Learning to Pick Up Objects Through Active Exploration

John Oberlin and Stefanie Tellex
Computer Science Department, Brown University

Robotics will assist at childcare, help with cooking, and provide service to doctors, nurses, and patients in hospitals. Many of these tasks require a robot to robustly perceive and manipulate objects in its environment, yet robust object manipulation remains a challenging problem. Transparent or reflective surfaces that are not visible in IR or RGB make it difficult to infer grasp points [4], while emergent physical dynamics cause objects to slip out of the robot's gripper; for example, a heavy object might fall out of the robot's gripper unless it grabs it close to the center.

To address these limitations, we propose an approach for enabling a robot to learn about objects through active exploration and adapt its grasping model accordingly. We frame the problem of model adaptation as a bandit problem, specifically the identification of the best of the arms of an N-armed bandit, [9] where the robot aims to minimize simple regret after a finite exploration period [2]. Our robot can obtain a high-quality reward signal (although sometimes at a higher cost in time and sensing) by actively collecting additional information from the environment, and use this reward signal to adaptively identify grasp points that are likely to succeed.

Existing algorithms for best arm identification require pulling all the arms as an initialization step [5, 1, 3]; in the case of identifying grasp points, where each grasp takes more than 90 seconds and there are more than 1000 potential arms, this is a prohibitive expense. To address this problem, we present a new algorithm, Prior Confidence Bound, based on Hoeffding races [6]. In our approach, the robot pulls arms in an order determined by a prior, which allows it to try the most promising arms first. It can then autonomously decide when to stop by bounding the confidence in the result. Figure **??** shows the robot's performance before and after training on a ruler; after training it grasps the object in the center, improving the success rate.

Formally, the agent is given an N-armed bandit, where each arm pays out 1 with probability μ_i and 0 otherwise. The agent's goal is to identify a good arm (with payout $>= k$) with probability c (e.g., 95% confidence that this arm is good) as quickly as possible. As soon as it has done this, it should terminate. The agent is also given a prior π on the arms so that it may make informed decisions about which grasps to explore.

Our contributed algorithm, Prior Confidence Bound, iteratively chooses the arm with the highest observed (or prior) success rate but whose probability of being below k is less than c. It then tries that arm, records the results, and checks to see if its probability of success is above k or within $[k-\epsilon, k+\epsilon]$ with probability c (in which case it terminates). (If the latter condition is not included, an arm with success probability equal to k will continue to be pulled indefinitely.) We need to estimate the probability that the true payout probability, μ, is greater than the threshold, c, given the observed number of successes and failures:

$$\Pr(\mu_i > c|S, F) \tag{1}$$

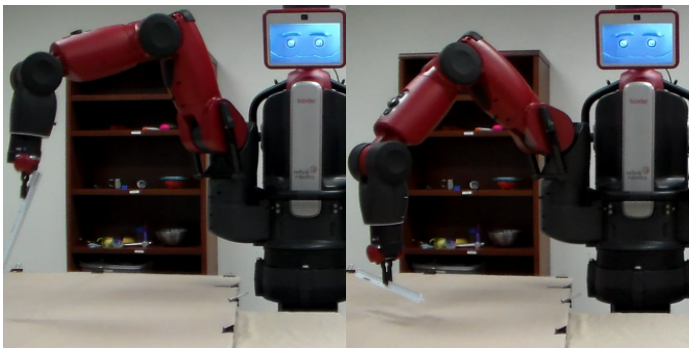
We can compute this probability using the law of total probability: $\Pr(\mu_i > c|S, F) = 1 - \int_0^k \Pr(\mu_i = \mu|S, F)d\mu$, assuming a beta distribution on μ: $\int_k^1 \mu^S(1-\mu)^F d\mu$. This integral is the CDF of the beta distribution, and is called the regularized incomplete beta function [7].

Evaluation

We implemented our approach on the Baxter robot. The robot acquired visual and RGB-D models for 30 objects using our autonomous learning system. The objects used in our evaluation appear in Figure 1(b). After acquiring visual and IR models for the object at different poses of the arm and camera, the robot performed the bandit-based adaptation step. After the robot detects an initially successful grasp, it shakes the object vigorously to ensure that it would not fall out during transport. After releasing the object and moving away, the robot checks to make sure the object is not stuck in its gripper. If the object falls out during shaking or does not release properly, the grasp is recorded as a failure. If the object is stuck, the robot pauses and requests assistance before proceeding.

Most objects have more than one pose in which they can stand upright on the table. If the robot knocks over an object, the model taken in the reference pose is no longer meaningful. Thus, during training, we monitored the object and returned it to the reference pose whenever the robot knocked it over. In the future, we aim to incorporate multiple components in the models which will allow the robot to cope with objects whose pose can change during training.

Our evaluation demonstrates that our adaptation step improves the overall pick success rate from 55% to 75% on our test set of 30 household objects, shown in Figure 1(b). Moreover, our approach also enables the robot to learn success probabilities



(a) Before learning, the robot grasps the ruler near the end and drops it; after learning, it grasps it near the middle.

(b) Objects in our evaluation, which improves grasping performance from 55% (before learning) to 75% (after learning).

	Prior	Training	Marginal
$\Delta = 0; training = 3$			
Garlic Press	0/10	8/50	2/10
Gyro Bowl	0/10	5/15	3/10
Helicopter	2/10	8/39	3/10
Big Syringe	1/10	13/50	4/10
Clear Pitcher	4/10	3/4	4/10
Sippy Cup	0/10	6/50	4/10
Red Bucket	5/10	3/3	5/10
Dragon	8/10	5/6	7/10
Ruler	6/10	5/12	7/10
Triangle Block	0/10	3/13	7/10
Bottle Top	0/10	5/17	7/10
Wooden Spoon	7/10	3/3	7/10
Icosahedron	7/10	7/21	8/10
Blue Salt Shaker	6/10	5/10	8/10
Epipen	8/10	4/5	8/10
Wooden Train	4/10	11/24	8/10
Stamp	8/10	3/3	8/10
Toy Egg	8/10	4/5	9/10
Yellow Boat	9/10	5/6	9/10
Vanilla	5/10	4/5	9/10
Round Salt Shaker	1/10	4/16	9/10
Packing Tape	9/10	3/3	9/10
Purple Marker	9/10	3/3	9/10
Whiteout	10/10	3/3	10/10
Syringe	9/10	6/9	10/10
Brush	10/10	3/3	10/10
Red Bowl	10/10	3/3	10/10
Shoe	10/10	3/3	10/10
Metal Pitcher	6/10	7/12	10/10
Mug	3/10	3/4	10/10
Total	165/300	148/400	224/300
Rate	0.55	0.37	0.75

(c) Quantiative results.

Fig. 1. Results from our evaluation.

for each object it encounters; when the robot fails to infer a successful grasp for an object, it knows this fact, enabling it to take active steps to recover such as asking for help [8].

Conclusion

Our robotic system gathers feedback from the environment, which it uses to learn models to detect, localize, and manipulate previously unseen objects. In the future, we plan to explore learning parameters for a wider variety of tasks. For instance, different objects can be located and oriented better or worse at different heights. One could learn which heights work well for which objects. Likewise, we use color gradients for localization, but some objects would work better with other quantities. One could learn the appropriate map to use when localizing each object, or even further, the map might also depend upon the robot's current environment.

Instance-based approaches mean that the robot must acquire data about a specific object before manipulating it; however by adapting itself to the object it can obtain higher accuracy. We aim to aggregate data collected via our instance-based system to create a new data set of images, RGB-D, and grasp success rates at various poses to provide supervision for general-purpose category models for grasping. Our long term vision is a system that can infer high quality grasps for any object, but adaptively recover if the initial grasp attempt is unsuccessful, leading overall to robust and accurate pick-and-place.

References

[1] Jean-Yves Audibert and Sébastien Bubeck. Best arm identification in multi-armed bandits. In *COLT-23th Conference on Learning Theory-2010*, pages 13–p, 2010.
[2] Sébastien Bubeck, Rémi Munos, and Gilles Stoltz. Pure exploration in multi-armed bandits problems. In *Algorithmic Learning Theory*, pages 23–37. Springer, 2009.
[3] Shouyuan Chen, Tian Lin, Irwin King, Michael R Lyu, and Wei Chen. Combinatorial pure exploration of multi-armed bandits. In *Advances in Neural Information Processing Systems*, pages 379–387, 2014.
[4] Ilya Lysenkov, Victor Eruhimov, and Gary Bradski. Recognition and pose estimation of rigid transparent objects with a kinect sensor. *Robotics*, page 273, 2013.
[5] Shie Mannor and John N Tsitsiklis. The sample complexity of exploration in the multi-armed bandit problem. *The Journal of Machine Learning Research*, 5:623–648, 2004.
[6] Oded Maron and Andrew W Moore. Hoeffding races: Accelerating model selection search for classification and function approximation. *Robotics Institute*, page 263, 1993.
[7] F. W. J. Olver, D. W. Lozier, R. F. Boisvert, and C. W. Clark, editors. *NIST Handbook of Mathematical Functions*. Cambridge University Press, New York, NY, 2010.
[8] Stefanie Tellex, Ross Knepper, Adrian Li, Daniela Rus, and Nicholas Roy. Asking for help using inverse semantics. In *Robotics: Science and Systems (RSS)*, 2014.
[9] William R Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, pages 285–294, 1933.