Massachusetts Institute of Technology

6.034 Artificial Intelligence

Solutions # **Final95** 6034 Item # **31**

Problem 1 Search

Part A

In the worst case, the search tree will have $3 \times 2^5 + 1 = 97$ leaves. In the best case, the tree will have 2 leaves.

Part B

Hill climibing will generate 2 leaves since the descendant of S that is closest to G is on the direct path.

Part C

Branch and bound with lower bound will generate 2 leaves since the actual path length to the two descendants of S are the same and the straight line distance estimate from the top node to the goal is longer than the actual path length to the goal of the bottom node.

Part D

Without a lower bound you will generate the same answer (there is only one) but you will generate a lot more leaves (because each path must be extended until it is longer than the actual best path length).

Part E

Branch and Bound with lower bound will generate $2^8 = 256$ leaves.

Part F

Without lower bound, one generates the same number of leaves (since the lower bound for all the nodes is about the same and useless) and the same path (the shortest one).

Part G

From S to G1 using best-first search you only expand the two intersections along shortest path.

Part H

From S to G2 best-first expands all intersections on the left block as well as the intersections along the best path from I to G2.

Part I

With the modified method, only the intersections along the best path get expanded from S to G2.

Problem 2 Constraints

Part A

1. **No**. This requires that lines in the picture have the same length as lines on Waldo, but length in the picture can vary due to changes in scale.
2. **No**. This is non-sense; it requires that lines all be equal length.
3. **Yes**. The angle between pairs of lines on Waldo and between the matching lines in the picture needs to be the same.
4. **Yes**. This is a bit subtle. This says that the angle between every line on Waldo and the matching line in the figure has to be the same. This follows because Waldo is rigid and when you rotate him, every line is rotated by the same angle.
5. **Yes**. The ratio of the line lengths between pairs of lines on Waldo and between the matching lines in the picture needs to be the same.
6. **Yes**. The reasoning is the same as in 4, except with scale instead of rotation.
7. **Yes**. The reasoning is the same as in 6. The distances behave as if they were line lengths for additional lines. Note that the ratios in this constraint (as in 6) are equal to the scale factor of Waldo in the picture.

8. **No.** This is non-sense; it requires that the ratio of line lengths on Waldo equal the scale factor of the picture.

Part B

The code here is slightly different from this year but closely related.

1. **True**, AC-4 will start by checking the constraints (arcs) involving the present variable but if any value is deleted from a domain it will propagate. Therefore, after the propagation is done all the values left in the domains are consistent with the current assignment.
2. **False**, AC-3 does more work (constraint checks) than AC-4 since it starts out checking every arc not just those relevant to the present variable, which is the only one that has changed.
3. **True (sort of)**. The idea is that if a variable did not cause a deletion from the current domain, then changing its value will not "add in" any new values to the domain and therefore not help us avoid the failure next time we get back to this variable. The (sort of) part is that this is not always guaranteed to work because of indirect propagation effects.

Part C

1. Backtrack is a form of depth-first search.
2. Note that, because of wildcards, any subset of a correct solution is also a solution. Also, once you find any pair of lines in the target that match those in the picture (even if it's the wrong match) that can lead to a consistent solution in which all the other target lines match the wild card. So, the odds of finding the longest solution during DFS is essentially zero.
3. One could do branch and bound search using a count of the number of wildcards in the current solution as a measure of path length. The resulting solution would be the longest consistent assignment, i.e. the one with the fewest wildcards. Note, however, that this search has no incentive to extend longer solutions first. So, we would probably want to sort entries on the queue that are tied on the basis of number of wildcards using the number of (non-wildcard) assignments.

Problem 3 Probabilistic Reasoning

Part A

We expect that 48 human faculty will be fired (0.96 x .05 = 0.048).

Part B

```
P(Robot | NoReaction) = P(NoReaction | Robot) P(Robot)/P(NoReaction)
P(NoReaction) = P(NoReaction | Robot)P(Robot) + P(NoReaction | Human)P(Human)
P(Robot | NoReaction) = 0.038/0.086 = 0.44 (not so small)
```

Part C

```
               0.2
  do nothing |------ -30M
 |-----------O <-10M>
 |           |------ -5M
 |              0.8
 |
 X <-5M>       0.75
 |           |------ 20M
 |           |             0.5
 |hire inv   |        sell |----- -100M (-100M + -50M + 50M)
 |-----------O<-5M> |------O <-80M>
             |      |      |----- -60M  (-100M + -50M + 90M)
             |      |        0.5
             |------X <-80M>
               0.25 |
                    | no sell
                    |-------- -100M
```

The expected values of the nodes is shown in brackets. The recommended action is to hire the investor and to sell the athletic fields if he loses the 100 Million.

Problem 4 Machine Learning

Part A

There are many possible sequences. The key is to make sure that there is only a single difference between the new sample and the evolving model. One possible answer is:

Sequence: 1, 4, 3, 5, 6, 2, 7, 9, 10, 11

In this sequence, 1 is a positive use to initialize the model. Then the subsequence (4, 3, 5, 6, 2) forces each of the (X Is-a Male) links to become (X Must-be Human). Note that the order of presentation of these is fairly constrained since the samples have different numbers of (X Is-a Female) links. The subsequence (7, 9, 10, 11) forces the (X Must-Parent Y) links. But, since there is only one link missing in each sample, they can be presented in any order. The order of these two subsequences (4,...,2) and (7,...,11) is arbitrary.

Part B

Each of the links (X Must-Parent Y) shows up as a 1 in the Parent(X,Y) i-link of the pattern. Similarly, the (X Must-be Human) shows up as a 1 in the Is-a(X,Human) i-link of the pattern. All the other entries in the pattern are don't cares (*).

Part C

Note that S keeps track of bit positions where all positive samples agree; any bit in which positive samples disagree is set to don't care (*). G keeps track of the Must and Must-not constraints, that is, the required values in bit positions where disagreement causes failure. Note that we need two vectors so as to differentiate between features that are in common among all the positives but don't matter from those features that actually explain why some negative sample failed.

With this understanding, we see that negative samples that match G do not violate any of the Must and Must-not conditions on the evolving model. So all the even choices (2, 4, 6) are not near-misses, since they do not match G, that is, they violate the existing model – they are contradictions. All the odd cases (1, 3, 5), matching G but different from S may represent near-misses. In fact, 1 and 3 are definitely near-misses. 5 is a near miss for the particular set of link types in this problem (Parent, Is-a) since the only way we can get multiple differences involving the same node is with the Is-a links. If we had a more elaborate class hierarchy, we would get a large number of correlated changes in the Is-a features of a sample and these all represent one "conceptual" difference between the samples.

Part D

1. **False** The Hamming distance is equivalent to the squared Euclidean distance.
2. **False** Nearest Neighbors predictions do not depend on order of presentation of the samples since the samples are simply stored in memory upon presentation.
3. **False** The more (likely irrelevant) features in the feature vectors the worse that Nearest Neighbor will perform since the differences on the irrelevant features will swamp differences on the relevant attributes.
4. **False** Normalizing using the average and standard deviation of feature values is relevant for values that have a Gaussian distribution, which is definitely not the case for binary variables.

Part E

Ideally, the tree should have a single long path leading to a leaf with all the positive samples with single branches going off to groups of negative samples.

In this tree, a 0 in the pattern shows up as a branch that tests the corresponding i-link and goes to a leaf with negative samples if the i-link is present in the sample and goes on to test another i-link if it is absent in the sample. Conversely, a 1 in the pattern shows up as a test for the presence of an i-link in the samples. A * in the pattern says that the i-link is irrelevant and it does not show up in the tree.

In practice, there may be spurious correlations in the data that may prevent us from finding this ideal tree. In particular, in this data set, for most of the Cousins, A Is-a Female, and this is not true for any not-Cousins. This might introduce some non-ideal branches in the tree.

Part F

1. **False** Large initial weights make it likely that some units in the neural net will saturate. In that case, the derivative of the unit's output with respect to changes in the weights is zero and no learning can take place.
2. **False** Backpropagation uses gradient descent to minimize the sum squared error. But this is a local minimization that is not guaranteed to find the global minimum except for very simple nets.

Problem 5 Miscellaneous

Some of these are based on guest lectures last year that were not repeated this year.

1. **False** The hippocampus seems to be involved in the transfer of short-term memory to long-term memory.
2. **False** There is evidence that during sleep the hippocampus "reviews" recent memories, possibly to consolidate them into long-term memory.
3. **True** The conflict resolution strategy determines which rule instance is fired next.
4. **True** The rete algorithms speeds up forward chaining.
5. **True** Alpha-Beta is guaranteed to choose the same move as straight mini-max, when operating on the same game tree.
6. **False** Beam search does not find optimal solutions since it discards most of the potential paths along the way.
7. **False** Three (independent) views of a transparent polyhedron are sufficient to predict its appearance from another view.

8. **False** The template approach does not require knowing orientation of the objects, but it does require the ability to match corresponding points on the two objects.
9. **False** Genetic algorithms can get caught in local maxima, most commonly because the population loses diversity i.e. one genotype dominates the entire population.
10. **False** Increasing selective pressure causes the GA to converge more quickly and makes it less likely that a GA will find the true global maximum.
11. **False** Children do not simply imitate; they infer rules based on the evidence.
12. **False** Thematic role frames try to capture the roles that objects, described in a sentence's noun phrases, play in the action specified by the sentence's verb.
13. **True** Thematic role frames capture the relations between nouns and verbs at the surface level while primitive act frames attempt to get at the underlying meaning of action words.
14. **True** Transition space views the world in terms of changes rather than in terms of states.
15. **True** As in PS 10.