Category Learning: Comparison of computational and human methods

Turner Bohlen Troy Astorino

December 31, 2012

Introduction

Rational analyses of human cognition seek to explain and quantify human behavior and thought processes under the assumption that they are an optimal adaptation to the constraints of the environment. Anderson [1] argues that categorization is a basic function of human cognitive processes, and that Bayesian statistical inference is a theoretically motivated and effective model for human categorization. Sanborn, Griffiths, and Navarro [2] further investigate Bayesian algorithms for category learning, finding a single-particle particle filter to be most effective, and most similar to human behavior.

To extend upon the previous work concerning particle filters and category learning, this experiment analyzes each individual move made by a human while sorting data, rather than simply analyzing the end result.

There are two hypothetical although quite appealing arguments for this more fine-grained approach. First, a single end sort can be achieved by a number of paths exponential in the size of the largest category. Analyzing the process step by step allows access to this vast amount of missed information, assuming it can be analyzed in some useful way. In this way, move-by-move analysis is an exponentially tougher test for human inference models. Secondly, by comparing the human and machine category learning at each step, a given machine inference algorithm could be decisively shown not to model all human logic being applied to the given problem if the parameters to the algorithm that result in the most "human" action vary from move to move. Phrased another way, any variation in the most "human" parameters indicates that the humans had to apply additional logic to modify his/her internal model of the system, and so the inference model in question is missing some human reasoning.

Experimental Methods

The task that trial subjects were faced with was designed to facilitate comparison between the subjects' decisions and those that would be made by a particle filter. Modeling sequential, online category learning was useful both because that was the type of categorization algorithm developed in previous work, and because it has a straightforward manifestation in a constrained human task. Trial subjects were presented with stimuli images, sequentially, and asked to assign each image into a group. Once the image was placed in

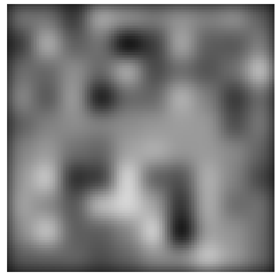

Figure 1: A sample stimulus to be categorized.

a group, the subject was not allowed to switch the image to another group. The instructions given to each subject can be viewed in the Appendix, Figure .

The stimuli used were smoothed, square, 10-by-10 pixel-blocks images. A sample stimulus can be viewed in Figure . Each pixel-block in an image is a shade of gray between white and black, with 256 possible grayscale values. This type of stimuli is convenient because it is directly interpretable for the inference algorithm: each pixel-block is a feature, and each feature can take one of 256 possible values. This type of stimuli was also considered attractive because of the difficulty for the human subject of categorizing the images. It was thought that this difficulty would compel the subjects to rely on subconscious processes for categorization. In retrospect, as is addressed in Section , the high dimensionality of the images may have made it too difficult for the brain to categorize the images effectively, and may have lessened the interpretive power of the results. Figure shows how the interface would appear for a mostly completed trial.

Workers from Mechanical Turk were hired to be trial subjects. They were presented with a website on which to conduct the trial: each worker was first shown the instruction page in Figure , and then were brought to the page on which the trial was conducted, show in Figure All the actions performed by the subject were collected in a database to be used for later analysis.

Particle Filter

The algorithm implemented was based on the single-particle particle filter described by Sanborn et al[2]. Justification for the use of a particle filter to perform probabilistic inference for sequential clustering can be found in

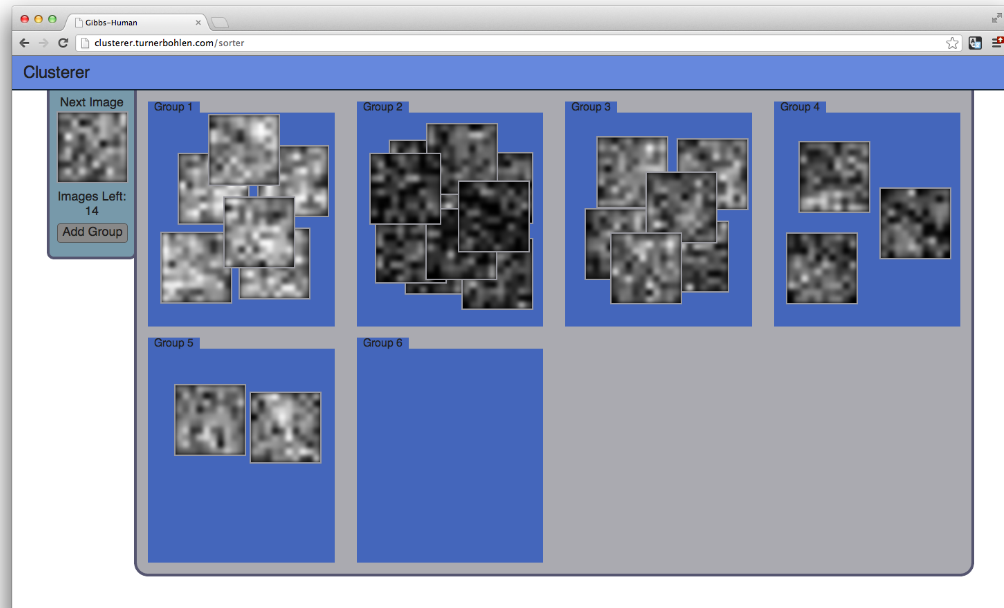

Figure 2: Interface through which a trial was completed. Images are dragged from the 'Next Image' box into one of the groups. Once an image was placed in a group, it could not be switched into another group. The subject was limited to creating 8 groups.

that paper, in addition to a development of the models used to build this particle. Presented below is only a brief description of the underlying probability distributions used, in addition to the modifications required to tailor the particle filter to this experiment.

The posterior probability that a stimulus is assigned to a group is proportional to a prior probability multiplied by a likelihood, as is always the case with Bayesian inference. The prior probability encodes a preference over group sizes: relatively how large groups should be, when new groups should be created, etc. This prior must protect against overfitting, which in this context of category learning would be creating a new category for each slightly different stimulus. The likelihood function encodes the probability that the stimulus is drawn from the same cluster that produced the stimuli already in the cluster. Using the primarily the same notation as the Sanborn et al. paper, the prior takes the form, this proportionality is represented by

$$P(z_i = k|\mathbf{X}_i = \mathbf{x}_i, \mathbf{z}_{i-1}) \propto P(\mathbf{X}_i = \mathbf{x}_i|z_i = k, \mathbf{z}_{i-1})P(z_i = k|\mathbf{z}_{i-1}) \qquad (1)$$

where $z_i = k$ notates assigning the i^{th} stimulus to group k, $\mathbf{z}_{i-1}$ refers to the assignment of groups that the previous $i-1$ stimuli have gone through, and M_k is the number of stimuli in group k after the previous $i-1$ stimuli had been sorted. In Equation 1, $P(\mathbf{X}_i = \mathbf{x}_i|z_i = k, \mathbf{z}_{i-1})$ represents the likelihood and $P(z_i = k|\mathbf{z}_{i-1})$ represents the prior.

A Dirichlet process models the prior distribution over the probability that any given input stimuli will be grouped with a given cluster, whether that is one of the existing clusters or would be a new cluster. This results in the following form for the prior probability:

$$P(z_i = k|\mathbf{z}_{i-1}) = \begin{cases} \frac{M_k}{i-1+\alpha} & M_k > 0 \\ \frac{\alpha}{i-1+\alpha} & M_k = 0 \end{cases} \qquad (2)$$

This says the probability that the i^{th} stimulus is placed in group k is proportional to the number of stimuli already in group k, or to a parameter of the Dirichlet process, α, if group k would be a new group. α is the dispersion parameter of the Dirichlet process; the larger α, the larger the probability that a stimulus will be assigned to a new group. The value used for this parameter, as well as the values used for other parameters of the particle filter, can be found in Figure .

The likelihood model for a stimulus being in a given group assumes that each feature in a group follows a Gaussian distribution. The prior on the variance of this Gaussian is modeled as an inverse χ^2 distribution, and the prior on the mean is modeled as another Gaussian. These priors result in the likelihood function over each feature having the form of a Student's t distribution with a_i degrees of freedom.

$$X_{i,d}|z_i = k, \mathbf{z}_{i-1} \sim t_{a_i}\left(\mu_i, \sigma_i^2\left(1 + \frac{1}{\lambda_i}\right)\right) \tag{3}$$

where

$$\lambda_i = \lambda_0 + M_k \tag{4}$$
$$a_i = a_0 + M_k \tag{5}$$
$$\mu_i = \frac{\lambda_0\mu_0 + M_k\bar{x}}{\lambda_0 + M_k} \tag{6}$$
$$\sigma_i^2 = \frac{a_0\sigma_0^2 + (n-1)s^2 + \frac{\lambda_0 M_k}{\lambda_0 + M_k}(\mu_0 - \bar{x})^2}{a_0 + M_k} \tag{7}$$
$$\tag{8}$$

$X_{i,d}$ is a random variable for feature d of the i^{th} stimulus. In Equation 3, $X_{i,d}$ is conditioned on the i^{th} being assigned to group k and the group assignments of the previous $i-1$ stimuli. M_k is again the number of elements in group k, but in this instance assuming that the i^{th} stimulus has been added to group k. The prior mean is μ_0, and the prior variance is σ_0^2, and the confidences in the prior mean and prior variance are λ_0 and a_0, respectively. μ_0 is set to midpoint of the potential values for each feature, and σ_0 is set to be $1/8^{th}$ the range of the potential values.

The input were 100 pixel square images, with each pixel taking on a grayscale value between 0 and 255. Each feature was treated as an independent feature. Because the range for each feature was limited and discrete, the Student's t distribution was discretized and renormalized along the valid range of the feature each time a feature likelihood value was calculated.

In order to match the methodology used by Sanborn et. al, the features are treated as being independently distributed, so the likelihood for the entire stimulus is simply a product over all the feature likelihoods:

Parameter	Value
α	30
μ_0	127.5
λ_0	0.5
σ_0	32
a_0	2.0

Table 1: The parameters used for the particle filter. α is the dispersion parameter for the Dirichlet process prior. μ_0 and σ_0 are the mean and standard deviation of the prior Gaussian distribution over each feature, and λ_0 and a_0 are the confidence in the prior mean and the confidence in the prior variance, respectively.

$$P(\mathbf{X}_i = \mathbf{x}_i | z_i = k, \mathbf{z}_{i-1}) = \prod_d P(X_{i,d} = x_{i,d} | z_i = k, \mathbf{z}_{i-1}) \tag{9}$$

Additionally, the particle filter was only allowed to create 8 groups, to match this limitation that was placed on human subjects. The restriction was enacted by not allowing the particle filter to consider placing a stimulus into a new group after 8 groups had already been created.

Results

Results from the analysis. Just presentation of the results and explanation, with interpretation limited to straightforward analysis.

Discussion

More full discussion of the results, with more allowances for speculative analysis. Potential directions for future work. Unanswered questions

Some potential issues are listed below:

Since each pixel is directly treated as a feature, the assumption of independence of the features does not fit well with how human beings will perceive images. This could be remedied by performing feature extraction on the pixelated images, and passing the extracted features the particle filter. The feature extraction could be designed such that feature independence assumption would safely be valid for humans. Alternatively, the likelihood model could be modified to account for possible covariance between dimensions.

Could have been some issues with the underlying probability distributions:

- problems with discretizing the student's T distribution? Assumed to be an acceptable approximation of a continuous distribution to a discrete domain

- limiting the number of groups to 8 groups without quantifying this in the prior probability distribution over groups

Layout could have biased choices (older choices covered up by newer ones, longer drags are performed less often, knew there would only be 8 groups allowed, group boxes all fixed in size -¿ trying to make groups approximately the same size, knowing how many stimuli needed to be categorized may have allowed people to preplan group sizes, etc) NOTE: we probably shouldn't include all of these, but there are just what came to my head haha

Checking the data to make sure the Turkers didn't do junk trials

Conclusion

Quick recap; significance of question, what the results were, etc.

References

[1] J.R. Anderson. The adaptive nature of human categorization. *Psychological Review*, 98(3):409, 1991.

[2] A.N. Sanborn, T.L. Griffiths, and D.J. Navarro. Rational approximations to rational models: alternative algorithms for category learning. *Psychological Review*, 117(4):1144, 2010.

Appendix

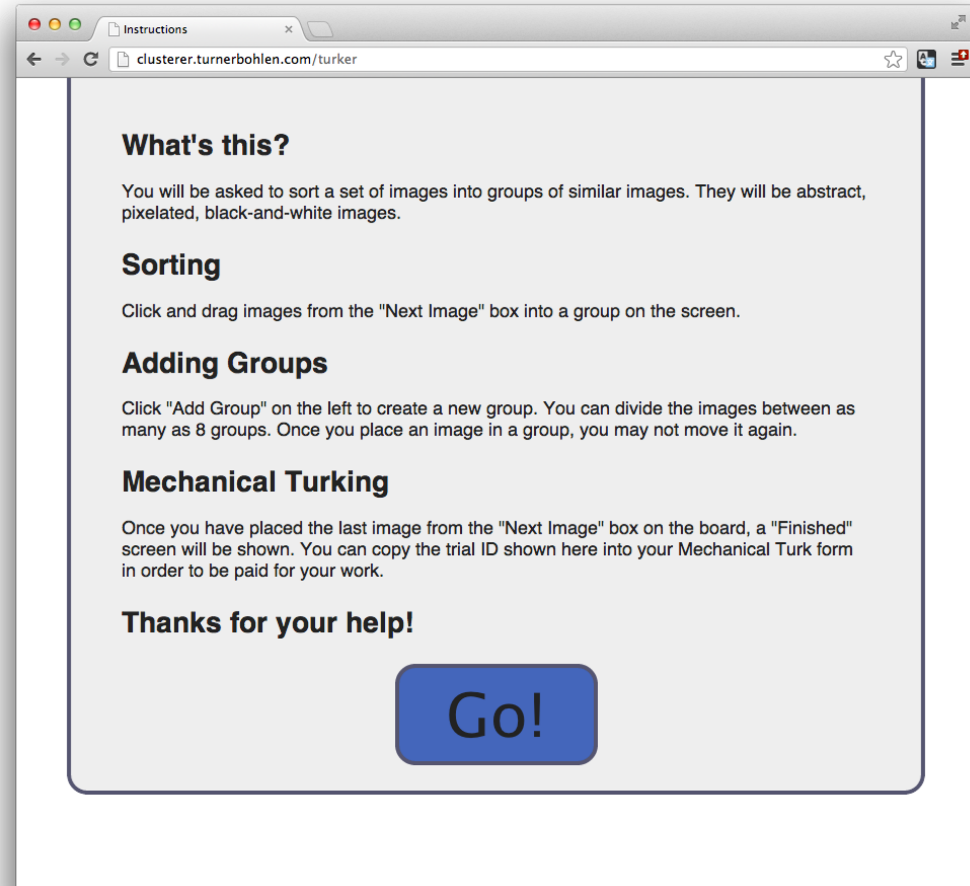

Figure 3: The instructions presented to the Mechanical Turk worker before starting a trial.