Advanced Algorithmics (6EAP)
Dynamic Programming

Jaak Vilo
2013 Spring

Example

• Wind has blown away the +, *, (, ) signs
• What’s the maximal value?
• Minimal?

2 1 7 1 4 3

• 2 1 7 1 4 3
• (2+1)*7*(1+4)*3 = 21*15 = 315
• 2*1 + 7 + 1*4 + 3 = 16

• Q: How to maximize the value of any expression?

2 4 5 1 9 8 12 1 9 8 7 2 4 1 1 2 3 = ?

Dynamic programming

• Avoid calculating repeating subproblems
• fib(1)=fib(0)=1;
• fib(n) = fib(n-1)+fib(n-2)
• Although natural to encode (and a useful task for novice programmers to learn about recursion) recursively, this is inefficient.


• Dynamic programming, like the divide-and-conquer method, solves problems by combining the solutions to subproblems.
  – Dynamically planeview.
• Divide-and-conquer algorithms partition the problem into independent subproblems, solve the subproblems recursively, and then combine their solutions to solve the original problem.
• In contrast, dynamic programming is applicable when the subproblems are not independent, that is, when subproblems share subsubproblems.
Structure within the problem

• The fact that it is not a \textit{tree} indicates overlapping subproblems.

• A dynamic-programming algorithm \textit{solves every subsubproblem just once} and then saves its answer in a table, thereby avoiding the work of recomputing the answer every time the subsubproblem is encountered.

Topp-down (recursive, memoized)

• \textit{Top-down approach}: This is the direct fall-out of the recursive formulation of any problem. If the solution to any problem can be formulated recursively using the solution to its subproblems, and if its subproblems are overlapping, then one can easily \textit{memoize} or store the solutions to the subproblems in a table. Whenever we attempt to solve a new subproblem, we first check the table to see if it is already solved. If a solution has been recorded, we can use it directly, otherwise we solve the subproblem and add its solution to the table.

Bottom-up

• \textit{Bottom-up approach}: This is the more interesting case. Once we formulate the solution to a problem recursively as in terms of its subproblems, we can try reformulating the problem in a bottom-up fashion: try solving the subproblems first and use their solutions to build-on and arrive at solutions to bigger subproblems. This is also usually done in a tabular form by iteratively generating solutions to bigger and bigger subproblems by using the solutions to small subproblems. For example, if we already know the values of $F_{41}$ and $F_{40}$, we can directly calculate the value of $F_{42}$.

Dynamic programming is typically applied to \textit{optimization problems}. In such problems there can be many possible solutions. Each solution has a value, and we wish to find a solution with the optimal (minimum or maximum) value.

• We call such a solution \textit{an} optimal solution to the problem, as opposed to \textit{the} optimal solution, since there may be several solutions that achieve the optimal value.

The development of a dynamic-programming algorithm can be broken into a sequence of four steps.

1. Characterize the structure of an optimal solution.
2. Recursively define the value of an optimal solution.
3. Compute the value of an optimal solution in a bottom-up fashion.
4. Construct an optimal solution from computed information.
**Edit (Levenshtein) distance**

- **Definition** The edit distance $D(A,B)$ between strings $A$ and $B$ is the minimal number of edit operations to change $A$ into $B$. Allowed edit operations are deletion of a single letter, insertion of a letter, or replacing one letter with another.
- Let $A = a_1 a_2 ... a_m$ and $B = b_1 b_2 ... b_n$.
  - $E_1$: Deletion $a_i \rightarrow \varepsilon$
  - $E_2$: Insertion $\varepsilon \rightarrow b_i$
  - $E_3$: Substitution $a_i \rightarrow b_j$ (if $a_i \neq b_j$)
- Other possible variants:
  - $E_4$: Transposition $a_i a_{i+1} \rightarrow b_j b_{j+1}$ and $a_i \neq b_j \neq a_{i+1}
  (e.g. lecture \rightarrow letcure)

**Dynamic Programming**

Recursion?

<table>
<thead>
<tr>
<th>i</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>_</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>i</td>
<td>_</td>
<td>d_{i-1,j-1}</td>
<td>d_{i-1,j}</td>
<td>d_{i,j-1}</td>
<td>d_{i-1,j}</td>
<td>d_{i,j-1}</td>
<td>d_{i,j}</td>
<td>d_{i,j}</td>
<td>d_{i,j}</td>
<td>d_{i,j}</td>
<td>d_{i,j}</td>
</tr>
</tbody>
</table>

| Algorithm Edit distance $D(A,B)$ using Dynamic Programming (DP) |
Matrix multiplication

\[ \text{for } i = 1 \ldots n \]
\[ \text{for } j = 1 \ldots k \]
\[ c_{ij} = \sum_{x=1}^{m} a_{ix} b_{xj} \]

The matrix multiplication problem can be stated as follows: given a chain \( <A_1, A_2, \ldots, A_n> \) of \( n \) matrices

- matrix \( A_i \) has dimension \( p_{i-1} \times p_i \)
- fully parenthesize the product \( A_1 A_2 \ldots A_n \) in a way that minimizes the number of scalar multiplications.

\[ \text{O}(nmk) \]

\[ \text{MATRIX-MULTIPLY}(A, B) \]
1 if columns \( [A] \neq \text{rows } [B] \)
2 then error "incompatible dimensions"
3 else for \( i = 1 \) to columns \( [A] \)
4 do for \( j = 1 \) to columns \( [B] \)
5 \( \text{do } C[i, j] = 0 \)
6 for \( k = 1 \) to columns \( [A] \)
7 \( \text{do } C[i, j] = C[i, j] + A[i, k] \times B[k, j] \)
8 return \( C \)
\[ A_1A_2A_3A_4 \]

- \( A_i(A_j(A_kA_l)) \)
- \( (A_iA_jA_k)A_l \)
- \( (A_iA_j)(A_kA_l) \)
- \( ((A_iA_j)A_k)A_l \)
- \(((A_iA_j)A_k)A_l) \)

Denote the number of alternative parenthesizations of a sequence of \( n \) matrices by \( P(n) \).

- Since we can split a sequence of \( n \) matrices between the \( k \)th and \((k+1)\)st matrices for any \( k = 1, 2, \ldots, n -1 \) and then parenthesize the two resulting subsequences independently, we obtain the recurrence

\[
P(n) = \sum_{k=1}^{n-1} P(k)P(n-k) \quad \text{if } n \geq 2.\]

Let’s crack the problem

\[ A_{i..j} = A_i \cdot A_{i+1} \cdot \ldots \cdot A_j \]

- Optimal parenthesization of \( A_i \cdot A_{i+1} \cdot \ldots \cdot A_n \) splits at some \( k, k+1 \).
- Optimal = \( A_{i..k} \cdot A_{k+1..n} \)
- \( T(A_{i..k}) = T(A_{i..j}) + T(A_{k+1..n}) + T(A_{i..j} \cdot A_{k+1..n}) \)
- \( T(A_{i..j}) \) must be optimal for \( A_i \cdot A_{i+1} \cdot \ldots \cdot A_j \)

Recursion

- \( m[i, j] \) - minimum number of scalar multiplications needed to compute the matrix \( A_{i..j} \)
- \( m[i, j] = 0 \)
- \( \text{cost}(A_{i..j} \cdot A_{k+1..j}) = p_{i-1}p_kp_j \)
- \( m[i, j] = m[i, k] + m[k+1..j] + p_{i-1}p_kp_j \).

This recursive equation assumes that we know the value of \( k \), which we don’t. There are only \( j - i \) possible values for \( k \), however, namely \( k = i, i+1, \ldots, j-1 \).

Let’s solve the problem

\[ A_{i..j} = A_i \cdot A_{i+1} \cdot \ldots \cdot A_j \]

- \( m[i, j] = 0 \)
- \( \text{cost}(A_{i..j} \cdot A_{k+1..j}) = p_{i-1}p_kp_j \)
- \( m[i, j] = \min \{ m[i, k] + m[k+1..j] + p_{i-1}p_kp_j \} \quad \text{if } i < j.\) (16.2)

To help us keep track of how to construct an optimal solution, let us define \( s[i, j] \) to be a value of \( k \) at which we can split the product \( A_{i..j} \cdot \ldots \cdot A_{n} \) to obtain an optimal parenthesization. That is, \( s[i, j] \) equals a value \( k \) such that \( m[i, j] = m[i, k] + m[k+1..j] + p_{i-1}p_kp_j \).
Recursion

- Checks all possibilities...

- But – there is only a few subproblems – choose i, j s.t. 1 ≤ i ≤ j ≤ n - O(n^2)

A recursive algorithm may encounter each subproblem many times in different branches of its recursion tree. This property of overlapping subproblems is the second hallmark of the applicability of dynamic programming.

Example

A simple inspection of the nested loop structure of MATRIX-CHAIN-ORDER yields a running time of O(n^3) for the algorithm. The loops are nested three deep, and each loop index (l, i, and k) takes on at most n values.

- Time Ω(n^3) = Θ(n^3)
- Space Θ(n^2)

Step 4 of the dynamic-programming paradigm is to construct an optimal solution from computed information.

Use the table s[1...n, 1...n] to determine the best way to multiply the matrices.

Multiply using S table

```
MATRIX-CHAIN-MULTIPLY(A, s, i, j)
1 if j > i
2 then X = MATRIX-CHAIN-MULTIPLY(A, s, i, s[i, j])
3 Y = MATRIX-CHAIN-MULTIPLY(A, s, s[i, j]+1, j)
4 return MATRIX-MULTIPLY(X, Y)
5 else return A

((A_1(A_2A_3))(A_4A_5A_6))
```
Elements of dynamic programming

• Optimal substructure within an optimal solution

• Overlapping subproblems

• Memoization

A memoized recursive algorithm maintains an entry in a table for the solution to each subproblem. Each table entry initially contains a special value to indicate that the entry has yet to be filled in. When the subproblem is first encountered during the execution of the recursive algorithm, its solution is computed and then stored in the table. Each subsequent time that the subproblem is encountered, the value stored in the table is simply looked up and returned. (tabulated)

This approach presupposes that the set of all possible subproblem parameters is known and that the relation between table positions and subproblems is established. Another approach is to memoize by using hashing with the subproblem parameters as keys.

Overlapping subproblems

Longest Common Subsequence (LCS)

Optimal triangulation

Two ways of triangulating a convex polygon. Every triangulation of this 7-sided polygon has 7 - 2 = 4 chords and divides the polygon into 7 - 2 = 5 triangles.

The problem is to find a triangulation that minimizes the sum of the weights of the triangles in the triangulation.

Parse tree

Parse trees. (a) The parse tree for the parenthesized product (A_bA_cA_d)(A_eA_fA_g) and for the triangulation of the 7-sided polygon (b) The triangulation of the polygon with the parse tree overlaid. Each matrix A_i corresponds to the side v_i, v_i for i = 1, 2, . . . , 6.
Optimal triangulation

\[ f(i,j) = \begin{cases} 0 & \text{if } i = j, \\ \min_{k=0}^{\min(i,j)-1} \{ f(i,k) + f(k+1,j) + w(\Delta a_n, a_k) \} & \text{if } i < j. \end{cases} \] (16.7)

Edit distance (Levenshtein distance)

- Smallest nr of edit operations to convert one string into the other

\[
\begin{array}{c|c}
\text{INDUSTRY} & \text{INTEREST} \\
\hline
\text{INDUSTRY} & \text{INTEREST} \\
\end{array}
\]

Definition

The edit distance \( D(A,B) \) between strings \( A \) and \( B \) is the minimal number of edit operations to change \( A \) into \( B \). Allowed edit operations are deletion of a single letter, insertion of a letter, or replacing one letter with another.

- Let \( A = a_1 a_2 \ldots a_m \) and \( B = b_1 b_2 \ldots b_n \).
  - E1: Deletion \( a_i \rightarrow \epsilon \)
  - E2: Insertion \( \epsilon \rightarrow b_j \)
  - E3: Substitution \( a_i \rightarrow b_j \) (if \( a_i \neq b_j \))
- Other possible variants:
  - E4: Transposition \( a_i \leftrightarrow a_{i+1} \rightarrow b_j b_{j+1} \) and \( a_i b_{j+1} \leftrightarrow a_{i+1} b_j \)
  (e.g. lecture \( \rightarrow \) lecture)

How can we calculate this?

\[
D(a,b) =
\begin{cases}
\alpha & \text{if } a = b \\
\beta & \text{if } a \neq b
\end{cases}
\]

\[
D(a, b) = \begin{cases}
\min(D(a[1..n-1], b[1..m-1]), & \text{if } a_n = b_m) \\
D(a[1..n], b[1..m-1]) + 1, & \text{if } a_n \neq b_m
\end{cases}
\]

How can we calculate this efficiently?

\[
D(S,T) =
\begin{cases}
\min(D(S[1..n-1], T[1..m-1]), & \text{if } S_n = T_m) \\
D(S[1..n], T[1..m-1]) + 1, & \text{if } S_n \neq T_m
\end{cases}
\]

Define:

\[
d(i,j) = D(S[1..i], T[1..j])
\]

Recursion?

\[
d(i,j) = \begin{cases}
\min(d(i-1,j-1), & \text{if } S_i = T_j) \\
d(i-1,j) + 1, & \text{if } S_i \neq T_j
\end{cases}
\]
Recursion?

Algorithm Edit distance D(A,B) using Dynamic Programming (DP)

Input: A=a₁, a₂, ..., aₙ  B=b₁, b₂, ..., bₘ
Output: Value dᵟₙ in matrix (dᵟₙ), Osijom, Osijsn.

for i=0 to m do dᵢ₀=1 ;
for j=0 to n do d₀ⱼ=j ;
for j=1 to n do
  for i=1 to m do
    dᵢⱼ = min( dᵢ₋₁,ⱼ₋₁ + (if aᵢ==bⱼ then 0 else 1),
               dᵢ₋₁,ⱼ + 1,
               dᵢ,ⱼ₋₁ + 1 )

return dᵟₙ

Dynamic Programming

Edit distance is a metric

• It can be shown, that D(A,B) is a metric
  – D(A,B) ≥ 0, D(A,B)=0 iff A=B
  – D(A,B) = D(B,A)
  – D(A,C) ≤ D(A,B) + D(B,C)
Path of edit operations

- Optimal solution can be calculated afterwards
  - Quite typical in dynamic programming
    \[ d_{i-1,j-1} \quad d_{i-1,j} \]
    \[ d_{i,j-1} \quad d_{i,j} \]
  - Memorize sets \( pred[i,j] \) depending from where the \( d_i \) was reached.

Three possible minimizing paths

- Add into \( pred[i,j] \)
  - \((i-1,j-1)\) if \( d_{ij} = d_{i-1,j-1} + (\text{if } a_i = b_j \text{ then } 0 \text{ else } 1)\)
  - \((i-1,j)\) if \( d_{ij} = d_{i-1,j} + 1\)
  - \((i,j-1)\) if \( d_{ij} = d_{i,j-1} + 1\)

Multiple paths possible

- All paths are correct
- There can be many (how many?) paths

Space can be reduced

![Diagram](image)

The path (in reverse order) \( \varepsilon \rightarrow c_6, b_5 \rightarrow b_5, c_4 \rightarrow c_4, a_3 \rightarrow a_3, a_2 \rightarrow b_2, b_1 \rightarrow a_1 \).

Calculation of \( D(A,B) \) in space \( \Theta(m) \)

<table>
<thead>
<tr>
<th>( A )</th>
<th>( B )</th>
<th>( C )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 1 2 3 4 5 6</td>
<td>a b a c b c</td>
<td>3</td>
</tr>
<tr>
<td>1 1 2 3 4 5</td>
<td>b 1 1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>2 1 2 1 2 3 4</td>
<td>a 2 1 2 1 2 3 4</td>
<td></td>
</tr>
<tr>
<td>3 2 2 1 2 3 4</td>
<td>c 4 3 3 3 2 3 3</td>
<td></td>
</tr>
<tr>
<td>4 3 3 3 2 3 3</td>
<td>b 5 4 3 4 3 2 3</td>
<td></td>
</tr>
<tr>
<td>5 4 3 4 3 2 3</td>
<td></td>
<td>C=2</td>
</tr>
</tbody>
</table>

\[ C[m] \]

Input: \( A=a_1 a_2 \ldots a_m \), \( B=b_1 b_2 \ldots b_n \) (choose \( m \leq n \))

Output: \( d_{mn} = D(A,B) \)

for \( i=0 \) to \( m \) do \( C[i]=i \)
for \( j=1 \) to \( n \) do
  \( C=C[0]; C[0]=j; \)
for \( i=1 \) to \( m \) do
  \( d = \min( C + (\text{if } a_i = b_j \text{ then } 0 \text{ else } 1), C[i-1] + 1, C[i] + 1 ) \)
  \( C = C[] \quad // \text{memorize new } \text{"diagonal" value} \)
  \( C[i] = d \)
write \( C[m] \)

Time complexity is \( \Theta(mn) \) since \( C[0..m] \) is filled \( n \) times
**Observations?**

- Shortest path is close to the diagonal
  - If a short distance path exists
- Values along any diagonal can only increase (by at most 1)

**Diagonal**

Property of any diagonal: The values of matrix \( d_{ij} \) can on any specific diagonal either increase by 1 or stay the same

Diagonal \( k \), \(-m \leq k \leq n\), s.t. diagonal \( k \) contains only \( d_{ij} \) where \( j-i = k \).
Extensions to basic edit distance

• New operations
• Variable costs
• ...

Transposition (ab → ba)

• E4: Transposition
  \[ a_{i+1}a_i \rightarrow b_ib_{i+1}, \text{ s.t. } a_i=b_{j+1} \text{ and } a_{i+1}=b_j \]
• (e.g.: lecture → letcure)

\[
\begin{align*}
d(i,j) &= \min \begin{cases} 
  1. & d(i-1,j-1) + (S[i]=T[j]|0 : 1) \\
  2. & d(i, j-1) + 1 \\
  3. & d(i-1, j) + 1 \\
  4. & d(i-2,j-2) + (\text{if } S[i-1]=T[j+1] \text{ then } 1 \text{ else } \infty) 
\end{cases}
\end{align*}
\]

Generalized edit distance

• Use more operations E1...En, and to provide different costs to each.
• Definition. Let \( x, y \in \Sigma^* \). Then every \( x \rightarrow y \) is an edit operation. Edit operation replaces \( x \) by \( y \).
  – If \( A=uxv \) then after the operation, \( A=uyv \)
• We note by \( w(x \rightarrow y) \) the cost or weight of the operation.
• Cost may depend on \( x \) and/or \( y \). But we assume \( w(x \rightarrow y) \geq 0 \).

Applications of generalized edit distance

• Historic documents, names
• Human language and dialects
• Transliteration rules from one alphabet to another
  e.g. Tõugu => Tyugu (via Russian)
• ...

Examples
näituseks – näiteks
Ahwrika - Afrika
weikese - väikese
materjaali - materjali

“kavalam” otsimine
Dush, dušš, dushš ?
Gorbatšov, Gorbatšov, Горбачов,
Gorbachev
režim, rezhim, rim

possible problems/tasks
• Manually create sensible lists of operations
  – For English, Russian, etc…
  – Old language,
• Improve the speed of the algorithm (testing)
• Train for automatic extraction of edit operations and respective costs from examples of matching words...

advanced dynamic programming
• Robert Giegerich:
  – http://www.techfak.uni-bielefeld.de/ags/pi/lehre/ADP/
• Algebraic dynamic programming
  – Functional style
  – Haskell compiles into C