关于Computational的三个方面

* Aggregate Model
* Agent-based Model
* Data

Model：

* 方法与思维
  1. *Why Model?* Joshua M. Epstein
  2. *Advancing the Art of Simulation in the Social Sciences,* Robert Axelrod
* Aggregate Model
  1. *Lotka-Volterra equations* (predator-prey equations)
  2. *scale-free networks* (complex networks, statistical dynamics)
  3. *SIR* in epidemic spreading
* Agent-Based Model (Simulation)
  1. Introduction (Slides)
  2. 与求解数学问题的区分: approach for **proof** v.s. approach for **explanation**
  3. 最经典的案例：诺贝尔经济学奖得主Thomas C. Schelling对人口隔离现象的**解释**
     + *Dynamic Models of Segregation*, Thomas C. Schelling
     + Netlogo:<http://ccl.northwestern.edu/netlogo/models/Segregation>
  4. Complexity Economics*:* <http://www.inet.ox.ac.uk/programmes/complexity>
  5. Social Simulations*:* <http://jasss.soc.surrey.ac.uk/JASSS.html>
  6. 模拟历史事件
     + *Spanish Flu Pandemic Model*: <https://www.openabm.org/model/3890/version/2/view>
     + Reconstruct Ancient Civilization: *Anasazi*
  7. 学习方法

图片包含 屏幕截图

已生成极高可信度的说明

* 1. 研究工具

图片包含 屏幕截图

已生成极高可信度的说明

* 1. 其它：详见Slides与相关材料

Data：

* 面向General Public的一种可视化demo/toy model
  1. *Robots Reading Vogue*
  2. 小型爬虫的数据展示
* 面向 工程问题/分析与决策 的一种理论/技术 (Engineering, Computer Science, Mathematics)
  1. 知乎/Quora对重复问题的自动检测与合并
  2. Recommendation System: *“你可能还喜欢……”*
  3. Topic Model：自动标注一篇文章可能包含的话题
     + *Probabilistic Topic Models*, David M. Blei
     + 给一个知乎/Quora的问题title, 自动归档topic
  4. 基于数据驱动的决策分析
     + 在医院开一张血检 -> 某个**指数**超标 -> **决定**去看什么医生/治疗方法
     + 互联网数据分析：Google Analytics, Parse.ly
* 作为证据的数据

*It is a capital mistake to theorize before one has* ***data****.*

Sherlock Holmes

* 1. 基于统计的定量分析：回归，方差，假设检验，显著性
  2. 社交网络行为数据与现实行为的关系 （待补充）
  3. Topic Model
     + 通过文本研究政治影响因子: ***Bayesian hierarchical topic model for political texts*, Justin Grimmer**
     + *A Language-based Approach to Measuring Scholarly Impact*
     + 记忆机制: *A Bayesian Analysis of Dynamics in Free Recall*
  4. 关于人文学科 (Comparative Literature, History)：***Propose new hypothesis***
     + 洋务运动期间存在两个流派：顽固派与洋务派。假如，我们现在并不知道洋务运动中有这样一个划分，只有一批文本，其中包括了这个时期重要人物的书信、言论等。此时，我们利用聚类分析 (data clustering)，发现文本可以被划分成两个簇(cluster)，进而提出新的假设。
     + 词频分析：
       - ***Women | Girls***在历史文献中的词频经历了交替变化，是否证实了某个历史进程的影响？
       - ***America***直到1873年才开始频繁出现在美国总统就职演讲中，这个现象是否和美国历史进程有一致性？
     + 数据的性质（characteristic），如密集/稀疏/相关性，能带来什么启发？

[图片包含 天空, 室内, 屏幕截图

已生成高可信度的说明](picture/dispersion_plot.png)

* 1. 如何用数据聚类来帮助发现Pattern, 使研究者提出曾未提出的Hypothesis，或是加强/削弱现有理论？