Using Inverse Reinforcement Learning to Predict Goal-directed Shifts of Attention

Gregory J. Zelinsky (Gregory.Zelinsky@stonybrook.edu)
Departments of Psychology and Computer Science, Stony Brook University
Stony Brook, NY 11794 USA

Abstract:

Understanding how goal states control behavior is a question intersecting attention, action, and recognition, and one that is ripe for interrogation by new methods from machine learning. This study uses *inverse-reinforcement learning* (IRL) to learn the reward function and policy underlying the simplest of goal-directed actions—shifts of gaze—in the service of the simplest of goals—finding a desired target category. Training this IRL model of categorical search required the creation of a large-scale dataset of images (4,366) that are labelled with the fixations of people searching for one of two target-category goals (microwaves or clocks). The IRL model is evaluated against a test dataset consisting of the fixations of 60 people searching for either a microwave (n=30) or clock (n=30) in the same images. The IRL model successfully predicted behavioral search efficiency and fixation density maps using multiple metrics. Moreover, reward maps and action maps recovered by the IRL model revealed target-specific patterns that reflect, not just attention guidance to target features, but also guidance by scene context (e.g., clocks are often on walls). Using methods from machine learning it is now possible to learn the reward functions that more broadly capture the target-object context.

Keywords: Inverse-reinforcement Learning; Imitation Learning; Goal-directed Behavior; Categorical Search; Fixation Prediction; Visual Search Dataset

Introduction

Most cognitively-meaningful behavior is made in the service of a goal, but computational models of goal-directed attention control are still in their infancy. Several models use *saliency maps* to predict fixations during free-viewing (Borji, Sihite, & Itti, 2013), but this task is largely *absent* a goal. Here we focus on the simplest goal-directed task, *categorical visual search*, the search for a target-object goal of a particular category (Schmidt & Zelinsky, 2009). Existing models of categorical search predict fixations by comparing learned target features to image features to compute a *priority map* (e.g., Zelinsky, Adeli, Peng, & Samaras, 2013; Adeli & Zelinsky, 2018). Here we conceptualize priority as *expected reward*. By assigning reward to labelled person-like behavior, we can learn a *reward function* and use it to predict fixations made during the goal-directed search for a target category.

Inverse Reinforcement Learning

Inverse reinforcement learning (IRL) is a method for learning a mapping from a *State* to an *Action*, termed a *Policy*, based on the selective application of reward. Our implementation makes the reward proportional to the model's ability to make State-Action pairings that mimic or imitate observed State-Action pairings (Ho & Ermon, 2016). Over training, and through the greedy maximization of total expected reward, the model learns a Policy (or reward function) that can be used to predict new Actions given new States. In the current context, the Actions are shifts in fixation location over an image (saccades), and the State is the *search context*, which can be understood as the totality of available information for use in the search task. This includes (but is not limited to) the input image and learned visual features of the target category. IRL assumes that the Policy learned through previous observations of people searching for a target category will predict the fixation behavior of new people searching for the same category in new images.

Model Methods

Model training can be broadly conceptualized as an *Actor* (A) and a *Discriminator* (D) locked in an adversarial process (Ho & Ermon, 2016), one that is driven by the maximization of total expected reward (Fig. 1). The Actor generates eye movements (actions) with the goal of fooling the Discriminator into believing that they were made by a person. The Discriminator tries to discriminate real fixations from the Actor's fixations, with greater reward assigned to person-like actions fooling the Discriminator. This reward-driven adversarial process plays out during training, with the result being an Actor who becomes highly adept in imitating the behavioral fixations made during categorical search. Specifically, at training, **D** provides immediate reward after each state-action pair by **A**, and from this a Policy is learned that maximizes the cumulative reward across all pairings. At testing, this trained Policy is used to predict fixations made to the target category in response to new States.

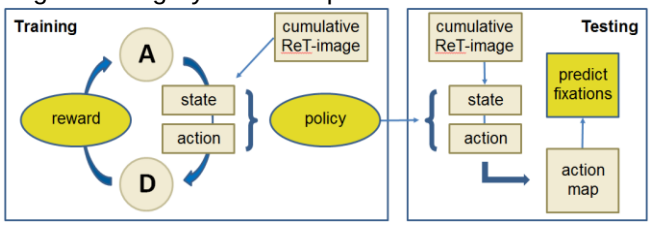

Figure 1: The adversarial IRL model of categorical search.

Cumulative Movements of a Foveated Retina

A novelty of our approach is that it integrates state-of-the-art IRL with a fixation-based state encoder, which accumulates high-resolution visual information with each movement of a 16×16 pixel fovea. We gave our model a fovea, without which changes in fixation would be unnecessary, by applying the method from Geisler and Perry (2002) to an input image to create what we call a *Retina-Transformed (ReT) image*. Figure 2 shows an example of cumulative ReT-images obtained at each of three fixation locations (0,1,2), with the sequence of these images comprising a dynamic state representation. For state encoding we used a pre-trained ResNet-50 that was dilated and fine-tuned on ReT-images. Over sequential fixations, a high-resolution state representation is thus created for predicting goal-directed attention control. To define the action space for the IRL model, a ReT-image is discretized into a 10×16 grid of 32×32 pixel cells, with the center of each grid cell corresponding to a potential fixation location. At each time step, one of these 160 possible locations is selected for an action.

Figure 2: A dynamic state representation from cumulative ReT-images. Note how each fixation (left to right) progressively de-blurs an initially blurred image input.

The Microwave-Clock Search Dataset

The most predictive models of fixation behavior are in the context of free-viewing, where models are trained on large image datasets that are annotated with fixations made during a free-viewing task (Kummerer, Wallis, Gatys, & Bethge, 2017). Training is also required to learn a target category's reward function, but there is spotty availability to suitably large datasets of fixations made during categorical search. Those that do exist are either limited to person search (Ehinger, Hidalgo-Sotelo, Torralba, & Oliva, 2009), or are broader but do not use a standard search task.

The MCS dataset consists of COCO2014 (Lin et al, 2014) images labeled as containing either a microwave oven or a clock, from which we created disjoint training and testing datasets. Image selection excluded scenes depicting a person or animal, and digital clocks in the case of the clock target category. This latter constraint was introduced because the features of analog and digital clocks are very different, and we were concerned that this would introduce unwanted variability in the search behavior. No additional exclusion criteria were used to select the training images, with our goal being to include as many images for training as possible. This left 1,494 analog clock images and 689 microwave images, which varied greatly in terms of search difficulty (Fig. 3, top). Test images were fewer in number (n=40), allowing their selection to be more highly controlled. Test images were further constrained to have: (1) depictions of *both* a microwave and a clock (allowing different targets to be designated in the identical images), (2) only a single instance of the target, (3) a target's size less than 10% of the image size, and (4) targets not appearing at the image's center (based on a 5×5 grid). These exclusion criteria resulted in the test images being fairly well controlled and aimed at a moderate level of search difficulty (Fig 3, bottom).

Figure 3: Top, examples of clock (left 3) and microwave (right 3) training images. Bottom, examples of test images, each depicting both a microwave and a clock.

The above-described selection criteria were specific to target-present (TP) images, but an equal number of target-absent (TA) images (n=2183) were selected so as to create a standard TP versus TA search context. These images were selected randomly from COCO, with the constraints that: (1) none depicted the target, and (2) all depicted at least two instances of the target's siblings. COCO defines the microwave siblings to be ovens, toasters, refrigerators, and sinks, under the parent category of "appliances". Clock siblings are defined as: books, vases, scissors, hairdryers, toothbrushes, and teddy bears, under the parent category of "indoor". Sibling membership was used as a selectin criterion so as to discourage target-absent responses from being based on scene type (e.g., a street scene is unlikely to contain a microwave).

The large size of the dataset required data collection to be distributed over groups of searchers. Each microwave training image was searched by 2-3 people (n=27); each clock training image was searched by 1-2 people (n=26). Test images were each searched by a new group of 60 people, 30 for a microwave target and 30 for a clock target in a between-subjects design.

Behavioral Search Procedure

A standard categorical search paradigm was used for both training and testing (Fig. 4). TP and TA images were randomized within target type, and searchers made a speeded TP or TA response terminating each trial. Search display visual angles were 54°×35° for testing; for training angles ranged from 12°×28.3° in

width and 8°×28.3° in height. Eye position was sampled at 1000 Hz using an EyeLink 1000 in tower-mount configuration (average spatial error < 0.5°).

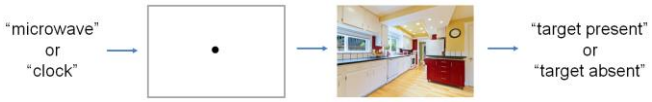

Figure 4: Categorical search paradigm.

Results

Search Behavior

Table 1 provides summary error and number-of-fixation data, but the key behavioral pattern is plotted in Figure 5. Fixations on TP trials were strongly guided to both the microwave and clock targets, as evidenced by the higher probability of gaze landing on the targets relative to object-based chance baselines. Given the importance of the first 6 fixations in the behavioral data, the model outputted fixed 6-fixation segments.

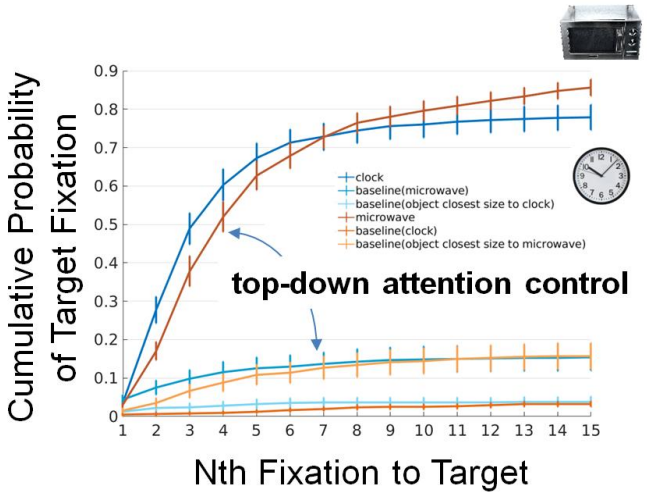

Figure 5: Cumulative probability of fixating the target in the target-present test dataset, relative to baselines.

Table 1: Errors and mean number of fixations before the button press, grouped by dataset, target type, and TP/TA.

Training Dataset	Target Category	Error (%)	Mean (SD) Fixations
target-present	microwave	18	5.46 (±2.6)
	clock	15	4.52 (±3.5)
target-absent	microwave	8	7.95 (±4.1)
	clock	10	11.14 (±6.8)

Testing Dataset	Target Category	Error (%)	Mean (SD) Fixations
target-present	microwave	9	6.76 (±2.1)
	clock	6	5.33 (±1.8)
target-absent	microwave	4	14.36 (±2.5)
	clock	5	15.85 (±2.3)

IRL Model

To determine whether the model's behavior is reasonable, we visualize in Figure 6 the reward maps (middle) and action maps (right) preceding and following the initial shift of its gaze (left, top to bottom) in a microwave search task. Greener colors on the reward maps indicate image locations associated with greater immediate reward, and bluer colors on the action map indicate greater total reward expected if fixation shifted to a given location. The clear alignment between the generated behavior and the distributions of both immediate and total reward suggests that the model has learned to associate a State (e.g., the features of a microwave in this image) with an Action.

Figure 6: IRL model searching for a microwave target (red box) in a test image. Right, action maps. Middle, reward maps. Left, cumulative ReT-images based on the first fixation (top row) and second fixation (bottom row).

Figure 7 is also a qualitative evaluation, this time comparing fixation-density maps (FDMs) from people searching for a microwave (n=30) or a clock (n=30) in two test images to FDMs generated (sampling from probabilistic policy) by the model as it searched for the same targets in the same images. Not only did the model find the targets, but depending on the target category it also prioritized actions to different scene regions, similar to what was observed in behavior.

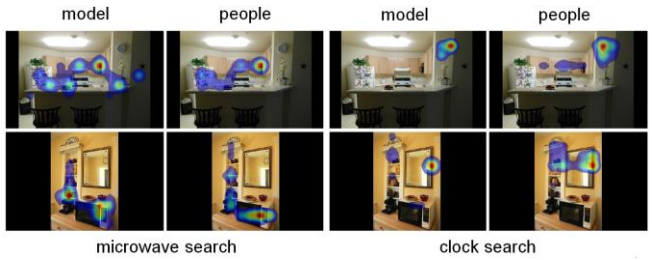

Figure 7: Fixation-density maps from the model and people searching for a microwave (left 4 panels) and clock (right 4).

We conducted several analyses comparing the model's search behavior to the behavior of people searching the test dataset. These analyses are summarized in Table 2 for search efficiency and in Table 3 for the model's success in predicting behavioral FDMs. A notable finding from Table 2 is that both model and behavioral search efficiency was high, with the better performing of the two depending on the target type. For microwave search, the model found the target more efficiently than the behavioral searchers by three different metrics (excluding errors). These include: Percentage of trials in which the target was located in the first six fixations, the average

number of fixations in these fixated-in-6 trials, and scanpath ratio, which is a measure of search efficiency defined as the distance between the starting fixation location and the target divided by the summed saccade distance. However, why this advantage occurs is uncertain, with possible causes ranging from error in the ReT-images to people failing to behave as optimally as an IRL model. Perhaps more informative is that for clock search the model outperformed people on only one measure of efficiency, on the other two measures the behavioral performance was equal to or more efficient than the model. This suggests that small differences in metric values may not be meaningful.

Table 3 reports how well the model's FDMs predicted the FDMs of the behavioral searchers. Prediction success was quantified using three measures: Pearson's Correlation Coefficient (CC), Normalized Scanpath Saliency (NSS; Bylinskii et al., 2016), and Area Under the ROC Curve (AUC). Higher values for all measures indicate better prediction. We also report predictions from a Subject Model, computed by having *n-1* searchers predict the behavior of the searcher that was left out, which establishes a practical upper limit on a model's ability to predict a behavior. Of these metrics, AUC yielded the best agreement between IRL and Subject models, while CC yielded the poorest IRL model predictions.

Table 2: Search efficiency

Microwave Search	Fixated in 6 (%)	Average Fixations	Scanpath Ratio
Subjects	69.1	4.3	.54 (±.23)
IRL Model	80.0	1.8	.78 (±.26)
Clock Search			
Subjects	81.3	4.0	.65 (±.22)
IRL Model	77.5	2.3	.66 (±.35)

Table 3: FDM prediction

Microwave Search	CC ↑	NSS ↑	AUC ↑
Subject Model	.885	1.680	.768
IRL Model	.435	1.149	.697
Clock Search			
Subject Model	.887	1.568	.754
IRL Model	.398	1.013	.669

Conclusions

We showed that a model driven by reward can predict the goal-directed control of attention. This demonstration required first assembling a dataset of search behavior large enough to train an Inverse Reinforcement Learning model. Creation of this dataset is itself a contribution, enabling other models of goal-directed search behavior to be trained and used for more extensive model comparison. Having learned reward functions for two target category goals, these functions were used to predict the fixations made by new searchers searching for those same target categories in a new testing dataset, where we found good prediction of both search efficiency and the spatial distribution of search fixations. Ongoing work is extending the Microwave-Clock Search dataset to include 18 target categories, thus enabling the relationship between attention and reward to be studied across a category structure.

Visual search is a goal-directed behavior of unique importance, shared by pigeons and people and most species in between. Perhaps because of its fundamental role in survival, we believe that search relies on the most basic of control mechanisms—*reward*. Using methods from machine learning it is now possible to learn reward functions that define the target goals used in the cognitive control of attention.

Acknowledgments

I would like to thank the National Science Foundation for their generous support (award IIS-1763981), and all of the members of the EyeCog and Computer Vision labs at Stony Brook University for their hard work and invaluable feedback.

References

Adeli, H. & Zelinsky, G.J. (2018). Deep-BCN: Deep networks meet biased competition to create a brain-inspired model of attention control. *CVPRw*, pp. 1932-1942.

Borji, A., Sihite, D.N, & Itti, L. (2013). Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. *IEEE Transactions on Image Processing*, *22(1)*, 55-69.

Bylinskii, Z., Judd, T., Oliva, A., Torralba, A., & Durand, F. (2016). What do different evaluation metrics tell us about saliency models? *arXiv*, 1604.03605.

Ehinger, K.A., Hidalgo-Sotelo, B., Torralba, A., & Oliva, A. (2009). Modelling search for people in 900 scenes: A combined source model of eye guidance. *Visual Cognition*, 17(6-7), 945–978.

Geisler, W.S., & Perry, J.S. (2002). Real-time simulation of arbitrary visual fields. *ACM ETRA*, pp. 83-87.

Ho, J., & Ermon, S. (2016). Generative adversarial imitation learning. *NIPS*, pp. 4565–4573.

Kummerer, M., Wallis, T.S., Gatys, L.A., & Bethge, M. (2017). Understanding low-and high-level contributions to fixation prediction. *CVPR*.

Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C.L. (2014). Microsoft COCO: Common objects in context. *ECCV,* pp. 740-755.

Schmidt, J., & Zelinsky, G.J. (2009). Search guidance is proportional to the categorical specificity of a target cue. *Quarterly Journal of Experimental Psychology*, *62(10),* 1904-1914.

Zelinsky, G.J., Adeli, H., Peng. Y., & Samaras, D. (2013). Modelling eye movements in a categorical search task. *Philos Trans R Soc Lond B Biol Sci*, 368:1-12.