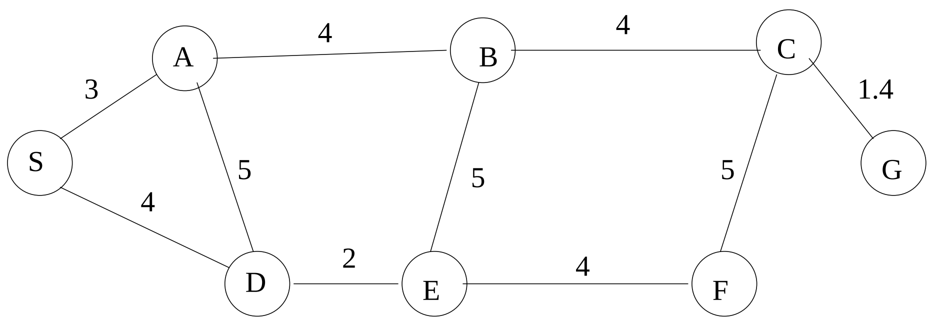
University of Pittsburgh
Introduction to Artificial Intelligence — Handout 2
Professor Milos Hauskrecht — September 24, 2009

Solutions to problem assignment 2

Problem 1

Consider the following graph that represents road connections between different cities. The weights on links represent driving distances between connected cities. Let S be the initial city and G the destination.



Part a. Show how the uniform search tree works. Is the path found by the algorithm optimal?

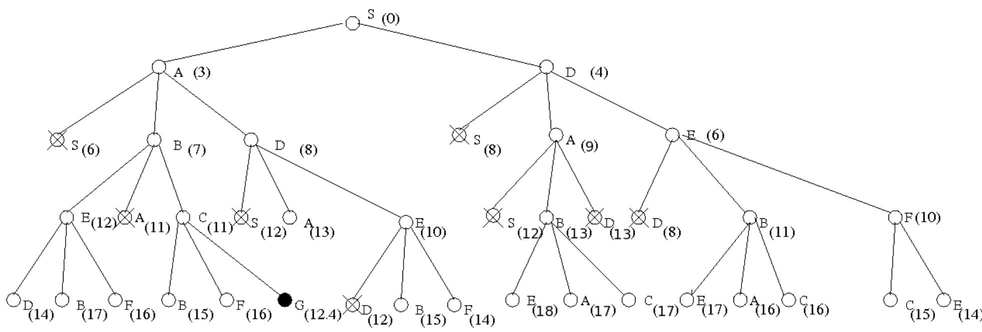
Part b. Assume the following set of the straight line distances between G and other cities.

S	A	B	C	D	E	F
10	10	6	1.4	9	7	2

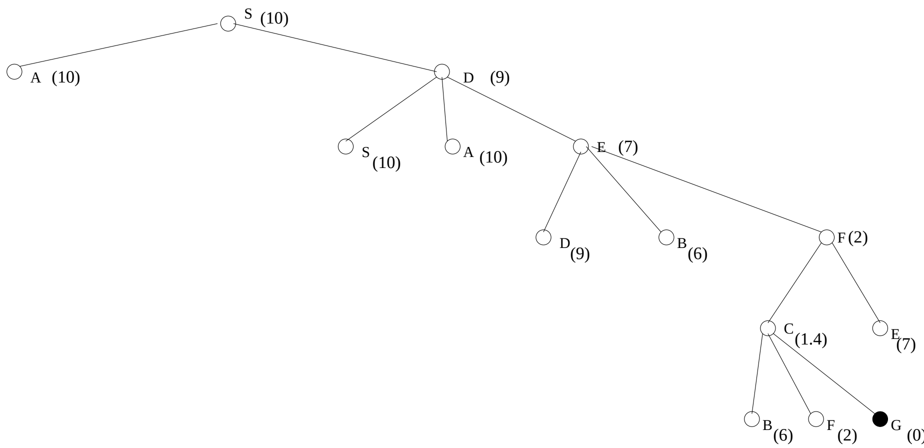
Show how the greedy search algorithm with the straight-line distance heuritic works. Is the path the algorithm finds optimal?

Part c. Show how the A* with the straight-line distance heuristic works. Is the path found optimal?

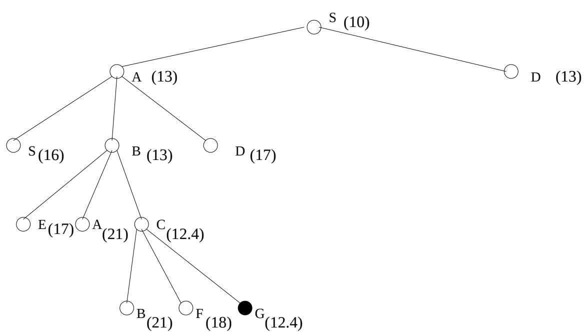
Solutions. Search trees generated by three different strategies are shown below. For the sake of simplicity we do not expand branches with cyclic state repeats. Crossed nodes represent nodes with cyclic repeats (checked at the time of the expansion). However, it is important to note that basic evaluation-function driven algorithms do not check for repeats of any kind.



Part a. Uniform cost search with the elimination of cyclic repeats.



Part b. Greedy search.



Part c. A* search.

Uniform cost search and A^* with a straightline distances are both complete and optimal. Greedy search with a straight line distances is not complete since it may enter an infinite loop and it is also not optimal. One way to prevent incompleteness of greedy methods is to use cyclic repeats. However, this fix does not solve the optimality problem and solutions we see may not be optimal. In our example, the path between S and G found by the greedy method is clearly suboptimal and of length 16.4 instead of 12.4.

Problem 2. Heuristic search for the 8-puzzle problem.

In this problem we will explore heuristic search solutions for the 8-puzzle problem. We will use evaluation-function driven search procedure to incorporate various exploration strategies. The procedure searches the space by expanding the nodes with the minimum value first .

Solutions for parts b-e. See program files on the course web page.

Part f. Analysis of results

Analyze the performance of all methods in terms of the collected statistics and include the analysis in the problem set report. You should:

- Summarize the results of the methods in three different tables, one table for every configuration tested.
- Write which method is the best in terms of respective statistics. Explain why?
- Write what heuristic would you suggest to use and why?

Solution. The statistics collected for EXAMPLE 1 are:

method	path cost	N. generated	N. expanded	Max q length
Uniform cost	3	44	16	22
A^* + MisTile	3	10	3	8
A^* + ManDist	3	10	3	8
Greedy + MisTile	3	10	3	8

The statistics collected for EXAMPLE 2:

Method	path cost	N. generated	N. expanded	Max q length
Uniform cost	8	476	169	204
A^* + MisTile	8	22	8	15
A^* + ManDist	8	22	8	15
Greedy + MisTile	8	22	8	15

The statistics collected for EXAMPLE 3:

Method	path cost	N. generated	N. expanded	Max q length
Uniform cost	10	1804	672	691
A^* + MisTile	10	71	26	39
A^* + ManDist	10	36	13	23
Greedy + MisTile	46	2531	932	659

The best method out of the evaluation-function driven methods is A^*.

The greedy search performs well but its performance depends heavily on the quality of the heuristics. In EXAMPLE 3 this effect is visible the most. This also distorted the effective branching factor, since the depth of the solution was large. It would be more adequate to compute the branching factor using the depth of the optimal solution for methods that are suboptimal (like greedy search).

Misplaced tile heuristic in combination with the greedy method is quite bad and it is outperformed by the uniform cost search. Note also that the method was not able to find the optimal solution of length 10.

Clearly, Manhattan distance heuristics is better than the misplaced tile heuristics. Both of them are admissible, but Manhattan distance values are larger, thus leading to better estimates of the remaining distance to the goal.

In addition answer the following questions.

- Would A* work without marking? Why or why not?
  Answer: Yes. A^* uses g-function that measures the cost of the path from the initial state. Thus, even if do not remove the repeats the nodes with the same underlying states differ in terms of their g-values. The paths that are worse get higher g-value and are preferred less. Thus one cannot get stuck in a cycle that corresponds to an infinite length path.

- Would greedy method work without marking? Why or why not?
  Answer: No. The greedy method needs to check at least for cycle repeats. Without the check it is possible it will get stuck and loop forever building and expanding an infinite length path.

- Assume we create a heuristic function h_3 such that it averages the values of the misplaced tile heuristic (h_1) and the Mahattan distance heuristic (h_2):

  $$h_3(n) = \frac{1}{2}\left[h_1(n) + h_2(n)\right].$$

  Is h_3 an admissible heuristic?
  Answer: Yes. Since both heuristics are admissible, i.e. $h_1(n) \leq h^*(n)$ and $h_2(n) \leq$

$h^*(n)$, it must hold:

$$\begin{aligned} h_3(n) &= \frac{1}{2}\left[h_1(n) + h_2(n)\right] \\ &\leq \frac{1}{2}\left[h^*(n) + h_2(n)\right] \\ &\leq \frac{1}{2}\left[h^*(n) + h^*(n)\right] \\ &= \frac{1}{2} 2h^*(n) \\ &= h^*(n) \end{aligned}$$