Greedy Algorithm

1 Definition

A greedy algorithm is a simple, intuitive algorithmic approach that makes the best or most optimal choice at each step as it attempts to find the overall optimal way to solve the entire problem. This means that it picks the best immediate option without considering the broader consequences and hopes that by choosing a local optimum at each step, a global optimum will be reached.

2 Interval Scheduling Problem

2.1 Problem Definition

Assume job j starts at s_j and finishes at f_j. Two jobs are compatible if they don't overlap. The goal is to find **maximum subset** of mutually compatible jobs.

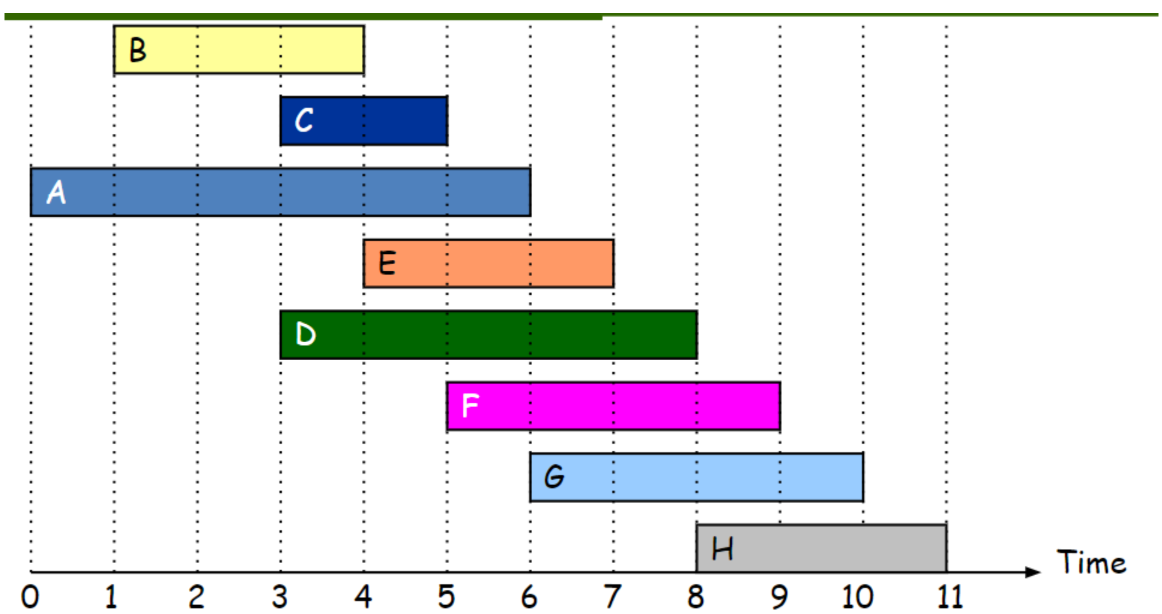

Figure 1: Interval Scheduling

The optimized solution is **choosing next job to add to solution as the one with earliest finish time that it is compatible with the ones already taken.**

2.2 Interval Definition

Let the set of all intervals be:

$$I = \{i_1, i_2, ..., i_n\} \tag{1}$$

where each interval i_j is represented by its start and finish times (s_j, f_j). Without loss of generality, assume these intervals are sorted by their **finish times**:

$$f_1 \leq f_2 \leq ... \leq f_n \tag{2}$$

2.3 Solution Definition

The set of intervals selected by the greedy algorithm, sorted by their finish times is defined as:

$$G = \{g_1, g_2, ..., g_k\} \tag{3}$$

The set of intervals in an optimal solution, also sorted by their finish times is defined as:

$$O = \{o_1, o_2, ..., o_m\} \tag{4}$$

2.4 Greedy Stays Ahead Argument

Need to prove that $f_{g_j} \leq f_{o_j}$ for all j where $1 \leq j \leq min(k, m)$. This implies that the j-th interval selected by the greedy algorithm finishes no later than the j-th interval in optimal solution.

2.5 Base Case

For $j = 1$, because g_1 is the interval with the earliest finish time, $f_{g_1} \leq f_{o_1}$

2.6 Inductive Step

1. Assume $f_{g_j} \leq f_{o_j}$ holds for particular j
2. Consider the next interval g_{j+1} chosen by Greedy, then g_{j+1} is the interval with the earliest finish time after g_j
3. Therefore, $f_{g_{j+1}}$ is the earliest possible finish time for any interval starting after f_{g_j}
4. Since $f_{g_j} \leq f_{o_j}$, the interval o_{j+1} in the optimal solution will not finish earlier than $f_{g_{j+1}}$, so we have:

$$f_{g_{j+1}} \leq f_{o_{j+1}} \tag{5}$$

2.7 Pseudo Code

$O(n \log n)$

```
Sort jobs by finish times so that f_1 ≤ f_2 ≤ ... ≤ f_n.
        set of jobs selected
A ← φ
f_{j*} = 0
for j = 1 to n {
   if (job j compatible with A : s_j ≥ f_{j*}) {
      A ← A ∪ {j}
      f_{j*} = f_j
   }
}
return A
```

$O(n)$

Running time: $O(n \log n)$

Figure 2: Interval Scheduling Pseudo Code