

Mini Project 4

Jerome Francis

jfrancis62@gatech.edu

1 HOW DOES YOUR AGENT WORK? DOES IT USE SOME CONCEPTS COVERED IN OUR COURSE? OR SOME OTHER APPROACH?

From a high level, my agent is just a classification agent. It starts with 3 objects: *is_monster*, *not_monster*, and *can_be_monster*. Each object is a dictionary with a key corresponding to a monster characteristic. Each key maps to a set (initialized to be empty).

We then loop through all the labeled monsters.

For each monster, we loop through all the characteristics and the corresponding value.

If the label was POSITIVE, we follow this logic:

- Check if the value of the characteristic for this monster was previously seen
 - If it was not, add it to the characteristic in *is_monster*
 - If it was seen, check *not_monster* if it is there
 - If it is in *not_monster*, remove it from that set and add it to *can_be_monster*
 - If it's in either of the other 2 groupings, don't do anything

We follow similar, but converse logic if the label was NEGATIVE:

- Check if the value of the characteristic for this monster was previously seen
 - If it was not, add it to the characteristic in *not_monster*
 - If it was seen, check *is_monster* if it is there
 - If it is in *is_monster*, remove it from that set and add it to *can_be_monster*
 - If it's in either of the other 2 groupings, don't do anything

Once we've done this for all possible values for each labeled creature, we then proceed to classify the unknown monster.

My agent starts by assuming the unlabeled creature is the specified type of monster by assigning a boolean to *True*. My agent loops through each characteristic and sees if the value for that characteristic is contained in *not_monster*. We do a logical *AND* between the tracking value above and the outcome that the current value is NOT in *not_monster[characteristic]*. Once we finish looping, we return the

tracking value.

A big thing to note here is that if a value has not been seen, it is effectively treated as a characteristic the monster could exhibit, because a value we haven't seen will not be in any of the sets.

2 HOW WELL DOES YOUR AGENT PERFORM? DOES IT STRUGGLE ON ANY PARTICULAR CASES?

My agent successfully solved all 40 of the Gradescope cases.

3 HOW EFFICIENT IS YOUR AGENT? HOW DOES ITS PERFORMANCE CHANGE AS THE NUMBER OF LABELED MONSTERS GROWS?

For each of the variables *is_monster*, *not_monster*, and *can_be_monster*, we keep a set of values for each characteristic. We perform set adds, set removes, map lookups, and set contain checks on these sets, which are all $O(1)$ constant time operations.

When I do the classifications, I loop through n creatures, with m characteristics each. Therefore, the classification is an $O(mn)$ operation.

Classifying the unknown creatures requires a follow-up loop through m characteristics, doing a constant time set lookup, so we have an $O(m)$.

Adding the two and reducing gets me a simple $O(mn)$ time complexity algorithm.

4 HOW DOES YOUR AGENT COMPARE TO A HUMAN? DOES YOUR AGENT SOLVE THE PROBLEM THE SAME WAY YOU WOULD?

My agent works somewhat similarly to how a human would. The one difference I could think a human would do is maybe use a case-based reasoning approach while classifying to save some time.

Nonetheless, my agent does look for characteristics that a POSITIVE example must have or could have, and look for characteristics that a POSITIVE example never has. This is very similar to what a human would do if they couldn't find a matching case. From there, both my agent and human see if any of the unlabeled creatures are characteristics a POSITIVE example would never have, and say "No" if that is the case.

Humans may take a different approach to values for characteristics they didn't

see at all in the labeled examples. Some may give the unseen value the benefit of the doubt and treat it as a value that could be seen, but others could treat the value as a characteristic the creature cannot have because it was not confirmed. Since both are valid approaches with their merits and drawbacks, I decided for my agent to choose the benefit of the doubt, since my logic was defined that a single negative characteristic would be grounds for rejection.

I believe a human would normally follow similar thinking since there isn't supporting data rejecting a POSITIVE classification.