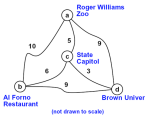
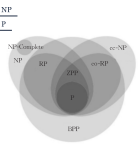
形式语言与计算理论

- 课程介绍
  - 回顾和复习下
    - 希伯特问题
      - 希尔伯特在1900年8月巴黎国际数学家代表大会上提出的最重要的23个数学问题。
        - 第1到第6问题是数学基础问题；
        - 第7到第12问题是数论问题；
        - 第13到第18问题属于代数和几何问题；
        - 第19到第23问题属于数学分析。
      - 希尔伯特的第10个问题
        - 能否通过有限步骤来判定不定方程是否存在有理整数解？
        - 扩展 → **问题是否有解/可解？** > *TM模型*
    - 写个程序：单词游戏
      - 需求
        - 有N个盒子，每个盒子上写着一个仅由小写字母组成的英文单词；
        - 需要给这些盒子安排一个合适的顺序，使得相邻两个盒子中，前一个盒子上面单词的末字母等于后一个盒子上面单词的首字母
        - 请写一个程序，判断是否能达到这个要求。数据规模：1≤N ≤100000
      - 抽象问题，建立模型，确认算法算法
        - 模型1
          - 以N个盒子作为顶点，如果盒子的末字母等于盒子的首字母，那么从向连一条有向边。
          - 例如，对于分别写有acm, malform, mouse的3个盒，按模型1可以将问题抽象为一个图。
            -

            - 这样，问题转化为在图中寻找一条不重复地经过每一个顶点的路径，即哈密尔顿路。然而，求哈密尔顿回路是一个困难的问题，这样的模型没有给我们的解题带来任何便利。
            - 需要遍历的信息——也就是每一个盒子——表示在顶点上,而顶点的遍历问题不易解决。能否将遍历信息表示在边上 呢？
        - 模型2
          - 以26个字母作为顶点：对于每一个 盒子，如果它的首字母为c,末字母为g,那么从c向g连一条
            -

          - 问题转化为在图中寻找一条不重复地经过每一条边 的路径，即欧拉路径。能够在O(N)时间内解决。
      - 可用的图模型
        - 哥德尼斯堡七桥问题
        - 汉密尔顿数学游戏
        - 哈密尔顿回路问题"与"欧拉回路问题"的比较；
      - 总结下 → **问题的解法如何？** > *计算模型*
        - 有限状态自动机
        - 下推自动机
    - 旅行商问题
      - Your boss says → **Get me an efficient traveling-salesman algorithm， or else**
      - your response
        - 

        - Parameters:
          - set of cities
          - set of inter-city distances
        - Solution
          - want the shortest tour through the cities
          - example: a,b,d,c,a has length 27.
      - 总结下 → **问题的解法使用的资源** > *复杂性问题*
        - 

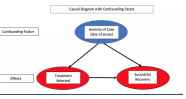
          各种经典算法复杂性关系
  - 课程讲述的内容
  - 课程成绩说明
- 第一部分：计算模型与形式语言
  - 有限自动机与正则语言
  - 下推自动机与上下文无关的语言
- 第二部分：可计算机理论
  - 图灵机与递归、递归可枚举的语言
  - 可判定性分析
  - 不可判定性分析
- 第三部分：复杂性理论
- 第四部分：其他计算模型-量子
- 第五部分：因果关系推理
  - 引言
    - 辛普森悖论
      - 辛普森悖论得名于英国统计学家E.H.辛普森（E.H.Simpson），这是处于1951年阐述的一种现象，当我们以分组和聚合两种方式统计同一数据集时，最后得出的两个趋势可能是完全逆转的。
      - 吃药与患病
        - 到底吃药还是不吃药？

	吃药患病（60人：40女，20男）	未吃药患病（60人：20女，40男）
男性	40%（8/20）	30%（12/40）
女性	7.5%（3/40）	5%（1/20）
聚合	18%（11/60）	22%（13/60）

        - 为什么会出现这种现象？ → 因果理论
          -

          - 混杂因子
            - 同时会影响原因和结果的因素，吃不吃药是原因，有没有得病是结果，那么是什么因素同时存在于扰这两点呢？性别!服药前后的发病率不一样，说明性别对服药结果有影响。混杂因子叫后门路径，像在这个因果做条件开启了一个后门，然后从后门进入干扰因果关系，意识混杂因子，需要做的就是阻断混杂因素。
            - 所谓阻断，是绝不能把跟混杂因子相关的因素直接加在一起，一定要分开计算，
              - 计算各自的得病概率，分开计算之后，这个混杂因素的干扰就被排除了
      - 肾结石治疗方案的取舍
        - 肾结石治疗方案的取舍

结石大小/治疗方案	方案A	方案B
小结石	81% 87/107	93% 234/270
大结石	192/263 73%	55% 55/80
总体	79%273/370	83%289/350

          - 你会选哪种治疗方案
            - 大结石肯定比小结石严重得多，而方案A比方案B更具侵入性（医学上带有一定创伤性的治疗措施）。因此，如果患者的肾结石很小，医生一般会保守起见，采用方案B；而如果患者的肾结石很大，医生就会直接用效果最好的方案A。
          - "病情严重性"称为混淆变量
            - 

      - Carlo's餐厅和Sophia餐厅

	推荐Sophia餐厅	推荐Carlo's餐厅
男性	50/150=30%	180/360=50%
女性	200/350=40%	36/60=60%
总计	250/500=42.5%	216/420=54%

        - 你会选择哪个餐厅？
        - 原因
          - 样本大小，当我们分组统计数据时，Carlo's餐厅的女性推荐率高达90%，但它的样本只有40个，只占总评论人数的10%；而Sophia餐厅的女性推荐率虽然只有80%，但女性评论者有250个，这显然会大幅拉高餐厅的总体好评率。
    - 中介因素 → 坏血病