CSC401 Assignment 3

Tutorial 3 of 4

2021-03-24

Based on the slides of previous years
Agenda

• General introduction
• Speaker Recognition, Fitting to data, Gaussian Mixture Models
• Levenshtein distance (← this tutorial)
  • Dynamic Programming
  • Levenshtein distance
  • Additional examples
• Misc. Q&A for A3
Dynamic Programming

• Overall idea: Break down a large problem into:
  • A smaller problem of the same type
  • An induction step.

• Implement (1) the initialization condition and (2) the induction rule
  • Then this problem is solved.

• Levenshtein distance (Edit Distance) is a traditional DP problem.
Fibonacci Sequence


• Method 1: To compute $A[n]$, call $\text{fib}(n-1) + \text{fib}(n-2)$ recursively.
  • $O(n^2)$ time

• Method 2: Memoization: remember $\text{fib}[i<n]$ once computed
  • $O(n)$ time, $O(n)$ space

• Method 3: Optimize for space
  • You only need $O(1)$ space with $O(n)$ time.
Edit Distance

- Problem: Given two arrays $X = [x_0, ..., x_{M-1}]$ and $Y = [y_0, ..., y_{N-1}]$
  
  Find the edit distance from $X$ to $Y$ (let’s write as $A_{M,N}$)

- $A_{i,j}$ is the edit distance from $[x_0, ..., x_{i-1}]$ to $[y_0, ..., y_{j-1}]$

- Smaller problem: Find $A_{i,j}$

- Induction step: Find $A_{i+1,j+1}$
  - Here, $A_{i,j}, A_{i,j+1}, A_{i+1,j}$ -- those cells “in front of $A_{i,j}$” are already known
Edit Distance

• Initialization:
  • \( A_{i,0} = i, A_{0,j} = j \)

• Induction:
  • We only have four options actually:
    • If \( x_i = y_j \): \( A_{i+1,j+1} \leftarrow A_{i,j} \)
    • Otherwise: \{insert, delete, replace\}

• Finally:
  • \( A_{M,N} \) is the answer, by definition.
More Dynamic Programming

• Partition equal subset sum (Leetcode)
  
  https://leetcode.com/problems/partition-equal-subset-sum/

• Coin changing (HackerRank)
  
  https://www.hackerrank.com/challenges/coin-change

• Memoization and Dynamic Programming (by Gayle Laakmann McDowell)
  
  https://youtu.be/P8Xa2BitN3I