Project 3(Summer 2020)

Anshuta Awasthi
aawasthi32@gatech.edu

1 INTRODUCTION:

The goal of this project is to implement an agent to solve 3X3 Raven's Progressive Matrices on Problem sets D and E. The plan is to extend the previous implementation of 3x3 RPM in this project and build it incrementally. Agent used in PR2 is already using visual heuristics methods like Dark pixel ratios and matching pixel ratios mentioned in (Joyner, 2015). Although, I was planning to work on implementation where agent acquires the ability to separate shapes and detect shapes from the image. But initial few attempts did not work out. Due to time constraints also, I decided to go with standard ATSI method mentioned in paper (Kunda, 2013), also because I have already dom=ne similar implementation in PR2.
For start, different combinations of collinear pairs and triplets mentioned in ATSI method have been used as reference. It was observed that Problem sets E and D have different kind of transformations. Problem set E transformation images are more of a type where one image is result of binary transformations (like Union, Intersection, XOR, subtraction and back subtraction) applied between rest of the 2 images. On the other hand, problem set D is having several shapes in an image which are changing across horizontal and vertical columns.
For calculating the image similarity, the ratio of total number of equal pixel values at the same coordinates with total number of pixels between two images has been used.
The complete production system code is structuredin nested if/else code blocks.

2 JOURNAL ENTRIES

2.1 Submission 1:

Date of Submission: 2020-07-14 01:19:46 UTC
Changes in this version:
My agent in PR2 was already based on visual heuristics using Affine transformations and Dark Pixel ratio, applied on the binary 1D numpy arrays. Looking at first few problems of problem set D, the horizontal and vertical pairwise dark pixel ratio of cells was compared and closest match was returned as per

ATSI method explained in (Kunda, 2013) paper however not all the combinations mentioned in the paper were implemented in this iteration. To get headstart in problem set E and verify its baseline performance, I first applied binary transformation on numpy arrays to get the resulting image array. Only AND, XOR and OR operations are applied between 2 pair of images in a row(and column), and the generated image is then compared against third one in the row(e.g. C = A AND B, B = A AND C, A = B AND C) . After verification of the transformation in next row, the resulting image is obtained by applying the transformation and then comparing against all the answer choices as per "Generate and Test" approach. The closest match is found by selecting the option which has maximum no. of dark pixels present at the same coordinates with the generated image.

Agent Vs. Human Approach:

The agent is performing in 2 ways - For problem set E, it is more similar to the way humans approach the problem, while problem set D approach is based on pixel ratios and hence is more computative.

For solving problems in E set, Agent is looking for best binary transformation (like AND, OR ,XOR etc.) between the pair of 2 images. This is how exactly I also solved the problems in my head by checking the result of overlapping of two images in different ways. For Problem set D, Agent is approaching the problem by comparing and calculating the number of dark pixels in each image. At implementation level ,this method is different from the way humans look at the problems. Humans can easily spot the difference in shapes, size, dark and white areas and numbers of figures in the image. While agent is using the aggregate of dark pixels and comparing the ratios, which is not needed for humans to establish the pattern.

Outcome ,Efficiency and Performance:

Execution Time : 21.152 Secs

Table 1—Autograder Results

	Basic	**Test**	**Raven's**	**Challenge**
Problem Set D	5/12	3/12	4/12	2/12
Problem Set E	5/12	5/12	1/12	5/12

With both the problem sets on the basic score of 5, I can say that the first version of the agent has set the good baseline. The test score on both the problem sets is also good and shows that the first version of agent is working on unseen problems too, hence implementation approach is on right direction.

2.2 Submission 2:

Date of Submission: 2020-07-14 22:48:15 UTC

Changes in this version:

The performance of agent in baseline version is almost equal for both the Problem sets. But now onwards, my agent is targeting to improve 1 problem set at a time to see the effect of separate code changes on the performance of agent on each problem set.

In this iteration, I have tried to include the dark pixel ratios on pairwise cells in the diagonal and inverse diagonal directions. The diagram displayed in the ATSI method in (Kunda, 2013) paper is very helpful to see more pairs and corresponding cells in different diagonals.

Agent Vs. Human Approach:

The agent in this iteration is trying to establish the pattern in diagonal directions.I feel it is more difficult for humans to establish the diagonally oriented pattern, since only 1 diagonal and inverse diagonal is clearly visible in 3x3 matrix. The Figure 2(taken from the (Kunda, 2013)) has been very helpful in visualizing the diagonal relationships among other cells which are not visibly located in diagonal in the 3x3 form of matrix. I must say, agent is performing better than

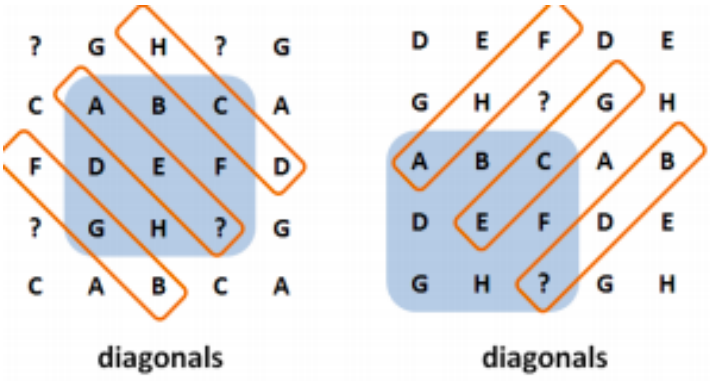

Figure 1—Diagonals

average human mind, when it comes to diagonal pattern. However fundamental difference where the agent is calculating pixel ratios while humans find differences by just looking at the shapes, number of sides of shapes, color, orientation, still remain the same

Outcome ,Efficiency and Performance:

Execution Time : 22.152 Secs

Table 2—Autograder Results

	Basic	Test	Raven's	Challenge
Problem Set D	7/12	3/12	4/12	2/12
Problem Set E	5/12	5/12	2/12	5/12

The resulting Agent improved on basic problem result for problem set D and shows no significant difference in set E score which is expected, since the agent was only improvised on Problem set D types of problem. This iteration helped me to achieve the correct 7 basic test cases milestone for problem set D. However it did not bring any change on Test score of set D, which was disappointing, since the change in this iteration is very generic and attempts to solve diagonal patterns. I was expecting it to solve few diagonal relationship problems in test set D too.

2.3 Submission 3:

Date of Submission: 2020-07-16 13:08:46 UTC
Changes in this version:
The main target in this iteration is to improve Problem Set E results by including more binary transformations between the numpy arrays of two images. I have implemented subtraction and back subtraction operations too along with AND, XOR and OR operations before. But when performing pixel wise transformations, I was having difficulty visualizing the resulting image. I changed my strategy and applied transformation on images directly using logical_or , logical_and,logical_xor,multiply and invert operation using ImageChops module of PIL library and see the resulting image on display. The generated image is then again pixel wise compared with answer choices by selecting the one with maximum number of matching pixel values on same coordinates.
Agent Vs. Human Approach:
The approach for solving problem set E is very similar to the way humans will perceive this. But the agent usually underperform on such problems due to slight variation of image alignment, shifts and size which affects the pixel calculation based results. Human brain mostly ignores the very slight shift or very slight changes in size of image .It focuses more on the shape of images. On the other hand slight change in image position can give very different results when agent is doing pixel wise comparison and that change accumulates when aggregation is applied on matching pixel calculations. ***Outcome ,Efficiency and***

Performance:

Execution Time :22.235 Secs

Table 3—Autograder Results

	Basic	Test	Raven's	Challenge
Problem Set D	8/12	3/12	4/12	1/12
Problem Set E	9/12	6/12	1/12	2/12

The above change improved the problem set E score and in basic and Test , but dropped in challenge set. The main difficulty agent was facing was to select which binary transformation among all best fits the problem. This is due to slight misalignment of images. The resulting image after binary transform on 2 images does not exactly fits the third image (although it looks the same with human eyes) and hence makes the comparison mostly threshold based. Fixing the threshold value solves the specific problem in basic test cases but results in overfitting agent, which does not perform well on unseen problems.

2.4 Submission 4:

Date of Submission: 2020-07-16 18:05:45 UTC

Changes in this version:

Since agent is very close to hit the passing milestone in test set of problem E, I decided to make it more generic and fix the issues encountered in previous iteration in order to improve its performance on unseen test. The Problem set E-03 was little tricky to solve, since looking at ABC and DEF patterns it gives an impression that C= A AND B is the right transformation and verifies it in F = D AND B. While if same transformation is applied on G and H , the resulting image do not match any of the criteria .The reason being, first two rows do not have overlapping dark pixels in the 2 cells as shown in the figure. The agent was improved by including the scenario, where more than one transformation satisfies the criteria the resulting images from all the applicable transformations will be compared with all the answer choices and the best match(where maximum no. of pixels at the same coordinates match) between the two images would be returned.

Agent Vs. Human Approach:

The problem described above is so easy for humans to solve. But the agent falls in the trap of choosing the correct binary transformation and this problem deceives the agent's generate and test approach. The first row establishes

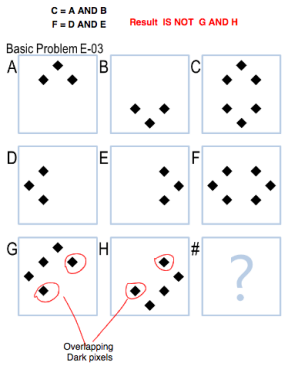

Figure 2—Basic Problem E-03

the "AND" transformation, second row verifies the "AND" transformation , but resulting value after applying the "AND" transformation gives the wrong results(has one redundant pair of dark pixels which were overlapping). While the spatial sense of shapes placement and symmetry makes this problem very easy for humans to solve.

Outcome ,Efficiency and Performance:

Execution Time :22.236 Secs

Table 4—Autograder Results

	Basic	**Test**	**Raven's**	**Challenge**
Problem Set D	8/12	3/12	4/12	1/12
Problem Set E	10/12	7/12	2/12	2/12

The change helped me achieve the milestone of passing 7 cases in test set of Problem Set E. It cannot be said that agent is generalized , as the score in Raven's and Challenge problems is very low.

2.5 Submission 5:

Date of Submission: 2020-07-17 00:05:04 UTC

Changes in this version:

Since the agent is performing well on Problem Set E, I shifted my complete focus to solve as many problems as possible in Problem set D. While pairwise DPR ratio approach was working well, few of the problems were still failing either due to comparison of threshold values or the agent falling into wrong if/else block. While trying to solve problem D-10, an interesting pattern was discovered that total no. of dark pixels in all the 3 cells across row or column(or diagonal) looks

almost same. In the problems like D-10 where the outer and inner shapes both are changing simultaneously across horizontal, vertical and diagonal directions, it was found that there is a pattern such that in one direction (either horizontally, vertically or diagonally) all 3 cells have different inner and outer images and similarly for other rows the same shapes exist in different combination in each cell and in different order. As shown in the figure,for such cases, the sum of dark pixels of all the cells horizontally(or vertically , or diagonally) should be very close to sum of the dark pixels of all the cells in next row(or column, or diagonal) as shown in the image below D-06. Surprisingly this approach works for almost all the problems in Basic set D After adjusting thresholds and little

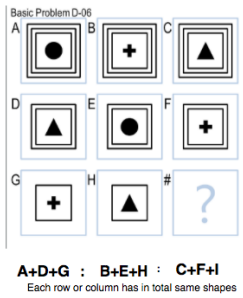

Figure 3—Basic Problem D-06

tweaking, this method worked wonders for almost all the problems in test Set of problem D

Agent Vs. Human Approach:

With 11 basic cases passed in problem set D, Agent is still not compatible to humans. Agent is very sensitive to small noise present in images,which causes it to work on specific threshold values.This may or may not work on unseen problems. However, this is not the case with humans. Human mind sees the shape and pattern at very high level and tends to ignore slight misalignment of images. This noise, (although very small) has an impact on the calculation and comparison of pixel values by agent but human approach is unaffected by the presence of small noises.

Outcome ,Efficiency and Performance:

Execution Time :22.228 Secs

The approach mentioned above boosted the basic problem set D score to 11 , while I observed very little improvement in test set score of D. It was still at 4. All basic tests in set D are passing except D-08 Having implemented most of the pair wise combinations and different comparisons, I was kind of hitting

Table 5—Autograder Results

	Basic	Test	Raven's	Challenge
Problem Set D	11/12	4/12	4/12	2/12
Problem Set E	10/12	7/12	2/12	2/12

roadblock at this point.The solution is clearly overfitting for both the problem sets.

2.6 Submission 6:

Date of Submission: 2020-07-17 00:05:04 UTC
Changes in this version:
In the method implemented in previous iteration explained above, I had missed including inverse diagonal relationship . I implemented the similar logic for inverse diagonal (i.e. F+H+A:G+C+E::G+c+E:B+D+?) in this iteration. Also, changed the hierarchy of nested if/else blocks in order to give more priority to Set D type of problems.
Agent Vs. Human Approach:
The same reasoning for human vs. agent approach described above applies here too. The only difference is, unlike horizontal and vertical patterns, "inverse diagonal" relationships are difficult to spot easily for humans too. Other than that same difference exists in both(agents and humans) of them as we described earlier.
Outcome ,Efficiency and Performance:
Execution Time :22.26 Secs

Table 6—Autograder Results

	Basic	Test	Raven's	Challenge
Problem Set D	10/12	5/12	4/12	2/12
Problem Set E	10/12	6/12	3/12	4/12

The above change got agent's score for set D improved to 5 . But in the process, few test cases which were passing earlier started failing. Test score of problem set E is down to 6 now. But that is expected since Problem set E has already achieved the minimum passing score and right now the priority is to get the minimum passing score for test set D.

2.7 Submission 7:

Date of Submission: 2020-07-17 17:24:01 UTC

Changes in this version:

In a desperate attempt to get the 7 cases passed in tests set D , I was looking to generalize the agent more. So far I had been comparing image similarity as the ratio of total number of equal pixel values at the same coordinates divided by total number of pixels between two images. This criteria for calculating image similarity may fail if images are shifted a bit and noise gets accumulated over aggregation.To overcome this issue, I introduced an extra filter to compare standard deviation of pixels between the two images.

Additionally, it was observed in gradescope results that few cases were giving skipped result for not falling under any if/else block and agent was returning -1. For that I tried to include the final block where threshold values were very relaxed, in order to catch any left out cases.

Agent Vs. Human Approach:

Apart from other similarities and differences explained above between human Vs agent approach , the agent now is also similar to how humans attempt a problem. In case none of the answer choices match we tend to select a choice which we think matches most closely instead of leaving the question unattempted. Similar way Agent is not returning "skipped" results for any of the cases and trying to return the closest match.

Outcome ,Efficiency and Performance:

Execution Time :25.096 Secs

Table 7—Autograder Results

	Basic	**Test**	**Raven's**	**Challenge**
Problem Set D	11/12	4/12	4/12	2/12
Problem Set E	10/12	7/12	2/12	2/12

Well, the hack did no good to overall performance of my agent and I was getting results almost similar to previous iterations.

2.8 Submission 8:

Date of Submission: 2020-07-17 23:28:38 UTC

This submission resulted in following error in bonnie:

Error": "Execution": "Your code caused RavensProject to crash." "
Although, it was not throwing any error on my local machine , but I was at fault here for not running error check before submitting.

2.9 Submission 9:

Date of Submission: 2020-07-17 23:32:07 UTC
Changes in this version:
As we had been told that test set problems are more or like similar to basic set. Getting 11 in basic set had left me clueless on where to improve. Any attempt to solve D-08(the only problem left in basic D set) was getting my other cases failing due to conflict in threshold values. Any changes in code were not helping me to get 7 passed in test set D. By this time agent file has all spaghetti code in many nested, related and unrelated if/else blocks which was causing me more confusion. The unnecessary code blocks where the code was very specific to previous project problems and was no good in Project3 has been cleaned up, suspecting some of the cases might be falling into these code blocks which may be resulting in error.
Agent Vs. Human Approach:
Unlike humans, Agent need to have clear path available in order to head in right direction to solve specific problems. Cleaning up the unnecessary code removed the possibility of agent to head in wrong direction. On the other hand humans have metacognitive skills to evaluate the self progress towards the completion of the task and ability to look ahead to get the bigger picture of problem.
Outcome ,Efficiency and Performance:
Execution Time :25.096 Secs

Table 8—Autograder Results

	Basic	Test	Raven's	Challenge
Problem Set D	8/12	5/12	5/12	2/12
Problem Set E	6/12	2/12	2/12	3/12

The Problem set E score has definitely suffered on removing some of the code blocks but that was intentional, in order to specialize the code to solve one type of problem(i.e. set D). This change may not have helped me achieve my desired goal, but I could see that it helped me solve 1 additional unseen Raven's problems taking the raven's score on problem set D to 5. I moved to one step closer

to solution which is more specific to cater set D type problems and more generic to handle the unseen set D type problem. Efficiency wise agent is doing OK and almost consistent across iterations.

2.10 Submission 10:

Date of Submission: 2020-07-18 11:37:33 UTC
Changes in this version:
By this time I was losing patience. This was my last attempt, so I made the Agent only for problem set D, and all other code was removed. All the if else block conditions refined to make clear demarcation of each code block. Relaxed the thresholds and made them more uniform, rather than varying thresholds which were adjusted to fit the problems in basic set. I was ok with losing few basic set problems in order to have more generic solution.
Agent Vs. Human Approach:
As said before, the big drawback Agent is experiencing in solving such problems is, its precision and high dependency on pixel calculation results. It is not able to sustain the minor noise or small shifts in pixels. While humans ability to figure out patterns and differences is immune to such minor noise and misalignment.
Outcome ,Efficiency and Performance:
Execution Time :16.338 Secs

Table 9—Autograder Results

	Basic	**Test**	**Raven's**	**Challenge**
Problem Set D	10/12	6/12	6/12	4/12
Problem Set E	1/12	1/12	1/12	0/12

As stated before , this agent is only targeted to solve the problem set D and hence a very poor score of problem set E should be ignored. I failed on my last attempt, to achieve the minimum expectation of passing test set D. However, the results show that my last attempt is the best generic solution for problem set D so far. With all the changes incorporated it did give me 6 cases passed not only in unseen test set but also in unseen Raven's set too. The challenge score achieved has also improved , although I did not attempt to solve any challenge problem specifically.

3 CONCLUSION:

a).Implementation Approach :

The agent was incrementally built up on the existing implementation of solving 3x3 RPMs in PR2. It has shown the continuous improvement trend on both the problem set till it hit the plateau due to Test set D. Different operations between different combinations of cells were tried and compared throughout the iterations. I started with generic implementation of ASTI algorithm with few cases described in paper (Kunda, 2013) while adding more scenarios incrementally. Later changes were done for one problem at a time for both the sets.

After achieving most of the problems correct in basic tests, my approach was mostly trial and error in order to get some problems passed in unseen test cases. I must say that tackling Problem sets E and D together needed a right balance of **Generalization and Specialization**. While we need to specialize our agent in order to tackle two separate sets of problems, the agent need to be generalized enough to have covered all the kinds of similar cases. The best score achieved in all the sets is given below:

Table 10—Best of All Sets

	Basic	**Test**	**Raven's**	**Challenge**
Problem Set D	11/12	6/12	6/12	4/12
Problem Set E	10/12	7/12	3/12	5/12

b). Cognitive Connection : I found the stark contrast between the differences and similarities between human and agent's problem solving approach. At high level, the approaches have more similarities like both are figuring out the relationships between shapes and patterns in different directions. Also, agent is learning through the series of iterations which is very similar to human cognition ability of learning incrementally.

Almost in all the problems we solved, the transformation is derived from the analogical reasoning by finding conceptual and logical analogies between the images. "Analogical Reasoning" is the core of human cognition too.

However,at the lower level, humans approach the problem based on their spatial understanding of shapes, color, alignment, symmetry and most importantly the ability to infer the pattern by just looking at different shapes transform.

The Agent's approach is very quantitative (based on pixel based calculations) and is highly precise. That is making the agent very sensitive to noise present in

the images. On the other hand, humans infer the pattern based on bigger picture and are immune to minor pixel level differences in the image.

c).Further Improvements: I feel there is a lot of scope for improvement and many important lessons learned. As someone has rightly said "We learn more from our failures and struggles than our successes." I achieved very good results in previous 2 projects using visual heuristics based approaches, but I could not utilise that much in this project. The reason being, my code for PR2 was not very scalable, structured and modular. The further addition of code related to Problem sets D and E together made it more complex. The resulting code had too many branches of if/else blocks. Later, I ended up cleaning the spaghetti code so that Agent will find a clear logical path.
The first learning is - keep the code readable, modular and structured for further improvement and better readability even though you are not tested for that.
Although there was so much help and advice available on slack and piazza from fellow students(big thanks to them), I found it very hard to leave the current approach and start on new approach from the scratch. I wasted my few attempts to keep debugging my existing code trying to fix through trial and error. I could have saved few bonnie attempts for testing new approaches.
Second learning- don't obsess over same problem and same approach too much. Leave the scope for trying new ways.
With so many methods, help and techniques available I could have tried more but I exhausted my bonnie attempts in order to get the test set D passing. Overall a very interesting and intense experience.

4 REFERENCES

[1] Joyner, D.A. (2015). "Using Human Computation to Acquire Novel Methods for Addressing Visual Analogy Problems on Intelligence Tests". In: *In Proceedings of the Sixth International Conference on Computational Creativity.* Provo, Utah.

[2] Kunda, Maithilee (2013). "Visual problem solving in Autism, Psychometrics, and AI: The case of the Raven's Progressive Matrices intelligence Test". In: Georgia Tech, Atlanta.