Project 3

Alonso Torres-Hotrum
atorreshotrum3@gatech.edu

Introduction

In this project, I will attempt to construct an AI agent that can solve advanced 3x3 Raven's Progressive Matrix problems using only visual data. The agents from both previous projects used only the verbal information provided, so neither agent will be able to be used as-is. I have decided to start from scratch with a completely different approach that was found while researching visual solutions to 3x3 RPM problems.

The new approach will be based on Agent 3 from *Using Human Computation to Acquire Novel Methods for Addressing Visual Analogy Problems on Intelligence Tests* (Joyner et al 2015). The agent described in the paper utilizes both dark pixel ratios (DPR) and intersection pixel ratios (IPR) in order to generate a best guess from the answer choices. The paper is somewhat vague in its definition of these two terms, which leaves room for experimentation of different interpretations. While this method does not closely follow the human approach to these problems, it correctly answered nearly 83% of the RPMs it was tested against.

If Agent 3 fails to provide adequate results, Agent 1 from the same paper will be adapted for use with my Project 2 solution. In this agent's knowledge representation "each frame in an RPM consists of objects, and each object consists of the following attributes: shape, size, fill, rotation, and relative-position to other shapes" (Joyner et al 2015). These attributes are exactly the same as those provided in the verbal descriptions of previous projects, and thus this knowledge representation can be applied to the previous solution in order to solve the new problem sets.

In the case that Agent 3 results in limited success, Agent 4 from the same paper can be implemented. This agent combines elements of Agent 1 & 3 by having the program look for visual relationships and transformations before implementing the DPR/IPR approach. This would allow for the Agent 3 solution to be combined with a method for quickly recognizing visual patterns, likely increasing the agent's overall effectiveness.

Submissions 1

Submitted on 2019-07-16 at 00:25:14 UTC.

Changes

For the first iteration of this agent, the main focus was to implement a working dark pixel ratio function. I interpret the dark pixel ratio to be:

$$DPR = \frac{\textit{dark pixels in A}}{\textit{total number of pixels in A}} - \frac{\textit{dark pixels in B}}{\textit{total number of pixels in B}}$$

In order to address the ambiguity of grey pixels, I developed a darkness threshold for what values were considered black.

The agent begins by removing the fourth layer of each pixel known as the transparency value, which allows the image support transparency. It then counts the number of dark pixels in each figure by iterating through every pixel, and if any of the values in the pixel are below the darkness threshold the pixel is considered black.

Next, the agent finds the DPR for each adjacent figure in the problem, including diagonals, and stores them in a dictionary. It then loops through all of the answer choices and finds the DPR between the answers and their surrounding figures (F, E, and H). The answer's DPRs are then compared to the appropriate DPRs from the problem figures (H is compared to other horizontal DPRs, E is compared to other diagonal DPRs, etc.) and if the difference between the values is below a specified similarity threshold a vote is cast in favor of that answer. The answer with the highest number of votes is assumed to be the correct answer.

The success of this implementation relies heavily on both the threshold of the dark pixel value and the similarity threshold. There are large variations of pixel values in the image, so defining what qualifies as a dark pixel has a significant effect on the dark pixel ratio. Likewise, when comparing ratios for similarity, the threshold at which two ratios are considered equivalent heavily impacts the agent's guess. Thus, both values will need to be optimized in order to provide the maximum number of correct answers. For this initial submission I have chosen threshold values that optimize the score of Basic D problems.

Human Similarity

This agent uses the least human-like approach to RPMs of any agent previously developed for this project track. It could be argued that the DPR is a way that humans initially check for the addition or removal of objects. However, it does not capture any of the pattern recognition methods people use to solve these problems.

Performance

This agent was able to achieve the following scores:

Basic D	Test D	Basic E	Test E
8	4	5	3

Table 1. A chart of scores for Submission 1.

The runtime of 1693.87 seconds (28.23 minutes) was exceedingly slow and was due to the lack of implementation of NumPy functions. This iteration of the agent goes through every pixel and checks each value, which results in an overall Big O value of $O(n^4)$. Future efforts will include research into ways of making the agent more efficient in its handling of large arrays without the use of nested loops.

The agent performed far better on Problem Set D than it did on Problem Set E, which is to be expected as the threshold values were set to optimize performance on D problems. The fact that the agent did not perform well on Test D problems shows that it is currently too specialized in its threshold values. Future efforts should work to find the optimum balance of both the dark pixel and similarity thresholds so that the agent maximizes its correct answer for all problem sets. The agent does not perform well on problems where multiple answers have the correct number of dark pixels or those that rely heavily on pattern recognition. This indicates that methods for pattern recognition will likely need to be developed in future iterations.

Submission 2

Submitted on 2019-07-16 at 03:52:54 UTC.

Changes

This version of the agent focused heavily on improving its efficiency through the implementation of NumPy functions. The first change was to have the images processed using Pillow's “convert” method, which changes the pixel values to grayscale mode (L) by taking the following ratio of the red, green, and blue pixel values:

$$L = R * 299/1000 + G * 587/1000 + B * 114/1000$$

The resulting array is two dimensional, meaning that it can be more quickly iterated through than the original 3D array.

The method for calculating the dark pixel ratio was also modified to operate more efficiently. Instead of using nested for-loops, NumPy's "sum" method was invoked to count the number of array values that fell below the darkness threshold. The final change was to set the darkness threshold equal to a value that was found to maximize results for all problem sets in local testing.

Human Similarity

The agent's approach has not changed, so it remains quite dissimilar to the human approach to RPMs. This iteration mainly focused on improving the agent's efficiency, the methodology was not changed at all.

Performance

This agent was able to achieve the following scores:

Basic D	Test D	Basic E	Test E
8 (+0)	6 (+2)	7 (+2)	1 (-2)

Table 2. A chart of scores for Submission 2. The change from the previous score is in parentheses.

Interestingly, the change in the similarity threshold did increase the agent’s score in the Test D and Basic E sets, but decreased the score for Test E. This means that while the agent has become slightly less specialized, future efforts should still focus on generalization. The agent still does not perform well on problems where multiple answers have the correct number of dark pixels or those that involve pattern recognition. It does now perform fairly well on the majority of problems that can be solved using DPR, so the next step will be to begin implementing production rules. The agent's efficiency dramatically increased, with a new Big O value of $O(n^2)$ and an execution time of 6.497 seconds.

Submission 3

Submitted on 2019-07-17 at 16:51:10 UTC.

Changes

This iteration of the agent includes the first implementation of a production rule, which stems from the observation that many problems in Problem Set E included progressive unions. This means that there would either be a horizontal union ($A \cup B = C$) or a vertical union ($A \cup D = G$) that could be used to generate a guess for the answer. If the agent recognizes that either of these situations is present in an RPM, it immediately searches for the answer that best completes the appropriate horizontal/vertical union and selects it.

Human Similarity

The addition of this production rule does make the agent's approach more human-like in that when people recognize union patterns they can usually apply that pattern to generate a guess the same way the agent does. However, this production rule applies only in a small number of cases, so the agent's overall approach is still not very similar to the human approach to RPMs.

Performance

This agent was able to achieve the following scores:

Basic D	Test D	Basic E	Test E
8 (+0)	6 (+0)	8 (+1)	4 (+3)

Table 3. A chart of scores for Submission 3. The change from the previous score is in parentheses.

As expected, the agent's performance increased in the E problem sets and did not change for the D problem sets. The agent now performs well on problems that can be solved with DPR and those that involve union patterns, and does poorly on all others. The agent performed slightly more efficiently, with a reduction in execution time of nearly half a second.

Submission 4

Submitted on 2019-07-17 at 23:00:09 UTC.

Changes

In this revision another production rule was added in order to be able to deal with subtraction transformations (A ∩ !B = C), which appear in Problem Set E. This was difficult to execute because the intersection between two images could not be computed due to misalignment of the figures (see Figure 1). The solution was to have the agent check if the difference in dark pixels between figures (A-B = C) and if a pattern of subtraction was found, the agent finds the answer that best fits that pattern.

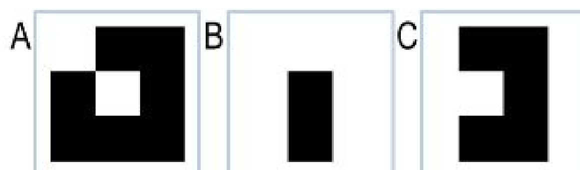

Figure 1. An extract from Basic E Problem 4. Notice that subtraction takes place, but that the dark pixels in B do not align with the corresponding pixels in A.

Human Similarity

Once again, the addition of this production rule does make the agent's approach more human-like in that when people recognize subtraction patterns and generate a guess the same way the agent does. The agent's overall approach is still not very similar to the human approach to RPMs as it relies mostly on DPR.

Performance

This agent was able to achieve the following scores:

Basic D	Test D	Basic E	Test E
8 (+0)	4 (-2)	9 (+1)	6 (+2)

Table 4. A chart of scores for Submission 4. The change from the previous score is in parentheses.

The agent's performance increased slightly in the E problem sets but decreased with Test D problems. The reason for this is unclear and will require some investigation, as the addition of this production rule should have no effect on Problem Set D. The agent now does better on problems that involve subtraction and has maintained its performance on all other problem types.

The Big O value remains at $O(n^2)$ and its execution time is now up to 6.972, an increase of almost a full second longer than the previous submission.

Submission 5

Submitted on 2019-07-17 on 23:31:14 UTC.

Changes

After investigating the agent's unexpected drop in Test D performance I found that I had changed the similarity threshold to optimize Problem Set E performance and never changed it back. For the sake of consistency, I decided to resubmit the code with the only change being the use of the threshold value that has been used for the majority of this project. This will allow me to decide if the changes in performance seen in Submission 4 were due to the addition of the production rule or the change in threshold value.

Human Similarity

This change has no impact on human similarity. The agent still solves problems using human-like production rules, but the majority of its problem solving methods do not follow the human approach.

Performance

This agent was able to achieve the following scores:

Basic D	Test D	Basic E	Test E
8 (+0)	6 (+2)	10 (+1)	5 (-1)

Table 5. A chart of scores for Submission 5. The change from the previous score is in parentheses.

The Test D score is now back to what I expected it to be, meaning that the production rule did not affect Problem Set D problems. Interestingly, the change in similarity threshold had a positive effect on the Basic E problems but a negative effect on Test E problems. This confirms that while some threshold values may cause the agent to perform better on some problems sets, the current value is optimum for performance across all problem sets. There was no change in efficiency.

Submission 6

Submitted on 2019-07-20 at 20:57:20 UTC.

Changes

This iteration of the agent aims to incorporate the intersection pixel ratio (IPR) to supplement the DPR calculations. The IPR is calculated as:

$$IPR = \frac{A \cap B}{(A \cup D) + (A \cap B)}$$

The agent now checks the IPR between figures in the same way it does the DPR, then chooses the answer with the highest similarity in both IPR and DPR.

Human Similarity

The incorporation of the IPR does not reflect the human approach to RPMs at all in that people do not calculate the IPR between figures. At this point the agent only reflects human thinking through its use of production rules, which are only utilized in Problem Set E.

Performance

This agent was able to achieve the following scores:

Basic D	Test D	Basic E	Test E
11 (+3)	6 (+0)	10 (+0)	6 (+1)

Table 6. A chart of scores for Submission 6. The change from the previous score is in parentheses.

These are the best scores the agent has been able to achieve for each of the problem sets. While the Basic sets are now able to score highly, the Test sets cannot seem to get over the 50% threshold. This likely means that the agent is too specialized towards the Basic problems, and future efforts will need to aim towards more generalization. The execution time was much higher at 9.55 seconds, which makes sense as the number of calculations has doubled.

Submission 7

Submitted on 2019-07-21 at 18:33:31 UTC.

Changes

This version of the agent aims to explore another interpretation of the DPR equation, where

$$DPR = \frac{\textit{dark pixels in A} - \textit{dark pixels in B}}{\textit{total number of pixels in A} + \textit{total number of pixels in B}}$$

Local testing has shown that this equation has the ability to increase the agent's performance on Problem Set E when coupled with the appropriate threshold parameters. However, these parameters have been shown to have a negative effect on Problem Set D performance. In order to test the effect of maximizing Problem Set E performance, the new equation and parameters will be used.

Human Similarity

This change has no impact on human similarity. The agent still reflects human-like problem solving when utilizing its production rules, but its use of IPR and DPR do not follow a human approach.

Performance

This agent was able to achieve the following scores:

Basic D	Test D	Basic E	Test E
7 (-4)	8 (+2)	11 (+1)	5 (-1)

Table 7. A chart of scores for Submission 7. The change from the previous score is in parentheses.

The agent performs almost perfectly on Basic E problems, but struggles with higher level Basic D problems, which often involve complex diagonal pattern recognition. Unexpectedly, the agent performed better on the Test D problem set and worse on the Test E problem set. This leads me to believe that each problem set has a set of optimum parameters, and that generalization rather than specialization is the only way to improve all problems set scores. The agent's efficiency had no significant improvement.

Submission 8

Submitted on 2019-07-23 at 00:26:15 UTC.

Changes

In reviewing the Basic E problems I noticed that several included some version of an intersection transformation (i.e. $A \cap B = C$). I decided to create a production rule that checks for either horizontal or vertical intersection transformations and finds an answer that fits accordingly.

Human Similarity

This addition makes the agent's approach more human-like because it utilizes production rules in the same way a person would. The agent still uses DPR/IPR calculations to solve the majority of problems, so for the most part it still does not follow the human approach to solving RPMs.

Performance

This agent was able to achieve the following scores:

Basic D	Test D	Basic E	Test E
7 (+0)	7 (-1)	11 (+0)	5 (+0)

Table 8. A chart of scores for Submission 8. The change from the previous score is in parentheses.

Surprisingly, the agent's performance decreased on Test D and stayed the same for all other problem sets. This leads me to believe that the DPR/IPR was already answering any intersection questions correctly. Furthermore, one of the Test D problems got mistakenly labeled as an intersection problem, which can be attributed to the agent having too high of an intersection threshold. The agent is no longer performing well on higher level problems that can be solved using DPR, presumably because it is mistakenly applying production rules to them. The execution time has decreased 5.947 seconds, the lowest yet. This is likely because more production rules are being implemented, which take less time than IPR/DPR calculations.

Submission 9

Submitted on 2019-07-23 at 02:04:02 UTC.

Changes

Upon further review of the Basic E problem set, I noticed that several of the problems included an "exclusive or" transformation ($C = (A \cup B) - (A \cap B)$). Another production rule was created to have the agent check for an exclusive or transformation in both horizontal and vertical directions, then find the corresponding answer.

Human Similarity

The addition of a production rule once again makes the agent's approach more human-like because it mimics a person's use of production rules. The agent now uses production rules to solve almost all Problem Set E questions, meaning that it approaches E problems similarly to how a human would. It still solves Problem Set D questions using IPR/DPR calculations, so the agent still does not have an overall human-like approach.

Performance

This agent was able to achieve the following scores:

Basic D	Test D	Basic E	Test E
7 (+0)	5 (-2)	10 (-1)	7 (+2)

Table 9. A chart of scores for Submission 9. The change from the previous score is in parentheses.

As expected, the agent's overall Problem Set E score improved with the new production rule. Unfortunately, its score on Problem Set D continued to decrease, once again suggesting that productions rules are being mistakenly invoked and the production rule threshold needs to be lowered. The agent performed well on problems that used exclusive or transformations and still did poorly on higher level problems that can be solved using DPR. The execution time increased by over a second despite the Big O factor remaining at $O(n^2)$.

Submission 10

Submitted on 2019-07-24 at 03:35:51 UTC.

Changes

The aim for the final version of the agent is to develop the most generalized and overall high-scoring agent possible. In order to do this, I have reverted back to the original DPR equation and have selected threshold values that optimize the agent's performance in the Basic D & E problem sets. Furthermore, I added functionality for having the agent execute the IPR/DPR calculations in the case of a production rule resulting in a tie. This ensures that the agent is not guessing between options that the production rules have found to be equally likely.

Human Similarity

This iteration has made the agent's approach slightly more human-like, in that if a person found that the execution of production rules resulted in a tie they would likely also use another method as a tie-breaker. However, the agent for the most part still uses the IPR/DPR approach that is not human-like, so overall it does not reflect a human-like approach.

Performance

This agent was able to achieve the following scores:

Basic D	Test D	Basic E	Test E
11 (+4)	6 (+1)	10 (+0)	7 (+0)

Table 10. A chart of scores for Submission 10. The change from the previous score is in parentheses.

While this iteration of the agent did score higher across the board than any previous agent, it is far from a working model that performs well on all problem sets. At this point it is difficult to say which types of problems the agent is struggling with, as it is mostly having issues with the test problem sets. The agent performs well in the Basic D & E problem sets, though it does struggle with the highest difficulty problems for each set as well as any problems where multiple answers have the same number of dark pixels. The agent's execution time drastically increased to 47.962 seconds, which is likely due to the fact that the agent now more often has to go through the production calculations as well as the IPR/DPR calculations. The agent finished with a Big O value of $O(n^2)$.

Conclusion

The overall process for designing this agent was to utilize deliberate improvement when implementing the DPR and IPR functionality. From there, the production rules were introduced to target one type of problem at a time. Finally, there was a significant amount of trial and error in the tuning of parameters to maximize the agent's performance.

While it can be argued that people do use some level of IPR/DPR detection when it comes to recognizing the addition or removal of objects, this method is not generally representative of how humans solve RPM problems. The agent does utilize a number of production rules that follow the human approach of pattern detection and application. However, the agent's overall approach heavily utilizes IPR/DPR calculations in a way that is not human-like, and is therefore not very representative of the human approach to RPM problems.

Given more time, I would aim to generalize the current agent so that it can perform well on all problem sets. I would equip the agent with the ability to identify shapes so that more production rules could be implemented, shifting away from the DPR/IPR centered approach. Ideally, the agent would utilize production rules developed in previous projects and use DPR/IPR calculations only as a check or a tie-breaking feature. Furthermore, I would aim to return the agent to the level of efficiency seen in Submissions 2-9.

References

1. Joyner, D. A., Bedwell, D., Graham, C., Lemmon, W., Martinez, O., & Goel, A. K. (2015, June). Using Human Computation to Acquire Novel Methods for Addressing Visual Analogy Problems on Intelligence Tests In ICCC (pp. 23-30).