COSE474-2019F: Final Project Report
Question Answering Network for Physical Reasoning

Hojoon Lee [1] **Suyeong An** [1] **Hyunseung Kim** [1] **Dooho Chang** **Taewan Lim**

Department of Computer Science & Engineering, Korea University

Abstract

Despite facing unknown situations in the real world, humans have the ability to predict the sequence of events based on general physical concepts acquired through experience. Similarly, to cope with unforeseen situations in a complex world, intelligent agents must understand the physical concepts and make physical reasoning. To develop and test the agent's capability of physical reasoning abilities, we used the benchmark called PHYRE which aims to solve simple classical mechanic puzzles unveiled by Facebook(Bakhtin et al. (2019)). Though various puzzles share similar physical concepts, methods of reaching a solution varies between each puzzles. It is essential to develop an algorithm that can understand the physical concepts to develop a generalized approach, applicable to unseen, novel puzzles. In this context, we created questions through the physical interactions of PHYRE, and let the agent answer the questions during training, expecting our algorithm to be generalized.

1. Introduction

If a child sees an apple hanging from a tree or a ball falling through a downhill path, it would be undue to contemplate the subsequent events to predict it. The reason a child can intuitively guess the future is not because he has seen the same situation before, but because he can understand the fundamental physical concepts that he experienced in similar situations. But until now, intelligent agents trained by deep learning are called black box models because they focus on improving the accuracy of their predictions without focusing on how they solved the problems. So, unlike humans, modern day agents lack the ability to make predictions in a new environment based on physical concepts.

What if these black box agents expanded to the real world? Think about self-driving cars in the field under aggressive research. In the real world, unexpected situations do happen. Just like a soccer ball that is suddenly kicked out of the other side. Under unforeseen circumstances, not conditions provided during training, the existing deep learning model cannot make accurate predictions, which can lead to serious accidents. Therefore, agents predicting a future behavior without aware of physical concepts will be an unpredictable and also unreliable.

So our question is, how can we let agents understand the physical concepts? Primitive people knew how to threw stones to hunt animals and roll out large stones from cliffs even if they are not aware of any physical formula. They were able to acquire physical concepts by repeating their physical experiments and questioning what would happen in the future. Physical concepts are derived from questions and experiences, not knowledge of physical formulas.

To investigate and test our hypothesis, we used a benchmark called PHYRE that was released on Facebook to research through physical reasoning. There are many physics puzzles in PHYRE. Each puzzle of PHYRE operates in the 2D world where problem is to place one or two red balls so that the green ball touches the blue ball or the wall. The goal of the training is to guide agent understand the physical concepts, so that they can clear puzzles with minimal effort even in unseen tasks. We aim to solve this problem with the basis of deep reinforcement learning to learn the behavior to reach the goal, but also created question & answer module to enable the agent to understand the physical concept. In summary our main goal is listed as below.

- **Understand physical concepts:** Agent should understand physical concepts through question & answer module.

- **Obtain generalized model:** By understanding the physical concepts, agent's behavior should be more generalized.

2. Task

The python package, PHYRE, provided by Facebook is a two-dimensional gaming environment with a Newtonian dynamic engine with gravity and some friction. Task has initial world state and goal. The initial world state refers to the static state of obstacles and balls with a fixed position,

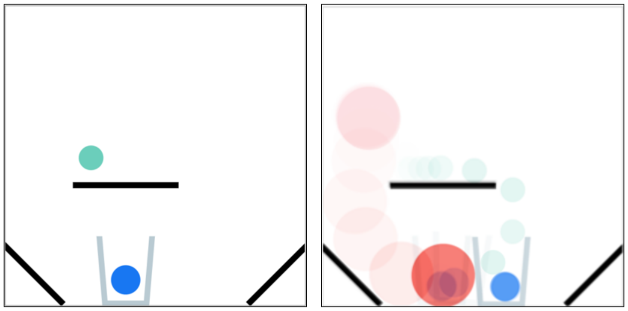

Figure 1. example of PHYRE task (left) and corresponding solution for the example(right). Black objects are static where other objects are dynamic. The task in the left pane requires placement of one ball to be solved. The right pane illustrates a solution (red ball) and the solution dynamics.

and the goal is to determine whether the red ball and the blue (bar, floor, wall, etc.) are touching each other when the all objects reached a static status. Given the initial state, agent can place a red ball on a two-dimensional plane. Some actions are invalid when the position of the ball comes out of the screen or overlaps with another object. Valid refers to all actions except for the above cases.

Benchmark. PHYRE benchmark is divided into the two tier accordance with the number of red balls that can be placed. For PHYRE-B tier, only one red ball can be placed. Action is 3-dimensional vector of continuous space where 1 dimension is for the location and the remaining are the width and height of the valid ball. PHYRE-2B consists of 6-dimensional action space similar as above, 3 dimension for each red ball. In this paper, we use PHYRE-B as our benchmark.

For each tier, 25 templates are given. Each template consists of 100 tasks where each task is quite similar. Task templates are used as two settings to evaluate agent's generalization capability: within-template and cross-template, respectively. For within-template settings, agent is trained with sub-sampled tasks in a particular template and gets evaluated through remaining tasks. For cross-template settings, templates that are used to evaluate are completely different from the templates that are used in training.

Performance measure. To evaluate agent in PHYRE, unique performance measure called **AUCCESS** is used. To measure the AUCESS score, first, we record a cumulative percentage of the success of all the test tasks for the number of trials. Next we measure the areas under the percentage curve up to the 100 trials. To give a higher weight to the smaller number of trials, for each number of trial k, $w_k = \log(k+1) - \log(k)$ is multiplied to each success percentage s_k and measuring the sample mean AUCCESS = $\sum_{k=1}^{100} w_k s_k / \sum_{k=1}^{100} w_k$.

3. Related work

Visual Question Answering (VQA). VQA is a task that produces the correct answer to a natural language question about an image, which requires a higher level of interpretability and complex reasoning compared to simple image captioning. CLEVR is a diagnostic dataset which tests an algorithm's ability to understand the structure of an image, or to perform multi-step reasoning; which is difficult to solve with existing deep learning models. (Johnson et al. (2017)). Recent research focus on utilizing feature-wise affine transformations with conditions, where images are conditioned with questions.(Perez et al. (2018)). In addition, this model offers an insight into how humans answer such questions, which allows the development of more effective models.(Lu et al. (2016)).

Learning physical representations. Learning Physical Representations aims to allow the network to derive physical concepts based on information given in the form of physical experiments. A network is trained to compress the information into a physical representation and is requested to answer the questions about the physical system. The compressed representation stores the physically relevant parameters to let network understand the physical law. A physicists' model can be created by simply answering questions through empirical data without prior knowledge. In these studies, the network was able to solve quantum mechanics and developed a heliocentric model of solar system. (Iten et al. (2018)).

4. Method

Three models were experimented; PHYRE baseline model, and two new models with question answering module.

4.1. Question Answering Module

In order to perform physical reasoning, the physical features needed to be extracted by the intermediate network.

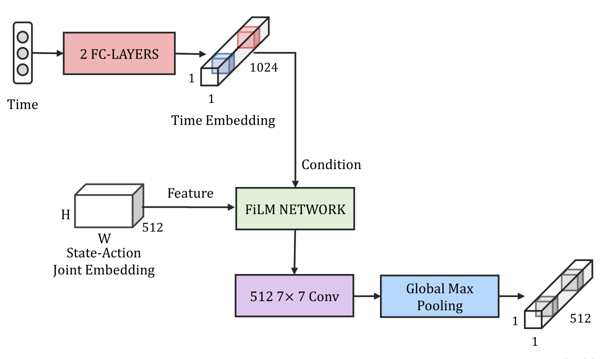

Figure 2. Question & Answering module

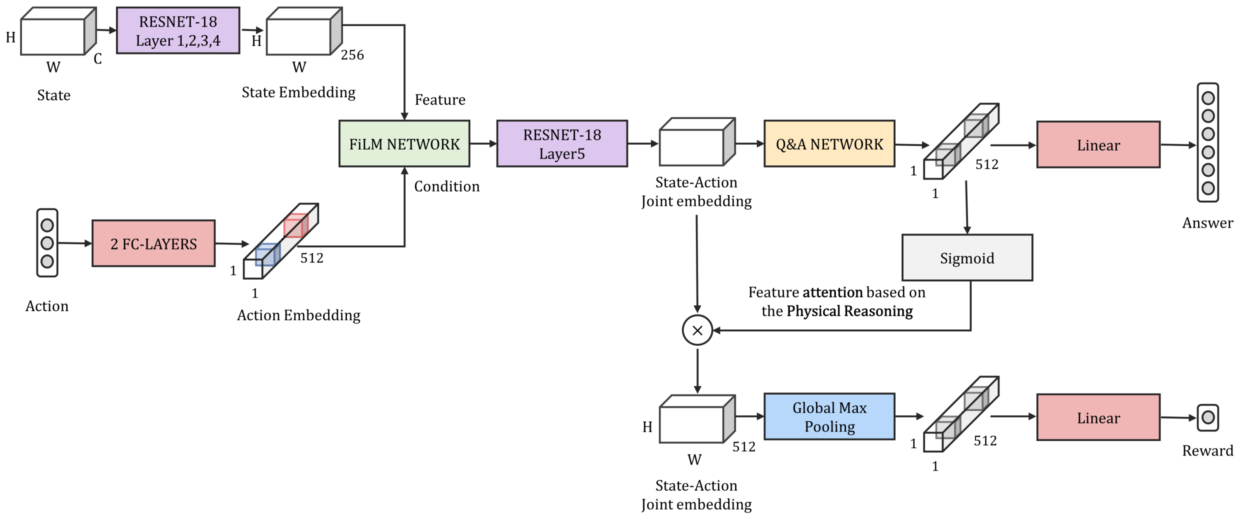

Figure 3. Attention QA Network's Architecture

Thus, the question fed to the network was modified for extracting physical features, which focused on tracking the position of objects at all times. This was the intended role of our Question Answering Module. This module has two inputs; time one hot vector, and state-action joint embedding tensor. The random entry time in one hot vector format is passed through the two fully-connected layers which converts it into time embedding vector. Then, the time embedding vector will be filmed with the state-action joint embedding tensor. The output of filming will be taken to global max pooling to create answer embedding. Through process, we expect that the module can extract physical feature of state-action joint embedding at exact time.

4.2. Baseline

Deep Q-network (DQN). The baseline model is DQN (Mnih et al., 2013; 2015) which is trained to learn the policy to maximize the expected rewards. Initial observation is fed into the first 4-layers of the resnet-18 network (He et al., 2016) to create the state embedding. Simultaneously, the three-dimensional action vector is passed through two layers of fully-connected layer, creating an embedded action vector. The feature map extracted from resnet-18 layers and the corresponding embedded action vector is passed through the FiLM Network in order to interpret the effects of the action to the initial observation. Subsequently, this FiLM Network's output is passed through the fifth layer of resnet-18 network to create a feature map where action is added. The resulting tensor of this process was named as the state-action joint embedding. This embedding tensor is passed through the global max pooling and fully-connected layer in order to predict the reward of the game. The binary cross entropy loss between the prediction and the reward was used, utilizing the philosophy of distributional reinforcement learning. (Bellemare et al. (2017))

4.3. DQN with the Question Answering module

Two-Headed Question Answering Network(THQAN). THQAN can be divided into two parts; location prediction of the balls, and reward prediction of the game. The loss function and reward prediction processes are the same with baseline DQN's. A question answering module was added as another head to predict the locations of the balls. The state-action vector embedding and the random-time one hot vector is passed into the question answering module. The agent can then extract features to answer questions about the position of the balls. Finally, the results of question answering module is passed to the fully-connecting layer to obtain a six dimensional vector for the position of the three balls. Through the above process, the layers of the baseline agent are expected to be taught to extract features to answer questions well. The loss function for the predicted locations of balls is the mean squared error.

Attention Question Answering Network(ATQAN). We hypothesized that the State-Action Joint embedding should be used to create a feed-back sequence in predicting the rewards of the games. Without such provisions, the gradients of back-propagation from question answering module, struggling to affect the reward prediction network. To expedite this feed-back sequence, the answer embedding vector is passed through a sigmoid function which is processed through the feature-wise multiplication with the state-action joint embedding to provide attention to the location relevant features.

5. Experiments

We've measured the success percentage of the test tasks for the corresponding number of trials. Afterwards, AUCCESS score was evaluated to compare model's generality and sample-efficiency. Further analysis was made to investigate whether our trained models were able to understand the basic physical concepts.

5.1. Experimental Setup

All of the hyperparameters used in training were set equally for fair comparison across the models. The model was trained end-to-end using stochastic gradient descent with the Adam optimizer (Diederik et al. (2015)). Learning rate was annealed from 0.0003 to 0 after 150,000 updates using a half cosine scheduler (Loshchilov & Hutter (2017)) without restarts. Since the number of success was significantly lower than the number of failures, we've balanced the positive and negative samples in the batches for stable training.

5.2. Evaluation

Models	Cross	Within
DQN	36.8 ±9.7	77.6 ±1.1
THQAN	-	78.4
ATQAN	**39.0**	**80.2**

Table 1. Area under the success-percentage curve (AUCCESS) of three models on PHYRE-B. For the baseline model, mean and standard deviation on the 10 folds are reported. For the others, performance for the 1 fold are reported.

In Table 1, we presented the AUCCESS score of PHYRE-B task for each model. The networks with Question & Answering module surpass the AUCCESS score of the DQN model where **ATQAN** has reached current **State-of-the-art**. Although we could not test all the 10 folds for time issues, ATQAN was better than the baseline at the significance level of 1% in the within template.

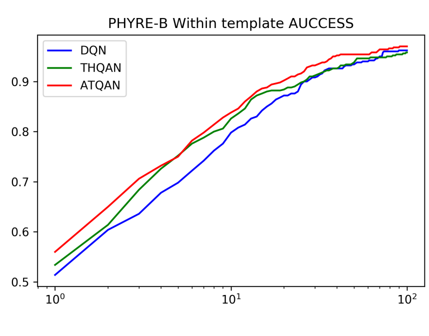

Figure 4. success percentage of the test tasks of 3 models.

Figure 4 illustrates the number of solved test tasks for the designated number of trials. Some of the unresolved tasks with the baseline model was solvable with attaching Question & Answering module.

5.3. Analysis

Figure 5 shows the loss curve of the question answering module for each ball. Our module had a hard time predicting the exact location of the red and green balls and this continues even if we increase the capacity of the network. One of the reasons we think of this mismatch is that the FiLM network is a feature wise attention rather than a spatial wise attention. It would be helpful if we consider alternative ways to help find a spatial wise patterns.

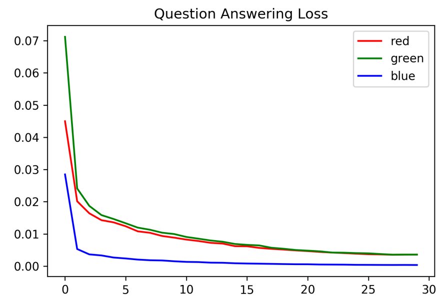

Figure 5. Question Answering Loss for each ball color.

6. Conclusion and Future Work

So, in summary, by guiding the model to answer the questions derived from physical interactions, our agent was able to learn a generalized and sample-efficient policy.

Recent attempts to use deep learning to discover physical laws have shown great potential in fields of the physics and will be essential for research in both directions.

For the future research directions, it will be a useful to have attempt at creating networks with the capability of automatically generating general physical questions with abundant interactions beyond the PHYRE challenge.

7. Roles

- **Hojoon Lee** Command center and Leader of Team.
- **Suyeong An** Research and implement the ATQAN.
- **Hyunseung Kim** Research and implement the THQAN.
- **Dooho Chang** Research and implement the QA module.
- **Taewan Lim** Analyze the environment with statistics.

References

Bakhtin, A., van der Maaten, L., Johnson, J., Gustafson, L., and Girshick, R. Phyre: A new benchmark for physical reasoning, 2019.

Bellemare, M. G., Dabney, W., and Munos, R. A distributional perspective on reinforcement learning. *ICML*, 2017.

Diederik, Kingma, P., and Ba, J. Adam: A method for stochastic optimization. *ICLR*, 2015.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. *CVPR*, 2016.

Iten, R., Metger, T., Wilming, H., del Rio, L., and Renner, R. Discovering physical concepts with neural networks, 2018.

Johnson, J., Hariharan, B., van der Maaten, L., Fei-Fei, L., Zitnick, C. L., and Girshick, R. Clevr: A diagnostic dataset for compositional language and elementary visual reasoning. *CVPR*, 2017.

Loshchilov, I. and Hutter, F. Sgdr: Stochastic gradient descent with warm restarts. *ICLR*, 2017.

Lu, J., Yang, J., Batra, D., and Parikh, D. Hierarchical question-image co-attention for visual question answering. *NIPS*, 2016.

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., and Riedmiller, M. Playing atari with deep reinforcement learning. *NIPS*, 2013.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., and Hassabis, S. L. . D. Human-level control through deep reinforcement learning. *Nature*, 2015.

Perez, E., Strub, F., de Vries, H., Dumoulin, V., and Courville, A. Film: Visual reasoning with a general conditioning layer. *AAAI*, 2018.