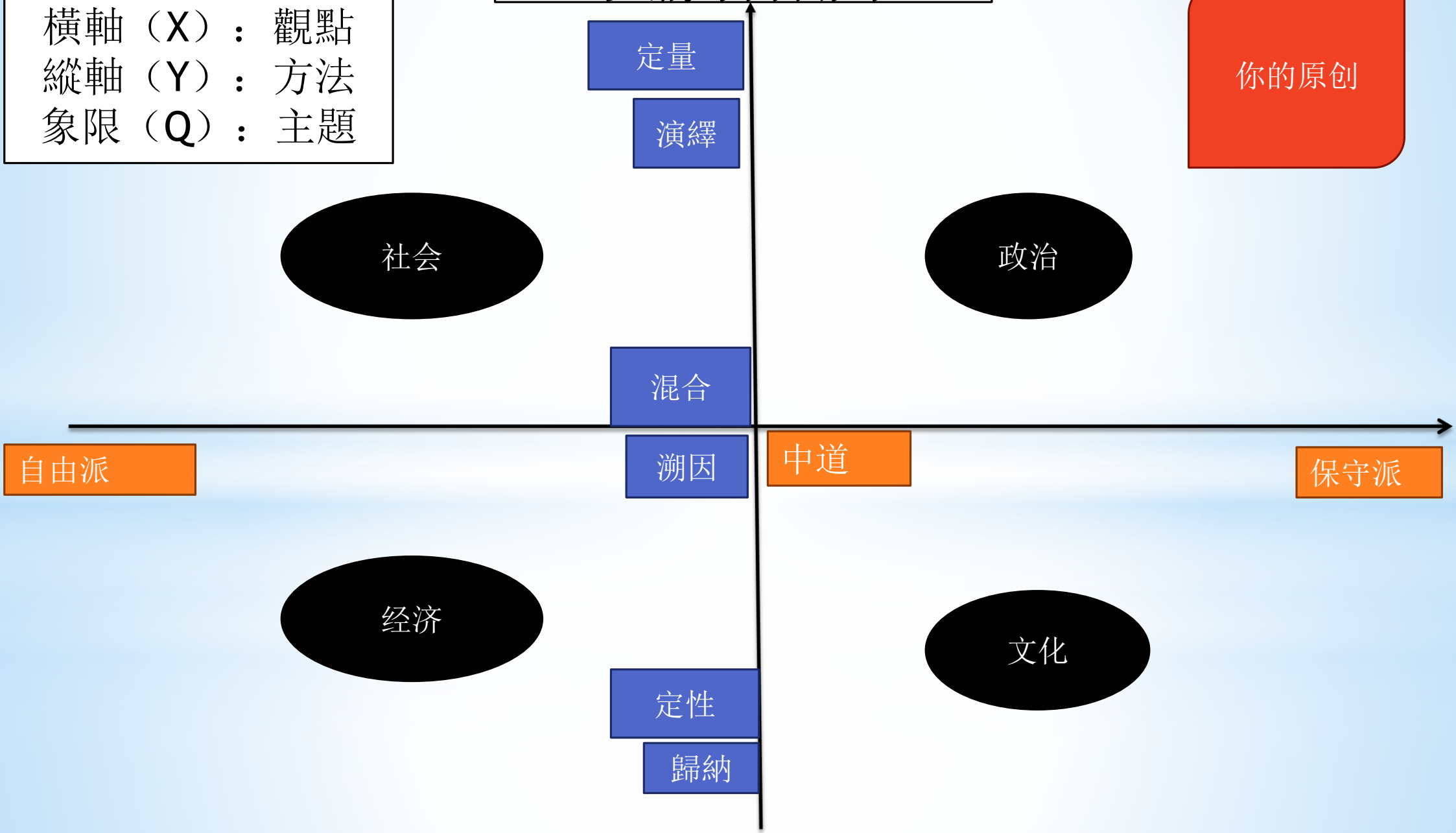
2）主题分类法：政治、经济、社会、文化
- 例七：美國口罩現象研究

3）观点分类法：自由、保守、中道
- 詮釋學（前理解）

4）方法分类法：定量、定性、混合
- 演繹、歸納、溯因（Abduction)

文獻分類圖

横軸（X）：觀點
縱軸（Y）：方法
象限（Q）：主題



4、普遍出现的问题

1）为文献而文献

2）引文不严谨、不规范

3）定义不明确

4）遗漏重要文献

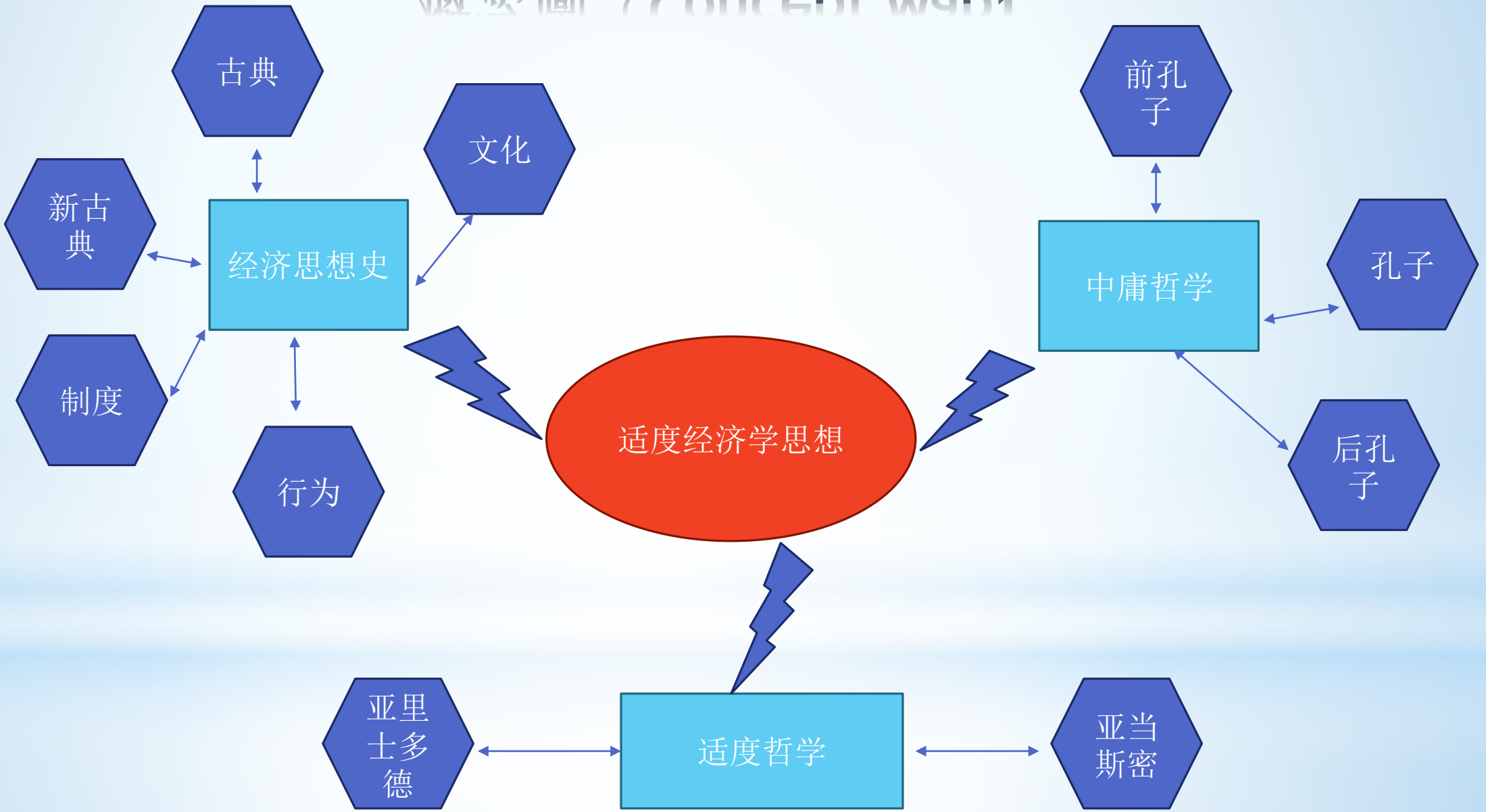
- 故意遗漏
- 无意遗漏
- 遗漏小人物文献

5）自以为原创而忽略文献引用

6）过度使用文献软件

- INCITES；EndNote Web；Reference Manager（Refman）；Reworks；NoteExpress; NoteFirst；PowerRef；Mendeley; Zotero）

概念圖（Concept Map）



三、 批判现有文献

1、为何要批判？

1）学术文化和习惯

- Critical thinking

2）深化研究的必需

3）讲真话的一大要素

4）寻找创新的机会

2、如何批判？

1）全面性、笼统性批判

2）具体性、精致性批判

3）三视角批判

- 康德的反例訓練

4）四维度批判

5）方法和数据批判

四、填补没有

1、何为填补或创新？

1）“Innovation means first **different**, then **better**.”

- 不同
- 更好

2) “Both the ‘different’ and the ‘better’ must be **significant** and **substantial**.” (Elliot Washor, *The Huffington Post*, October, 2009).

- 重大不同
- 本质性更好

3)创新“不同”与“更好”：皆取决于比较

2、如何填补与创新？

1）重点创新：新观点、新方法、新数据

2）复合式创新

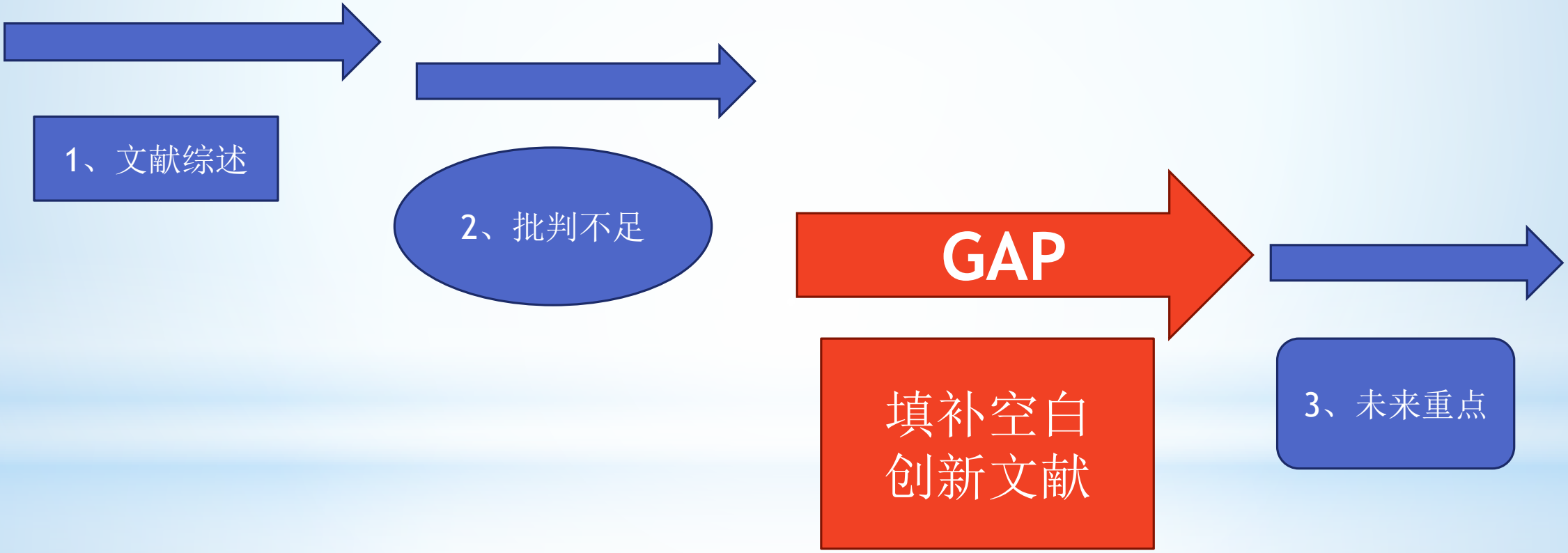
- 复杂经济学(The Complex Economics)
- 复合式进化(Evolution by Combination)

3）质疑式创新

- 例八：文革宗教场所空间化研究

4）图示式创新

文献研究三要求



文献研究案例讨论

探索数据低估对COVID-19时空分布的影响

鲍曙明
中国数据研究所

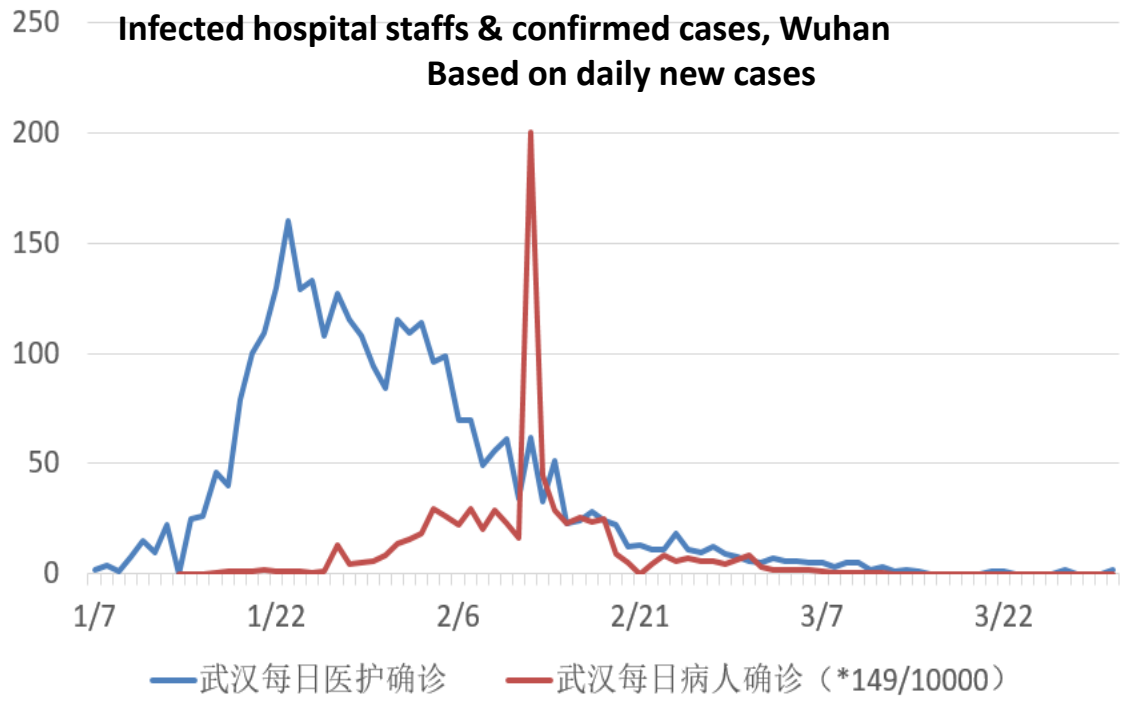
文献研究讨论要点

- 本项问题研究的理论、应用或现实**意义**?
- 本项研究**聚焦**的核心科学问题（抽象性、代表性、科学性）？
- 是否具备科学问题研究**特征**：可重复、可复制、可扩展?
- 文献研究**三要素**分析:
  - 发现已有
  - 批判现有
  - 填补没有

研究意义与科学问题

1. 研究意义：新冠确诊统计人数与实际感染人数存在巨大差异

2. 科学问题：如何利用相对精确的小样本（医护感染人数）估计误差较大的大样本（确诊统计人数）统计分布(感染峰期与区间)



文献研究三要素分析：发现已有

- **COVID-19数据低估相关研究：**
  - Lau et al. (2020)估计了多个国家实际的感染人数，发现疫情早期阶段西班牙、法国和美国等国家的报告数据被严重低估
  - Golding et al. (2020)重建了COVID-19在全球真实的空间分布
  - Prado et al. (2020)考虑了死亡率在不同年龄段间的差异，估计了巴西整体及各州的数据报告率，发现疫情早期巴西实际感染数量是报告数量的11倍
  - Saberi et al. (2020)发现忽略数据低估会严重影响基本再生数（effective reproduction number）的估算并高估控制措施的作用
  - Lachmann et al. (2020)发现数据低估现象将导致估计的死亡率与真实的死亡率存在较大偏差
- **数据缺失（低报）估计相关研究：**
  - 在流行病学、生物医学等公共健康领域（public health），数据低估（Under-reporting in data）是一种普遍现象。数据低估：报告的数量低于实际的数量[Amanda, L.；2016]。
  - 尤其是新发传染病COVID-19，由于疫情早期对传播机理不清晰、检测能力不足等原因，数据低估现象更加普遍[Cabaña A,2020]。
  - 数据被低估的危害很多[Alejandra,C.；2020]，数据低估现象对COVID-19的时空分布有什么影响？

文献研究三要素分析：批判现有

- **病人死亡数据依然是一种报告数据，报告的死亡人数与实际因COVID-19死亡的人数可能存在一定的偏差**
- **疫情早期阶段大量研究采用reported cases研究COVID-19的时空规律，但很少有研究探索数据低估对COVID-19时空分布的影响**

文献研究三要素分析：填补没有

- **数据：**精确小样本医护感染人数
- **方法：**
  - 综合不同方法（贝叶斯估计与蒙特卡罗统计模拟）提出一个峰值期差异综合估计权重
  - 空间分析方法扩展
- **发现：**
  - 肯定了其他研究关于新冠疫情确诊统计人数与实际感染人数的差异，对新冠早期感染峰值期和扩散区间结合时间和空间分析方法提出了科学估计

相较于病人死亡数据，医护感染数据是一个更加精确的小样本。疫情早期阶段，医护感染易发现，从而被统计。因此，本文采用医护感染数据，从一种回顾性的视角，分别从武汉市、湖北省（除武汉市）两个尺度分析了under-reporting cases对COVID-19时空分布的影响。从而为采用reported cases探索COVID-19时空分布相关的研究提供经验借鉴。