In 1994 International Machine Learning Conference -
Workshop on Robot Learning, Rutgers University, New Brunswick NJ, July 1994.

Active Exploration Based ID-3 Learning for Robot Grasping

Marcos Salganicoff **Len G. Kunin**
Lyle H. Ungar
Department of Computer and
Information Science
University of Pennsylvania
Philadelphia, PA, USA
{sal,kunin}@grip.cis.upenn.edu
ungar@central.cis.upenn.edu

Abstract

We present a technique for active robot learning by taking the Interval Estimation (IE) exploration heuristic and combining it with the inductive learning algorithm ID-3. We use the resulting ID-3 tree as an indirect predictor by selecting actions that index leaves of the tree that have a high expected probability of success conditioned on the current perceived attributes. We motivate this technique with a real robot learning system for learning to grasp novel objects where each grasping trial can be lengthy and it is desirable to reduce unnecessary experiments. The resulting system has an on-line learning capacity which interleaves its learning and exploitation of learned knowledge smoothly. This allows the learner to benefit from learning as quickly as possible and focuses its data acquisition on profitable areas of action space for the grasping task.

1 Introduction and Motivation

Traditionally, machine learning algorithms and connectionist approaches have assumed a batched learning paradigm in which a complete set of data is given to the learner which attempts to generalize and predict based on that static learning set. However, in the robotic domain, as in many others, the learner (robot) can often control their inputs (e.g. active perceptual systems) and outputs (actions) to facilitate learning. In active learning or learning with queries, the learner can select the which inputs are presented to it. It has been shown that in some situations, the ability to control inputs improves learning speeds by orders of magnitude [Moore, 1991] and may even make learning tractable [Baum, 1991] where it is otherwise not possible. This is primarily due to the fact the learner can control inputs so that they are informative with respect to the current task.

1.1 Control of Exploration is Critical to Success

The cost of exploration often has a major impact on the success of robotic learners. This is especially the case in robotic environments, where each action, either a perceptual or motor primitive, might take a significant amount of time to invoke, and might have concomitant risks as well. Often, the computational complexity of invoking the learning algorithm is small in comparison to the cost in time and material of gathering the data set. This problem becomes particularly acute in high-dimensional learning tasks, since the number of samples necessary to uniformly sample the attribute space increases exponentially with the dimensionality. Even if the dimensionality of the task is relatively small, uniform random exploration may not be feasible because the cost of acquiring all exemplars may be too much relative to manually programming a solution to a problem.

We present a technique for active robot learning which takes the Interval Estimation (IE) [Kaelbling, 1990] exploration heuristic and combines it with the inductive learning algorithm ID-3 [Quinlan, 1986]. The resulting ID-3 tree predicts successes as a function of the perceived attributes of the object being grasped and of the grasping strategy (the approach angles). Actions are then chosen using an indirect indexing scheme which indexes the leaves on the tree that have a high expected probability of success (or reward) [Salganicoff and Bajcsy, 1992] conditioned on the current perceived attributes. We motivate this technique with a real robot learning application for learning to grasp novel objects. Each grasping trial can be lengthy (approximately two minutes in duration) and so we want to eliminate unnecessary experiments. We demonstrate a system that has an on-line learning capacity which interleaves its learning and exploitation of learned knowledge in a smooth fashion. This allows the learner to benefit from learning as it occurs, and more importantly, focuses its data acquisition on areas of action space most likely to improve performance.

2 Combining ID-3 and IE

Several researchers have investigated intelligent experimentation strategies for learning [Thrun and Moller, 1992; Moore, 1990; Schneider, 1993; Sutton, 1990; Atlas *et al.*, 1990; Cohn, 1994]. These are briefly summarized in section 4. We focus here on Kaelbling's Interval Estimation (IE) Algorithm for control of experimentation [Kaelbling, 1990]. The IE algorithm is designed for control of experimentation during learning in situations with multiple discrete actions in a given state. The algorithm works by keeping statistics on the number of times a given action has been executed and the proportion of times that it has succeeded in terms of yielding a fixed reward. A confidence interval is computed for the underlying probability of reward for each of the feasible actions in a given context. The action whose upper bound confidence interval is highest is the one chosen. The rationale is that an action may have a higher upper bound for two possible reasons, either because the underlying state of nature is that the action is actually a good one to take, or that very few trials of the action have been executed, in which case there is insufficient empirical evidence to decide the goodness of the action in that context. Most importantly, this approach rapidly stops executing actions which have a very low likelihood of success and focuses on sampling those actions whose successful outcome cannot statistically be ruled out.

It is interesting to note that although the Interval Estimation (IE) algorithm was originally proposed in the context of a finite number of discrete actions [Kaelbling, 1990] for reinforcement learning, the approach can be generalized to other learners. We apply it with the ID-3 Algorithm [Quinlan, 1986] as follows. Let T be the current tree partitioning, based on binary outcome O, of the joint perception-action space $(P, A) = ((p_1, \ldots, p_l), (a_1, \ldots, a_m))$, where p_i and a_i are the individual perception and action attributes. In other words T is a tree created by inducting on a learning set of $((P_i, A_i, O_i)$ for $i = 1, \ldots, N)$ where N is the current number of observations in the learning set. To bind percepts to actions, the system senses the current percept P vector. Let F be the set of leaves of the ID-3 tree that have perceptual interval values that contain the currently sensed attribute values of P. The set F represents a partitioning or binning of the attribute space by the current decision tree subject to the currently perceived attributes, P. As we vary A over all possible values with P fixed, we move over all possible leaves which predict outcomes for different action parameters A (see Figure 1). Since the leaves induce a partition on the space, we now have a finite and discrete number of action outcome bins given the observed P. Given this the partitioning, we are now in a position to directly apply the IE algorithm since we have a finite and discrete number of possible actions based on that partition. The statistics necessary for computing the confidence intervals (the number of success and failures contained in a leaf partition) are already available because they are needed to form the tree using the ID-3 algorithm.

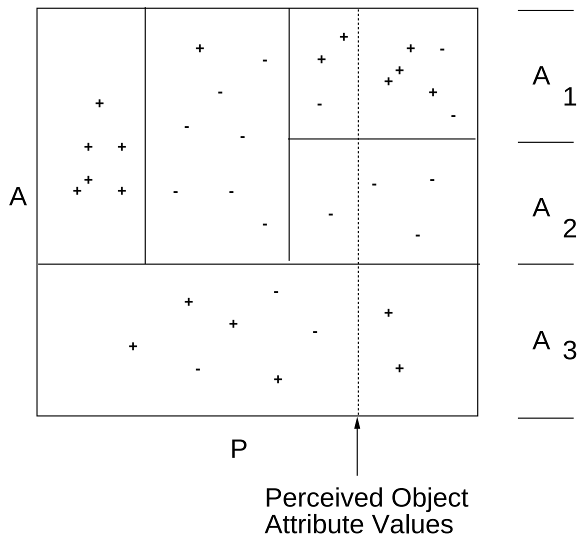


Figure 1: A schematic of the exploration and action selection algorithm. The solid lines represent the ID-3 partitioning of the joint perception-action space (P, A). For the purposes of illustration assume a single degree-of freedom A action and a single sensed value P, although the approach generalizes to arbitrary dimensions. When a new percept is sensed, this determines the discrete intervals for action values (the dotted line) and their associated statistics of task success (+) and failure (-). An action in interval corresponding to the leaf with the highest upper bound confidence value for probability of success is executed.

To conclude the description of the method, for each leaf l_i in F we compute the upper bound probability of success according to the same binomial confidence interval formula as utilized by Kaelbling [Kaelbling, 1990; Larson and Marx, 1986]. The selected action leaf l^* is the leaf with highest upper bound of success probability. Let A_{min}, A_{max} be the lower and upper action attribute intervals for l^*. The next action is chosen as random within l^* using a uniform distribution bounded by the intervals A_{min}, A_{max}.

3 The Grasping Task

The task of the system is to learn to pick up unknown objects based on perceptual measurements of the height, width and length extents of the object. The system must learn to select the azimuth θ and elevation ϕ of the approach axis of the two fingered gripper in a fashion which it predicts is most likely to succeed given the observed object description, in h, w and l (see Figure 2).

The actual robotic system which carried out the experiments consists of two PUMA 560 robots: the perceptual robot and the grasper. The pertinent aspects of the system are summarized in Figure 3. The range image provided by the scanner is processed using a superquadric data reduction procedure [Solina, 1987] which fits a superellipsoid to the dense range data. This procedure returns the object position orientation and shape parameters. Only the extents of h, w, l are used as perceptual inputs to the learner. An object centered action frame is chosen so that the θ and ϕ are relative to the direction of the w and h axes, respectively.

The ID-3 algorithm takes as its input 5 real valued attributes for all of the combined attempts on the objects with given actions so far. The attributes are the h, w, l perceptual parameters of the superellipsoid fit to the object, and the action values θ and ϕ (see Figure 2). The output variable O is binary: whether the grasp succeeded or failed. The success or failure of a grasp is assessed by using the force torque sensors on the gripper. During the grasp attempt, the system moves the gripper fingers inward from their full open position to an interfinger distance of 5 mm. After the approach and grasp phase, the robot moves the gripper upwards, hopefully lifting the held object. It then measures the normal forces on the gripper fingers. If the force is below a threshold value, then the fingers are characterized as empty, otherwise if there is a large normal force, the grasp is labelled a success.

3.1 Previous Work in Learning to Grasp

This work differs from previous efforts in learning for robotic grasping in terms of action and perceptual representation, as well as the learning methods employed. Dunn [Dunn and Segen, 1988] employed a two dimensional polygonal representation, and a random search for successful grasps during learning, followed by a 2-D model matching during execution. Tan [Tan, 1990] employed a feature-based sonar depth representation and a cost sensitive extension of ID-3 with 3-D objects. Bennett [Bennett, 1991] worked in robotic grasping of polygonal 2-D puzzle piece task using explanation-based learning and domain theories about uncertainty and grasping. Mel [Mel, 1991], Ritter [Ritter *et al.*, 1991] and Cooperstock have used $\Sigma - \Pi$ neural-networks, self-organizing feature maps and backpropagation, respectively, for learning visually-guided control of robot arms for grasping. Kamon [Kamon *et al.*, 1994] uses a contour-based two dimensional representation of the perimeter of objects and measures several attributes such as the finger contact angles relative to the surface normals, distance of line connecting the contact points to the estimated center of mass, and distance of the grasp points to the symmetry axis of the object. The system uses a nearest-neighbor based approach to index similar previously successful grasps points and make estimates of the quality of the grasp. Salganicoff [Salganicoff, 1992; Salganicoff, 1993] developed a system for grasping objects based on three-dimensional information from a laser range scanner to select from a set of canonical approach directions using a density-adaptive learning algorithm.

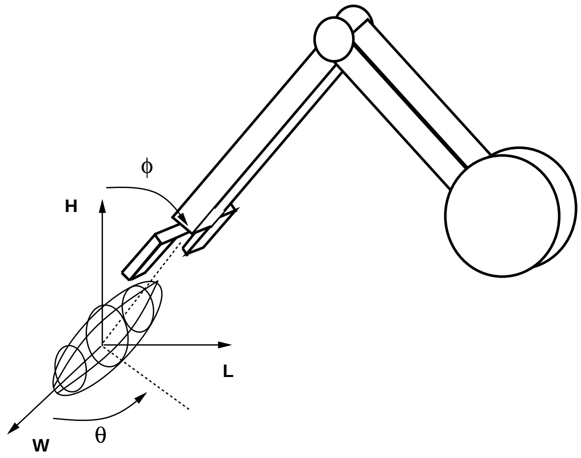


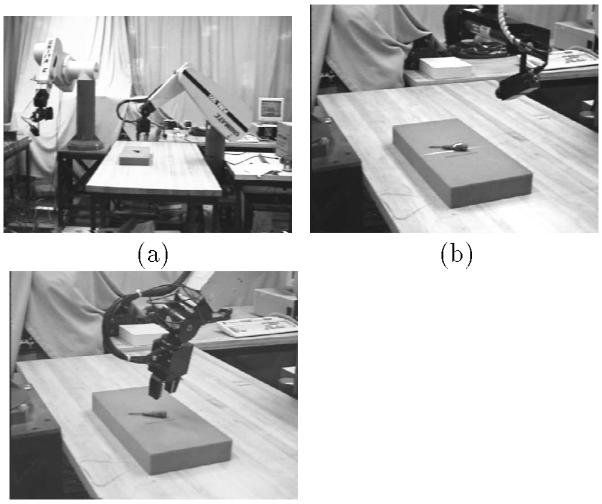
Figure 2: The perceptual and action representation for the grasping system.

3.2 Experimental Protocol

In summary, the operation of the system consists of the following sequence:

1. The robot encounters the object and the laser scanner scans the object.
2. Data reduction algorithms refine the data and produce the h, w and l of attributes that describe the object [Solina, 1987].
3. A decision tree is generated based on the database of points using the standard ID-3 algorithm [Quinlan, 1986] [1].
4. The tree is pruned.
5. Interval estimation is used to select the best θ, ϕ using the approach described in section 2. Here $P = (h, w, l)$ and $A = (\theta, \phi)$.
6. The robot attempts the grasp and assesses the outcome.
7. The new data point (P, A, O), attributes plus the outcome, is recorded in the database.

[1] Incremental versions of ID-3 such as ID-5 [Utgoff, 1988] could have been used, but the non-incremental algorithm executed rapidly enough that it was not a hindrance to regenerate the full tree including the new observation



(a) (b) (c)

Figure 3: The experimental system, (a) consisting of the perceptual robot (left) and the grasping robot (right). (b) The vision system consists of a mobile laser range scanner mounted on the end of the perceptual robot which is swept over the object to be grasped. (c) The gripper is a simple two finger gripper with force torque sensors to sense grasp force. Compliant foam is placed under the object to prevent damage to the gripper or object in situations where the fingers would be stubbed.

8. Goto step 1.

3.3 Experimental Results

A total of six objects were chosen for grasping, a balsa block, whiteboard eraser, small flashlight, wire spool, screwdriver and a stapler (see Figure 4). For each of these objects, the protocol of section 3.2 was carried out, starting with an empty learning set for each new item.

Figure 5 shows the average success (or hit) rate over the thirty trials. The hit rate is the windowed proportion of successful grasps to attempts using a sliding window 10 attempts wide over the 30 instances in a learning run. For example, the hit rate for trial 10 consists of the proportion of consecutive grasp attempts numbered 10 through 19 that succeeded. It can be seen that the robot rapidly converges on a strategy for selecting approaches that succeeds reliably (.9) in a small number of trials.

The scatter plot of Figure 5 also illustrates the effect of the exploration strategy. This figure shows the selected azimuth θ and elevation ϕ values for the balsa block trials in the range $0° < \theta < 90°$ and $0° < \phi < 60°$. Each point in the θ, ϕ space is marked by its outcome, 0 for failure or 1 for success. As is evident by the plot, the action parameters are not uniformly distributed in the space, but concentrated in the region which has a small θ value, which is the constraint for alignment for this object. Outside of the constraint values there are very few attempts. The exploration algorithm has rapidly focused exploration on the region of success.

Figure 4: The objects for which learning was attempted.

In the case of the flashlight, the learning curve increase more slowly (see Figure 6) since the width of the screwdriver is quite close to the maximum width of the gripper and it takes more attempts to find a first successful grasp which can form the basis of a tree partition.

The learning curve for the stapler is also fairly rapid (see Figure 7), converging after about 20 attempts. The system succeeds almost at once and maintains a high level of performance. The scatter plot once again shows that the actions are highly concentrated in the favorable part of the action space. Similar learning performance and action distributions were observed for the eraser, screwdriver and spool.

A simulation study was undertaken to explore the scaling properties as the number of objects presented to it increased. In the simulation we modelled the gripper by taking the necessary coordinate transformations on the object and determining if the projection of the object in the 2 dimensional plane was smaller than the maximum span of the fingers. If the gripper cleared the object, it was assumed to be graspable. Figure 8 shows the performance of the learner over 20 objects presented sequentially to the simulation system. The system generally performed very well, except for an occasional object which had extents very close to the maximum graspable dimensions for the gripper. When this occurs, the system may have to search for a while until it finds an action which succeeds and provides a new possible split for the ID-3 algorithm. It was observed that the pruning level of the ID-3 algorithm did not have a major impact on system performance. The distribution of actions in Figure 8 (b) illustrates

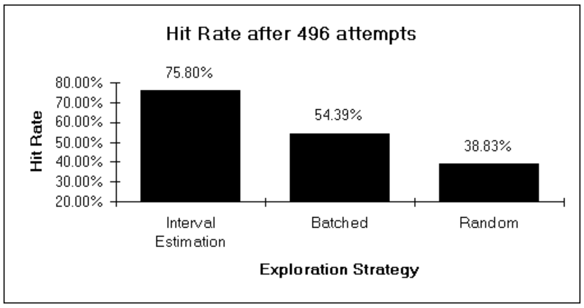


Figure 9: Comparison of IE based ID-3 and batched ID-3 in simulation. Random refers to the probability of succeeding with a completely random strategy. The results are the average of 100 runs and represent the probability of grasping the next object after 496 attempts.

that the attempted actions are grouped closer to angle values that are aligned with the major axes of the objects, as one would expect.

3.4 Active Exploration vs. Batched Learning

We were also interested in comparing the performance of ID-3 in this task with and without the IE strategy. Figure 9 compares the probability of succeeding with a randomly presented object with and without the benefit of the IE strategy after 496 attempts. It can be seen that IE increases the performance of the learner dramatically in comparison to the random experimentation strategy followed by batched application of ID-3.

4 Discussion

The previous results show that the IE algorithm provides an efficient mechanism for focusing the activity of an ID-3 based learner on profitable areas of the action space. The method learns to grasp new objects very rapidly, usually within about 15 attempts. The scatter plots illustrate the fact the experiments are highly focused in the correct areas of the attribute space.

One possible drawback of the approach is that it is possible that other successful areas of the perception/action space might be missed, since the method could greedily identify a good region and cease exploration in other regions. For example, an initial split based on the trials seen so far might yield a partitioning which covers only a subset of the true success region. The subsequent attempts would then be focused in that restricted region. However, the method will always identify some region of the joint perception action space which is predicted to succeed, assuming such a region exists. If we are willing to tradeoff optimality for rapid achievement of a sufficient solution, this is acceptable.

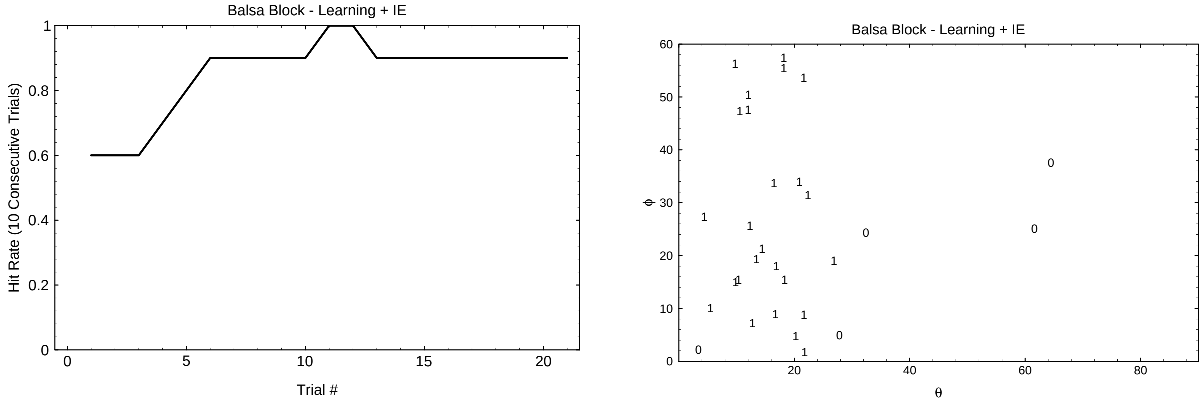
Another possibility is that the IE algorithm may get stuck in actions which are sub-optimal due to pathological data orderings. See ([Kaelbling, 1990].) for a discussion and analysis of this phenomenon. Fortunately, the probability of this occurring becomes quite small for any large difference in the underlying probability of success. Therefore, for practical purposes, if the sticking phenomenon occurs, it is quite likely that the efficacies of the different alternatives will be almost indistinguishable.

Although the action space is continuous, using the indirect indexing method with the ID-3 partitioning affords a natural and meaningful partition of that space. The resulting intervals provide a range of feasible action which can be intersected with other high level constraints, such as approach limits that might be due to obstacles around the object.

4.1 Comparison to Other Approaches

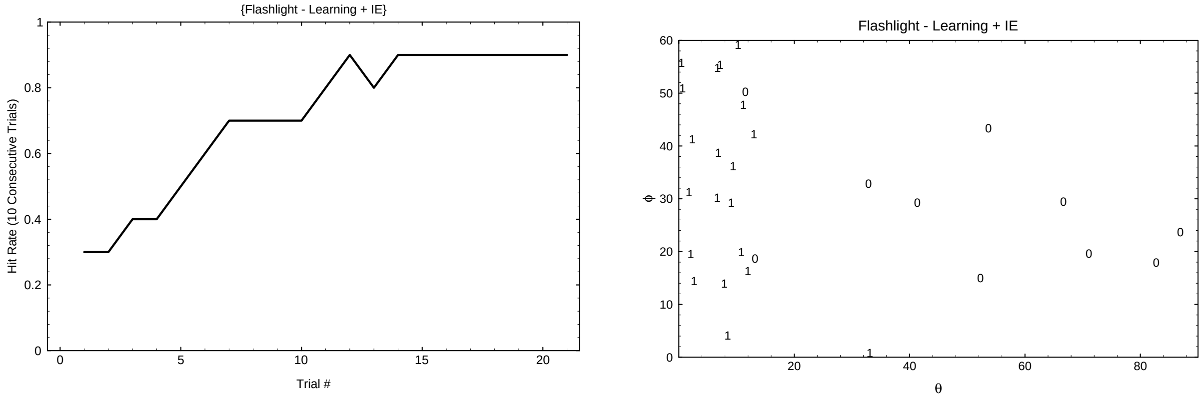
Thrun [Thrun and Moller, 1992] proposes a directed experimentation method for neural networks that selects actions which maximize a linear combination of exploration and exploitation utilities. Gradient descent in the action space is done by taking the derivatives of the two neural networks which compute exploration and exploitation utility and hill climbing to a local optimum for the combined utility. The action at the local optimum is then executed. The exploration utility measure is based on the expected error of the learner in the current state for a given action and the exploitation is based how close the action is expected to brings the system to next desired state. In order to prevent the combined measure from rapidly oscillating between preferring exploratory and exploitatory actions, a *selective attention* hysteresis mechanism is utilized. Although the gradient approach is not practical for use with discontinuous partitioning approaches such as ID-3, it does point out that prediction accuracy can be a powerful indicator of the potential gain from an experiment.

Moore [Moore, 1990] devised a heuristic for trading off exploration versus exploitation in learning forward models of a specific trajectory for a robot arm. The method works by evaluating a set of candidate actions favoring those that are predicted to achieve the desired next state with high likelihood, and discouraging those actions which are predicted to yield a next state which is far from the desired with high likelihood. The predictiveness is based on the heuristic of the Euclidean distance of the action under evaluation to previous observed points in the learning set. Schneider [Schneider, 1993] has extended and modified this approach for action selection in very high dimensional spaces. It is not clear if Moore's heuristic is alway valid. The IE method used here is advantageous since it makes an es-



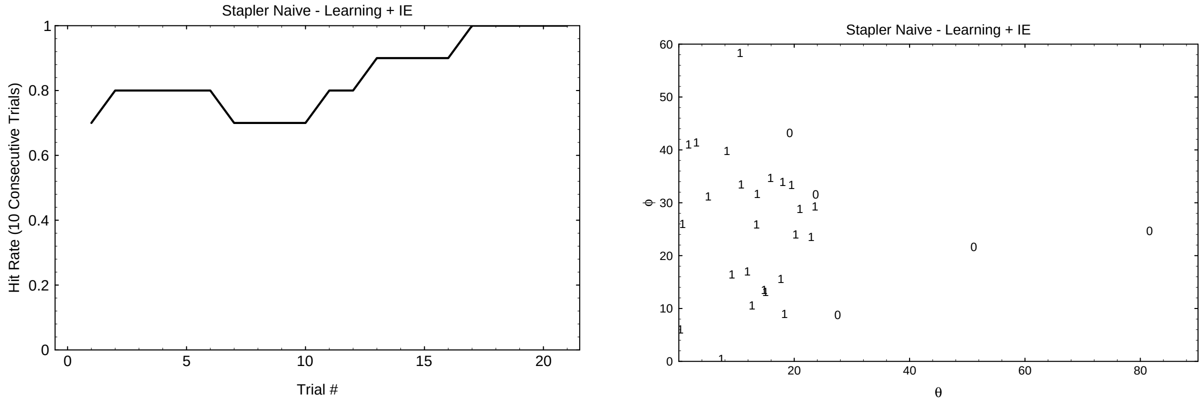
(a) (b)

Figure 5: The hit rate for the balsa block over the first 30 trials and (b) corresponding scatter plot of chosen actions in the action space. The cumulative probability for successful grasping is .83 for the learning condition and .43 for random approach.



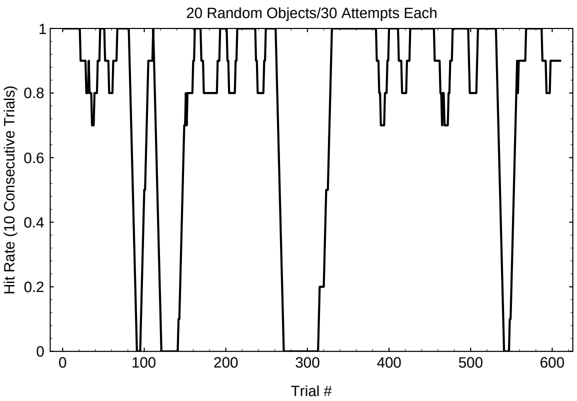
(a) (b)

Figure 6: The hit rate for the flashlight over the first 30 trials and (b) the distribution of actions in the action space. The cumulative probability for successful grasping is .67 for the learning condition and .23 for random approach.

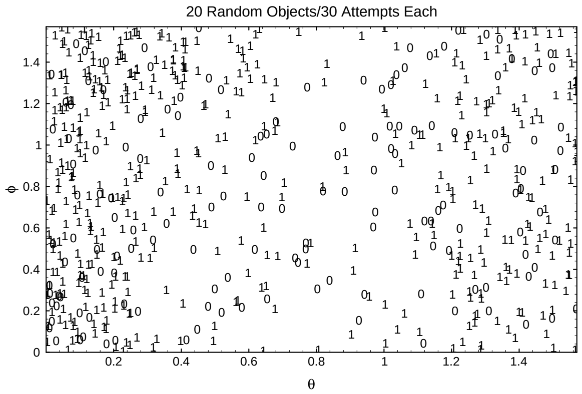


(a) (b)

Figure 7: The hit rate for the stapler over the first 30 trials and (b) the distribution of actions in the action space. The cumulative probability for successful grasping is .83 for the learning condition and .4 for random approach.



(a)



(b)

Figure 8: Simulation study for scaling to larger number of objects. The run consisted of a sequence of 20 objects, each of which is presented for 30 consecutive grasping attempts. It can be seen that the system generally performs well, but occasionally is presented with a difficult object which require many attempts before the success rate rises. (b) shows that the exploration strategy focuses grasping attempts in regions that represent approaches aligned with the major axes of the objects, which are, on average, more likely to succeed.

timate of predictiveness based on empirical statistics, rather than an *a priori* heuristic.

Cohn has proposed two methods for control of experimentation. Selective sampling [Atlas *et al.*, 1990] attempts to identify the region of uncertainty between the most general and most specific concepts that are consistent with the training set. The distribution of query points is then biased to favor those points that lie within the region of uncertainty. Cohn has also proposed using statistical measures of experimental action based on Optimal Experimental Design to hill climb in the parameter space to a point which is predicted to be maximally informative [Cohn, 1994] in terms of minimizing the expected generalization error over the entire domain. Note that [Cohn, 1994] measures the informative of an action in terms of learning the function over its entire domain. Ideally, we would only want to learn part of the domain of the function that is relevant to succeeding at the task at hand in order to avoid the curse of dimensionality. Characterizing the entire domain may be an unnecessary or impossible luxury.

Other techniques are recency based [Sutton, 1990], tending to favor actions that will bring the learner to parts of the perception/action space that have not yet been visited or have not been visited for a long period. The notion of recency may be hard to implement in continuous state spaces, since the probability of revisiting the same point has zero probability, but *is* possible to achieve using a tree based partitioning

4.2 Future Work

The approach described here is relatively general. We can also consider applying the IE algorithm as an action selector to any batched algorithm which can provide estimates of the uncertainty of its predictions and turn it into an on-line algorithm. As long as the method can partition the space into regions that provide statistics about the total number of observation contained per region and the relative proportion of successes it is a candidate for IE. Candidate learning algorithms would include Radial Basis Functions, Nearest-Neighbor Approaches and Hierarchically Gated Mixtures of Gaussians. Additionally, we can explore using some of the other exploration approaches described above or other statistical measures for efficient action selection with ID-3 or other batched learners.

5 Conclusion

We have presented a technique for active robot learning which combines the Interval Estimation (IE) exploration heuristic [Kaelbling, 1990] with an inductive learning algorithm, ID-3 [Quinlan, 1986]. We demonstrated the success of this technique with a real robot learning application for learning to grasp novel objects where each grasping trial is lengthy so that minimizing the number of trials to convergence is crucial. The resulting learner was able to benefit from learning as quickly as possible by avoiding random exploration and focusing its data acquisition on profitable areas of action. The results also bear out the fact that using an exploration strategy dramatically improves performance over an uninformed batch learner.

Acknowledgments

This work was supported by Navy Grant N00014-92-J-1647, AFOSR Grants 88-0244, AFOSR 88-

0296; Army/DAAL 03-89-C-0031PRI; NSF Grants CISE/CDA 88-22719, IRI 89-06770, and ASC 91 0813 and an NSF HPPC Postdoctoral Associateship for MS (CDA-9211136). We thank Ruzena Bajcsy for many productive discussions and ideas related to the development of this work.

References

[Atlas *et al.*, 1990] L. Atlas, D. Cohn, R. Ladner, M.A. El-Sharkawi, R.J. Marks II, M.E. Aggoune, and D.C Park. Training connectionist networks with queries and selective sampling. In *Advances in Neural Information Processing Systems 2*, pages 567–573. Morgan Kaufmann, December 1990.

[Baum, 1991] E.B. Baum. Neural net algorithms that learn in polynomial time from examples and queries. *IEEE Transactions of Neural Networks*, 2(1), 1991.

[Bennett, 1991] S. Bennett. Planning to address uncertainty: An incremental approach employing learning through experience. In *Proceedings of the DARPA Workshop on Advanced Planning and Control Strategies*, pages 313–324, 1991.

[Cohn, 1994] D.A. Cohn. Queries and exploration using optimal experiment design. In *Neural Information Processing Systems 6*, Denver, CO, 1994. Morgan-Kauffman.

[Dunn and Segen, 1988] G. B. Dunn and J. Segen. Automatic discovery of robotic grasp configuration. In *Proceedings, 1988 IEEE International Conference on Robotics and Automation*, pages 396–401, 1988.

[Kaelbling, 1990] L. P. Kaelbling. *Learning in Embedded Systems.* PhD thesis, Stanford University, Stanford, CA, 1990. Dept. of Computer Science.

[Kamon *et al.*, 1994] I. Kamon, T. Flash, and S. Edelman. Learning to grasp using visual information. Technical report, The Weizmann Institute of Science, Rehovot, Israel, 1994.

[Larson and Marx, 1986] R. Larson and M. Marx. *An introduction to mathematical statistics.* Prentice Hall, Englewood Cliffs, NJ, 1986.

[Mel, 1991] B. Mel. *Connectionist robot motion planning: A neurally inspired approach to visually guided reaching.* Academic Press, San Diego, CA, 1991.

[Moore, 1990] A. W. Moore. Acquisition of dynamic control knowledge for a robotic manipulator. In *Proceedings of the seventh international conference on machine learning*, 1990.

[Moore, 1991] A.W. Moore. Knowledge of knowledge and intelligent experimentation for learning control. In *Internation Joint Conference on Neural Networks*, Seattle, WA, July 1991.

[Quinlan, 1986] J. R. Quinlan. *Learning efficient classification procedures and their application to chess end games*, pages 463–481. Morgan-Kauffman, Los Altos, Ca., 1986.

[Ritter *et al.*, 1991] H. Ritter, T. Martinetz, and K. Schulten. *Neural Computation and Self-Organizing Maps: An Introduction.* Addison-Wesley, Reading, Massachussetts, 1991.

[Salganicoff and Bajcsy, 1992] M. Salganicoff and R. Bajcsy. Robot sensorimotor learning in continuous domains. In *1992 IEEE Conference on Robotics and Automation*, Nice, France, May 1992.

[Salganicoff, 1992] M. Salganicoff. *Learning and Forgetting for Perception-Action: A Projection-Pursuit and Density-Adaptive Approach.* PhD thesis, University of Pennsylvania, 1992. Dept. of Computer and Information Science.

[Salganicoff, 1993] M. Salganicoff. Density-adaptive learning and forgetting. In *Proceedings of the Tenth International Workshop on Machine Learning.* Morgan-Kauffman, June 1993.

[Schneider, 1993] J. Schneider. Open loop motor skill learning. In *AAAI Fall Symposium Series: Machine Learning in Vision: What Why and How?*, Raleigh, N.C., 1993.

[Solina, 1987] F. Solina. Shape recovery and segmentation with deformable part models. Technical Report MS-CIS-87-111, University of Pennsylvania, Dept. of Computer and Information Science, Philadelphia, Pa., 1987.

[Sutton, 1990] R. S. Sutton. Integrated architectures for learning planning and reacting based on approximating dynamic programming. In *Proceedings of the Seventh Internations Conference on Machine Learning.* Morgan-Kauffman, 1990.

[Tan, 1990] M. Tan. A cost-sensitive learning system for sensing and grasping objects. In *Proceedings, 1990 IEEE International Conference on Robotics and Automation.* IEEE Press, 1990.

[Thrun and Moller, 1992] S. Thrun and K. Moller. Active exploration in dynamic environments. In *Neural Information Processing Systems 4*, pages 531–538. Morgan-Kauffman, 1992.

[Utgoff, 1988] P. Utgoff. ID-5: An incremental ID-3. In *Proceedings of the Fifth International WOrkshop on Machine Learning.* Morgan-Kauffman, 1988.