

less 04 - 05 Basic NLP methods

less 06 Model, Validation, Test

less 07 Logistic Regression, Linear Regression

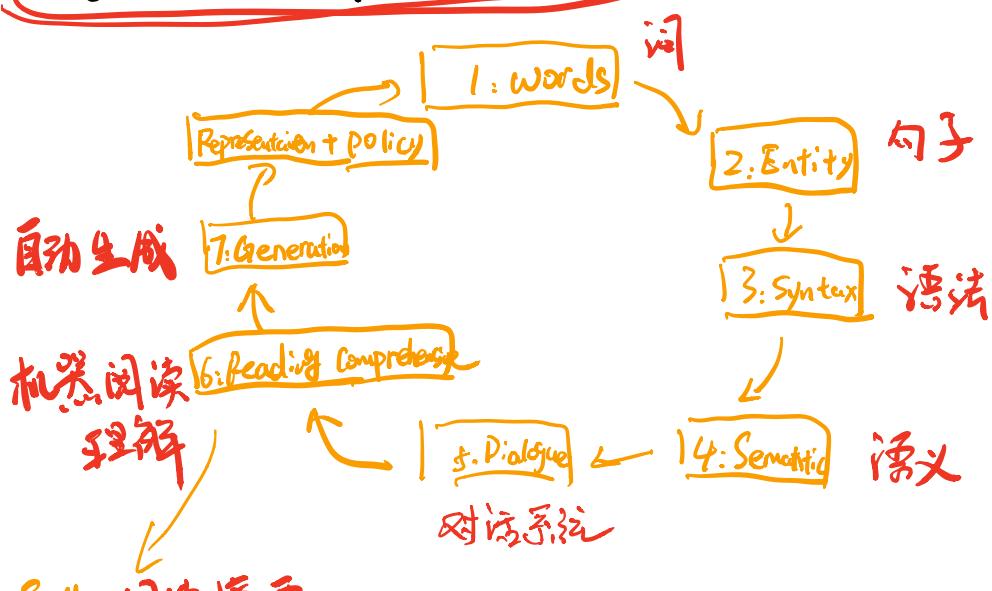
less 08 KNN, SVM, Bayes

less 09 Unsupervised Learning

less 10 ~ less 14 ML

less 14 - less 16 Dive into ML.

## What NLP concern



Representation: 信息如何表示  
policy: 信息如何计算

## sub-fields of AI

- 1. Computer Vision
- 2. Prediction
- 3. Optimization

4: Self Driving

5: Recommend System

6: etc.

---

## Why NLP

- 1: Information chaos 信息杂乱
- 2: Unstructured 信息非结构化
- 3: Discrete 信息不连续的
- 4: Unconventional 非常规的
- 5: OOV 未见过的词汇

困难点

- Language is the representation of mind
- the most classical field of AI
- the most immature field of AI
- the most sophisticated field of AI

精妙的

---

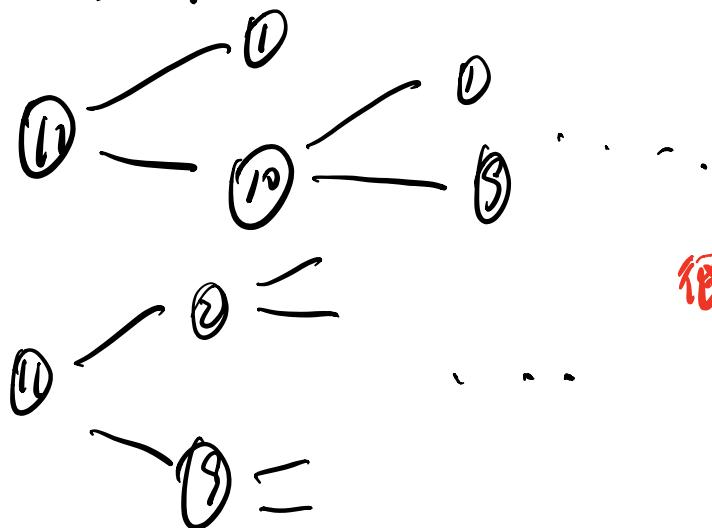
## Dynamic Programming

- 1: 切割问题 瓷砖如何切割卖的钱最多
- 2: 矩形覆盖问题 找到最大的矩形覆盖
- 3: 路线规划问题 经过所有点并且代价最小

## 一维切割问题

一根钢管，可以切成不同长短，不同长短价值不一样。  
如何使得卖价最高？

比如长度为 11



很多种解决方案

$$f_n = \max(P_n, f_1 + f_{n-1}, f_2 + f_{n-2}, \dots)$$

不切割时  
价钱

$f_x$  需要继续分解

两个问题：

1. 复杂度过高，导致大量的内存保存
2. 无法保存方案，中间停在打印

python3 自带 lru-cache  
装饰器可以解决

⇒ 动态规划

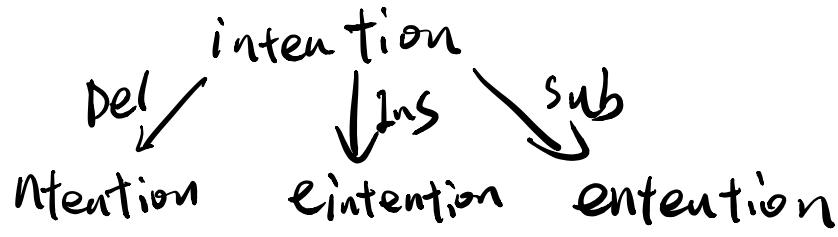
1. 分拆子问题重叠
2. 子问题存储
3. Solution 进行拆分

为什么不用机器学习，比如这种情况

- ① 无法拟合，比如逐步数据训练
- ② 没有必要，训练耗时可能比规划慢很多
- ③ 动态规划出来的是精确解

而拟合的无法确定是否为精确解

How similar are two strings



通过三种变化变成另外一个单词

并且变化可以持续递推下去.

- ⇒ Insert : 距离定义为 1  
 Delete : 距离定义为 1  
 Substitute : 距离定义为 2

$$D(i, j) = \min \left\{ \begin{array}{l} D(i-1, j) + 1 \\ D(i, j-1) + 1 \\ D(i-1, j-1) + 2 \quad \text{if } x(i) \neq y(j) \\ \text{else } D(i-1, j-1) \end{array} \right.$$