Successful Task and Motion Planning with Hindsight Experience Reinterpretation

Abstract

In this paper we describe an approach for improving the performance of robotic Task and Motion Planning (TAMP) systems on complex, long-horizon manipulation problems. In an uncertain environment, unexpected situations are all but guaranteed to arise, causing the robot to risk failure to plan or complete a task. HEReinterpret allows the robot to reinterpret its initial goal under uncertainty, claiming a more believable goal given the current state. We show how our Deep Reframing Net allows the robot to reinterpret challenging goals into more achievable ones, resulting in a 100% success rate on a variety of TAMP problems. We also propose alternative planning methods that do not require deep reframing but still achieve a 100% success rate through goal reinterpretation.

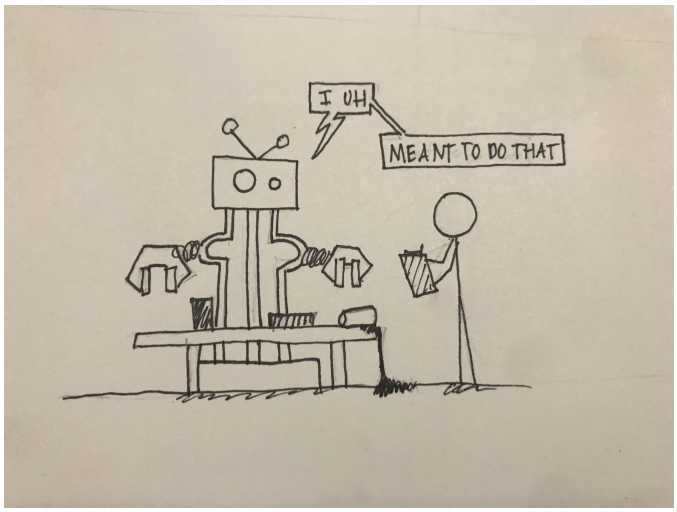

Figure 1. The authors were informed that having a picture of a robot on the first page of your paper correlates with higher acceptance and citation rates.

1. Introduction

Throughout all of history, roboticists have struggled with the pervasive issue of their algorithms not working all of the time. Our goal is to build robots that are so intelligent, they can adapt and reinterpret goals to ensure a level of robustness never seen before in general-purpose robotics. Many industries have been slow to adopt robotic automation because they cannot successfully fulfill enough tasks, or fail too often to rely on. For instance, two Chinese restaurants employed robots, but found them to be too terrible that they caused those restaurants to close [1].

A longstanding long-term goal of general-purpose robotics is to make humans happy to employ robotic systems. In the face of slow motion-planning unable to find trajectories, RRT dances, state uncertainty, numerous false-positives and false-negatives in perception, and massive engineering feats to perform relatively simple tasks like human recreational games.

For instance, if a robot is given the goal to "make a cup of coffee", but drops the spoon on the ground, current robust approaches might try to recover the spoon. However, perception on the ground is difficult. It is far simpler to design a system that reinterprets the goal to "make the human exercise and do a squat." As the human reaches down to grab the spoon, the human accomplishes the task for the robot, ensuring reliable completion.

Our contribution is a framework ensuring 100% completion of the goal in complex general-purpose task and motion planning problems. We demonstrate a prototype of the system, results demonstrating the reliability of HERinterpret and discuss the lack of limitations.

2. Related work

Task and Motion Planning remains an open problem in robotics, due to the challenges of integrating state-of-the-art perception, planning, and controls systems for solving difficult, real-world problems. Some approaches have proved immensely successful at tasks involving dropping and knocking objects over, but struggle to generalize to a wider variety of goals. Other approaches, like [], may find a solution to a given problem only after contemplating for a significant period of time - depending on the problem and how much money you're willing to spend on servers [2], this might even approach the amount of time it takes your PI to respond to your 'urgent' emails.

But the real limiting factor Robotic TAMP systems must contend with today is the damage done to the robot's self esteem when it fails to successfully complete a task. As demonstrated by the modern success of 'reinforcement learning', robots learn best in an environment where they are rewarded, not punished.[3] Our goal-reinterpreting approach is based on an existing body of evidence that the narrative reframing of an intended goal can be a great benefit to an agent's self esteem. This form of reinterpretation has been shown to be effective in therapy settings, such as in [4], [5], or [6]. Shown here to be effective for robots, there is even evidence that the method will extend to help the grad students who work with them. This could be an interesting point of future research.

3. Technical Approach

Hindsight Experience Reinterpretation, or HEReinterpret, is a robust framework to online planning in task and motion planning manipulation domains that requires minimal online computation. First, the robot executes an offline closed-loop plan. Then, the robot *reinterprets* the goal to ensure successful completion. Our framework

DOI: http://dx.doi.org/10.1145/(to come)

is modular and can work successfully with multiple planning and reinterpretation strategies.

3.1 Planning

Here, we describe our recommended planning approaches depending on the constraints and objectives of the system.

3.1.1 STRIPStream

STRIPStream is a symbolic planning framework built using the STRIPS planning language using blackbox samplers to find a low-cost solution in highly constrained domains. The blackbox samplers were hand-written in Python. Then, the agent executes the plan from STRIPStream. Since our framework is always successful, there is no re-planning needed. Ever.

3.1.2 FF-Giveup

Our planning approach is based on FF-Replan [7], an effective method for planning in uncertain domains by planning in a simplified, determinized version of the problem. But in our approach, if the planner is unable to find a plan due to an unexpected observation, instead of replanning starting from the current state as FF-Replan does, it simply gives up. Given a domain description, initial goal, and initial state, FF-Giveup returns a plan and a final reinterpeted goal. First, the domain is determinized using single-outcome determinization where each outcome is chosen randomly. Then, an off-the-shelf FastForward (FF) is run to find an open loop plan to the initial goal. If FF is unable to find a plan to the initial goal, the reinterpreted goal is set to the initial state, and the problem is solved.

The major distinction between FF-Giveup and FF is that FF-Giveup returns a plan with 100% success. Then the controller executes the plan - if the original goal is not achieved, through HEReinterpret, the goal is reinterpreted, and the approach maintains a 100% success rate. The proof of this method's efficacy has been omitted due to constraints on space and the authors' apathy.[1]

3.2 Un-actuated task and trajectory optimization

First, we find a task plan independent of the motion plan. Separating the two allows us to find a plan without worrying about constraints. We use forward chaining to find a series of tasks that may or may not satisfy the preconditions and post-conditions of each task depending on whether the actions were formulated correctly. If there does not exist a plan to the initial goal found through forward chaining, then we set the reinterpreted goal to the initial state.

To execute the motion planning, we transform the system into an un-actuated one, meaning no actuators. We set the control inputs, $u_0, \ldots u_N$ values at knot points $x_0, \ldots x_N$ for a finite length trajectory of length N The optimal trajectory is one that minimizes the following objective function:

$$\sum_{n=0}^{N-1} (x - x_{goal}) + |u|^2) \tag{1}$$

subject to:

$$\forall n \; u_n = 0 \tag{2}$$

At least one plan is guaranteed to exist: to not move any non-existent actuators. This optimization is then guaranteed to return nothing, because nothing is actuated

3.3 HEReinterpret

In task and motion planning, it is common for robots to fail to achieve a goal, either because the robot fails to find a plan, or fails to successfully execute the skills needed to achieve it. This is sometimes called the "MRDMTMOTT (Most Robots Drop Most Things Most Of The Time)" Theorem [8]. In order to improve our approach's success rate, we took inspiration from deep learning's Hindsight Experience Replay, a method for making use of unsuccessful executions by replaying them as successful executions of a different goal. Because our approach does not incorporate learning, we see no benefit from replaying such scenarios, but instead proposed HEReinterpret as an effective approach to TAMP problems. With HEReinterpret, instead of calculating the cost of a state relative to the initial goal, the cost is calculated with respect to a *reinterpreted* goal that optimizes for both similarity to the current state and 'believability', b.

Intuitively, if the robot has just attempted to pick up a cup, and instead knocked the cup onto the floor, a claim to have achieved the "pick" goal will not be very believable. But, achieving the goal of "making sure gravity is still working" is in this situation, a much more believable claim. The cost function incorporating this term is included below.

$$|s_t - g_{reinterpreted}|^2 + b \tag{3}$$

We take a supervised approach to learning the believability of various goal claims given the end state of the world after execution and the goal provided at planning time. In our experiments, the believability of such a goal is labeled by the human operator after a plan has been executed. We use this data to train a neural network to predict the believability of reinterpreted goals, and pick the most believable of a set of sampled possible reinterpretations based on this predictor.

3.4 Experiments

We tested our approach on a variety of long-horizon manipulation problems in an approximated Kitchen Domain. Each problem begins with a set of objects on a table in some initial state, and requires the robot to manipulate said objects in sequence in order to achieve some pre-specified goal state. The PR2 uses its two 7-DOF manipulators to execute open-loop manipulation skills. By detecting the configuration of objects on the table through the use of an RGB-D sensor, it can estimate the state of the world at any point in planning and deter

We also test the accuracy of our novel deep reframing network. The input to the network is an image, which is then resized to a 30x30 thumbnail. A deep neural network, architecture shown in Figure 3 is used to classify the end goal as one of six potential natural language strings, which are the reinterpreted goal. In order to improve generalizability over potential final states, we use data augmentation to expand our dataset from 6 to 12.

3.5 Results

3.5.1 Deep Reframing Network

Deeper Reframing Network The deep re-framing network shows remarkable classification accuracy in very few iterations. In Figure **??**, we show the maximum accuracy seen so far during training. In order to measure variance, we run the learning algorithm N times. The value of N can be found in the appendix. The validation set was sampled from the training set. The model trained on images and annotated reasonable reinterpreted goals.

Other state-of-the-art approaches to real-world robotic Task and Motion Planning, demonstrate low success rates, due to the complexity of managing uncertainty in a real-world domain. In comparison, our approach demonstrates a 100% success rate regardless of the planning algorithm used.

[1] That is to say, the authors Gave Up

```
def build_model():
    model = Sequential()
    model.add(Conv2D(32, (3, 3), activation='relu', input_shape=(imshape[0], imshape[1], 3)))
    model.add(GaussianNoise(0.01))
    model.add(Conv2D(32, (3, 3), activation='relu'))
    model.add(MaxPooling2D(pool_size=(2, 2)))

    model.add(Conv2D(64, (3, 3), activation='relu'))
    model.add(Conv2D(64, (3, 3), activation='relu'))
    model.add(MaxPooling2D(pool_size=(2, 2)))
    model.add(Flatten())
    model.add(Dense(64, activation='relu'))
    #model.add(Dropout(0.9))
    model.add(Dense(64, activation='relu'))
    model.add(Dense(32, activation='relu'))
    model.add(Dense(9, activation='softmax'))
    model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])
    return model
```

Figure 2. Screenshot of Neural Network Architecture

3.5.2 Reinterpretation user study

We presented the results from our validation set to evaluate how believable were the reinterpreted goals. We asked five users to ensure statistical significance, though omit rigorous statistical analysis for clarity. Using the test set shown in Figure 4, we asked two users how plausible they thought the re-interpretation was. In Figure 5 we show the average rating. For clarity, we pick the highest result for each picture. To avoid needing to file for IRB approval, the authors of this paper were the users in the user study.

Here, we show the success of our algorithm on various planning problems compared to the state-of-the-art baselines:

	HEReinterpret	Anything Other than HEReinterpret
Success Rate	100%	<100%

4. Conclusion

In this paper, we show that Hindsight Experience Reframing is an efficient and effective method for achieving high success rates when attempting real-world Task And Motion Planning. It achieves a goal with a 100% success rate, a vast improvement over other state-of-the-art approaches. Additionally, we show how the improved believability of re-framed goals as generated by our Reframing module greatly eased the worries of Grad Students operating the robot, introducing a potentially interesting path of research for future scholars of Human-Robot Interaction. The ability to use un-actuated robots greatly expands the capabilities of inanimate objects to achieve goals while saving energy.

References

[1] R. Price, "A line of robot waiters did such a terrible job that they forced 2 restaurants to close down," *Business Insider*, 2016.

[2] e. a. Andrychowicz, Marcin, "Learning dexterous in-hand manipulation," *arXiv preprint arXiv:1808.00177*, 2018.

[3] R. S. Sutton and A. G. Barto, *Introduction to reinforcement learning*. Cambridge, Massachusetts: MIT press, 1998.

[4] J. Powell, *Reflection and the evaluation of experience: prerequisites for therapeutic practice*, pp. 26–42. Boston, MA: Springer US, 1991.

[5] A. Stevens

[6] R. Stouffer

[7] R. G. Sungwook Yoon, Alan Fern, "Ff-replan: A baseline for probabilistic planning.," *ICAPS*, 2007.

[8] ltchin. something ltchin says sometimes.

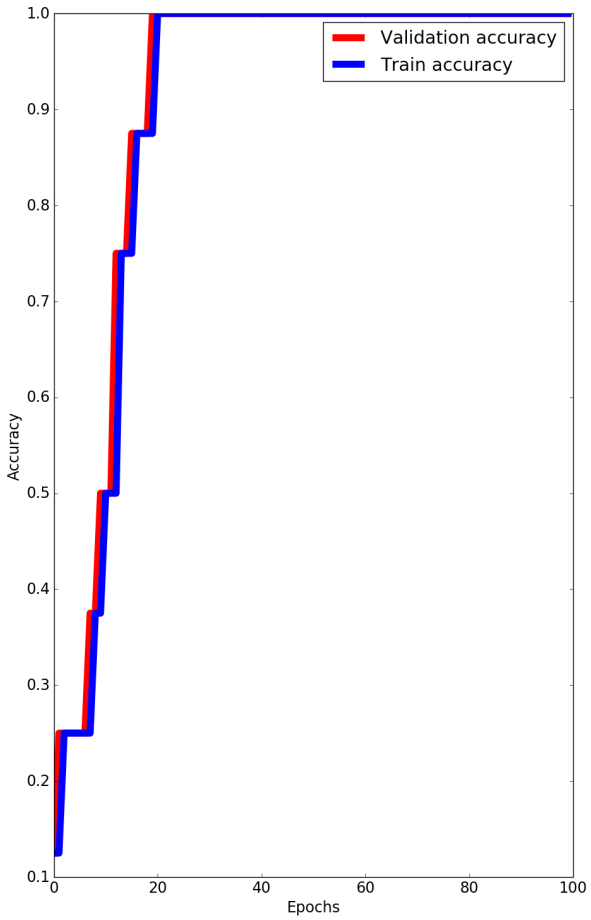

Figure 3. Learning curve of maximum accuracy seen so far over training epochs

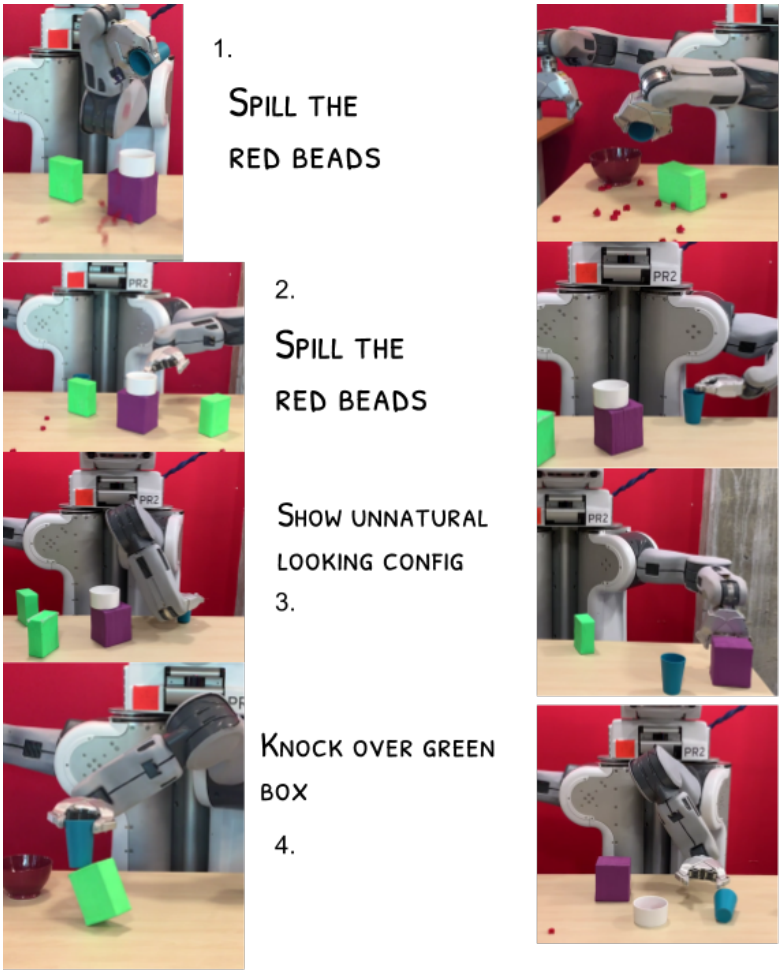

Figure 4. Test set of labeled images with annotations

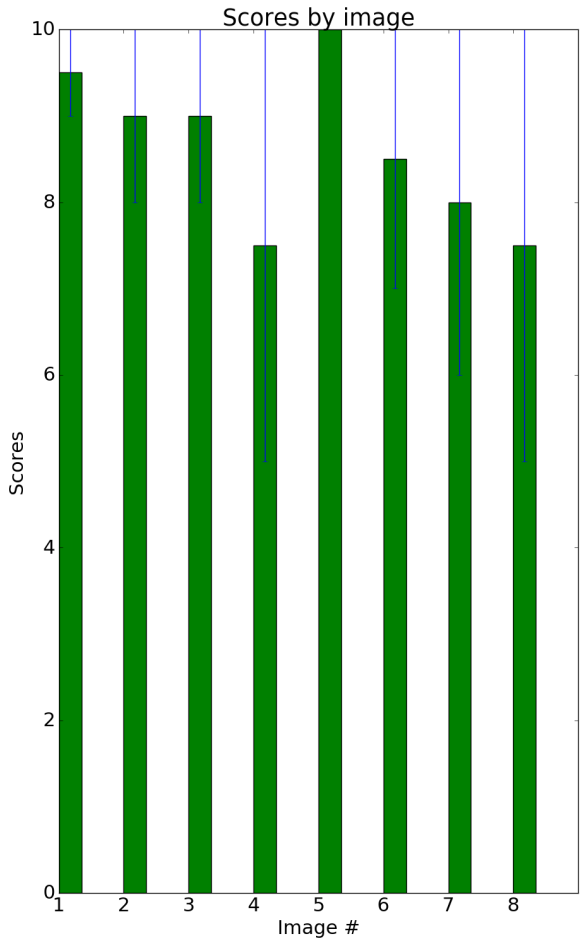

Figure 5. User Study results surveying the authors about how plausible the reinterpreted goal is where 1 is implausible and 10 is the most plausible