Extreme Value Theory and Visual Recognition

Rachel Moore
Department of Computer Science
University of Colorado, Colorado Springs

***Abstract* – The fields of machine learning and psychology have begun to merge, particularly in the subject of vision and recognition. This paper proposes an experiment on human recognition and categorization, using arbitrary images as stimuli. The data will be fitted to an Extreme Value Theory based model, which we hope will give clearer incite into the ways humans categorize novel information.**

***Index Terms* – Recognition, Category Learning, Machine Learning, Extreme Value Theory, Cognitive Psychology.**

I. Introduction

Training set selection is one of the most crucial steps in machine learning. If one wishes for a machine to identify images of apples, only providing images of oranges during training is, in most cases, counter productive. Traditionally, training sets have been selected so that was a wide array of data, so that the training set would closely match a gaussian, or normal, distribution [1]. However, this can become expensive, particularly in tasks that require labeled data for supervised learning. An extensive amount of data is also needed, as smaller sets are less likely to have a normal distribution, which can cause high variance responses and inconclusive results [1].

There have been many advances in the area of training data selection, but there is still a need for methods that select the best training data from small datasets. Simple methods like bootstapping can be used to generate new data in cases of small data sets. Many of these methods do not work with certain learning models [1] [2]. To understand what makes an effective training set, it is important to study how training affects the ultimate categorization. One way of doing this is to study recognition and categorization in humans.

The ability to categorize is one of the most crucial skills we develop as children. Despite its importance, the way we organize information is still a mystery. There are many models of categorical learning in psychology, and more are in development. Studies, such as the one done by Hsu and Griffiths [3] (discussed in Section II), have given some insight into the category learning process, and have yielded interesting results. However, the Gaussian models currently being used on these types of experiments are not capturing the extremes in the data, or the participants' bias towards one category or another. From our research, Extreme Value based models in machine learning have been shown to be a better predictor of human response frequency than Gaussian models. In summary, there are 4 main contributions of this paper:

- Extreme Value Theory and its application.
- Empirical evaluations and metrics for this research.
- Experimental results
- The future of this work.

II. Background

In this section, we discuss Extreme Value Theory (EVT) and its applications, as well as studies involving categorization and EVT modeling. We will also explore psychological research involving categorical learning, which will be the basis for our experiments.

A. Extreme Value Theory

The extreme value theorem states that a function with a continuous and closed interval will have a minimum and maximum value [4] [5]. EVT has been implemented as a statistical model in many different fields of research. Hugueny, Clifton, and Tarassenko [6] used EVT as the basis to create a new model for intelligent patient monitors. The current monitors they reviewed set off false alarms constantly, to the point that hospital staff ignored them. The model they proposed would be less likely to do this, as the EVT-based model would be able to differentiate between truly non-extreme changes in vitals and clear abnormality. EVT has also been used in machine learning to normalize recognition scores [7], which may skew distributions due to outliers.

This research seeks to establish a new EVT-based model of visual recognition and categorization. Particularly, this model may be instrumental for tasks that wish to replicate human information processing. There are three types of extreme value distributions:

Type 1, Gumbel-type distribution:

$$PR[X \leq x] = exp[-e^{x-\mu/\sigma}]. \tag{1}$$

Type 2, Fréchet-type distribution:

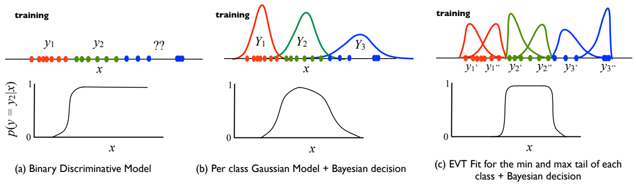
$$PR[X \leq x] = \begin{cases} 0, & x < \mu, \\ exp\left\{-\frac{x-\mu}{\sigma}^{-\xi}\right\} & x \geq \mu. \end{cases} \tag{2}$$

Type 3, Weibull-type distribution:

$$PR[X \leq x] = \begin{cases} exp\left\{-\frac{x-\mu}{\sigma}^{\xi}\right\}, & x \leq \mu \\ 0 & x > \mu \end{cases} \tag{3}$$

where μ, $\sigma(> 0)$ and $\xi(>$s $0)$ are the parameters [5].

EVT-based models can be used as replacements for Binary and Gaussian models, as EVT-models are able to include multiple classes, and do not rely heavily on norms (see Fig. 1). This can also be helpful in the case of training set selection. For example, say there is a set images of apples that need to be categorized into 2 groups: green granny smith and red delicious. While the first and last apple groups have green and red skin tones, respectively, with slight variations in color. However, in this set of apples are a few fuji apples, whose colors range from ruddy green to orange red, and might be categorized into either of the other apple groups. To make the best predictions on which category each apple belongs to, we can use the EVT to find the apples at the groups' decision boundaries, i.e. the most and least red red delicious apples, and the most and least green granny smith apples. From this we can create a training data set. When these clear decision boundaries are known, anything that lies outside of them, say a greenish red fuji apple, can be categorized as a true outlier or part of a third class in the data.



(a) Binary Discriminative Model

(b) Per class Gaussian Model + Bayesian decision

(c) EVT Fit for the min and max tail of each class + Bayesian decision

Fig. 1. Example of data selected with EVT. Courtesy of Boult

B. Prior Research in Human Visual Recognition

In a two part study, Cohen, Nosofsky, and Zaki [8] examined the effects of class variability on categorization. They hypothesized that the generalized context model (GCM), used to calculate the probability that an item will be categorized into one class or another, would substantially underestimate the degree to which participants would classify stimuli into the categories of high variance (we discuss GCM in greater detail in Section V). They found that the middle stimuli (items that were in between the low variance and high variance classes) were classified into the higher variance category, with the probability of up to .73. The GCM estimated the probability to be as low as .35, significantly below what was indicated by the data.

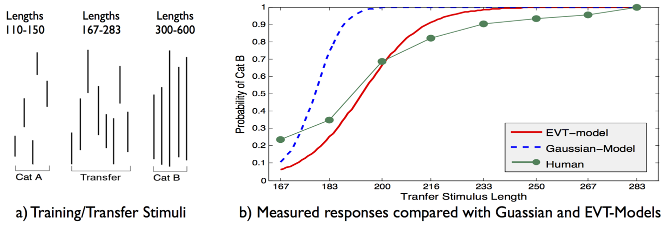
Hsu and Griffins [3] conducted a study in which the participants were taught two alien "languages", consisting of simple images of line segments. Class A had short, low variance line segments, which only differed slightly from one another. Class B had much longer, high variance line segments, in which each line's length was very different from the others. Participants were put into either a generative learning condition or a discriminative learning condition, which varied by the way the training images were presented. In the generative condition, two different cartoon aliens would appear on the screen to indicate which line belonged to which tribe's language. In the discriminative condition, one cartoon alien appeared as a single translator, indicating which language was language was on the screen. After training, participants were shown line segments that were between the lengths of the low and high variance classes and asked to categorize them.

As with Cohen, Nosofsky, and Zaki's [8] study, the results showed that the participants had a strong bias toward the high variance class (Class B), clustering the middle stimuli with the more diverse lines. They found that their Gaussian-based model did not fit their data accurately, and therefore wondered if the Gaussian assumption did not reflect this type of human recognition.

III. Empirical Evaluations

In this section, we discuss our experimental designs, as well as the metrics and technical approach of our study.

In a pilot study, Boult et. al [1] analyzed the data from Hsu and Griffins' [3] study using an EVT model. Because of the bias toward the high variance class, they believed that an EVT-based model would match human data in a more concise way (see Fig. 2) than Gaussian models.



a) Training/Transfer Stimuli

b) Measured responses compared with Guassian and EVT-Models

Fig. 2. Comparison of Gaussian and EVT-based models with human data. Courtesy of Boult.

For the second half of this pilot study, we have collected our own data. Our experiment expanded on Hsu and Griffins' [3] study, but used EVT-based models. We hope our model will paint a clearer and more accurate picture of the way humans categorize unfamiliar stimuli. Another possible extraneous factor in Hsu and Griffiths' [3] study is the way the alien interpreters (the categories) were presented. Those in their generative group were clearly shown when the category had changed, as the aliens changed depending on the sign. In the discriminative group, there was a single alien which never left the screen, and so the participants may not have noticed the sign change. We have duplicated some of their stimuli to test for this factor (see Fig. 3).

A. Metrics and Design

Our research uses models based on the extreme value distributions. Scheirer et al. [7] define extreme value distributions as "... limiting distributions that occur for the maximum (or minimum, depending on the data) of a large collection of random observations from an arbitrary distribution." In the case of visual recognition and categorization in humans, instead of removing the outliers or having them skew the results, one can normalize them, possibly allowing for a better fitted prediction.

For our experiment, we referred to the generalized extreme value (GEV) distribution, or the combined Gumbel, Frechet, and Weibull distributions. GEV is defined as

$$GEV(t) = \begin{cases} \frac{1}{\lambda} e^{-v^{-1/k}} v^{-(1/k+1)} & k \neq 0 \\ \frac{1}{\lambda} e^{-\left(x+e^{-x}\right)} & k = 0 \end{cases} \tag{4}$$

[1] Personal Communication

where x is equal to $\frac{t-\tau}{\lambda}$, v is equal to $(1+k\frac{t-\tau}{\lambda})$, and k,λ, and τ are the shape, scale, and location parameters.

For stimuli, we created a set of 2 dimensional Non-uniform rational B-spline (NURBS) shapes. NURBS are mathematically based shapes, and can be manipulated through functions and interpolation. In a NURBS parametric form, "... each of the coordinates of a point on a curve is represented separately as an explicit function of an independent parameter" [9]

$$C(u)=(x(u),y(u)) \qquad a \leq u \leq b \tag{5}$$

Where "$C(u)$ is a vector-valued function of the independent variable u", which is within the interval [a,b] (usually normalized to [0,1]) [9]. The NURBS we created look similar to ink blots. Each group of images had points that were interpolated to create a set with two clear classes, and another that was somewhere between those two classes (see Fig. 3 and 4). Four groups of shapes were used, and each group contained 17 images. These stimuli acted as distractor tasks. The other stimuli were white lines of varying lengths, place inside of a black circle. Each set had a total of 18 images. These were based on the stimuli in Hsu and Griffiths' [3] study (see Fig. 5). These stimuli were placed into 3 conditions: generative, discriminative, and enhanced tails.

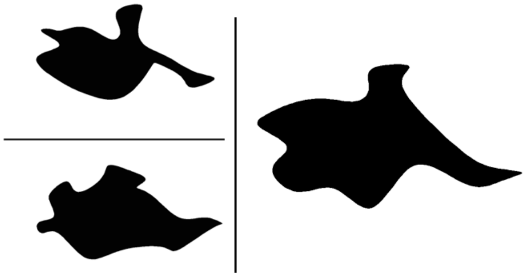


Fig. 3. Example of images duplicated from our NURBs stimuli.

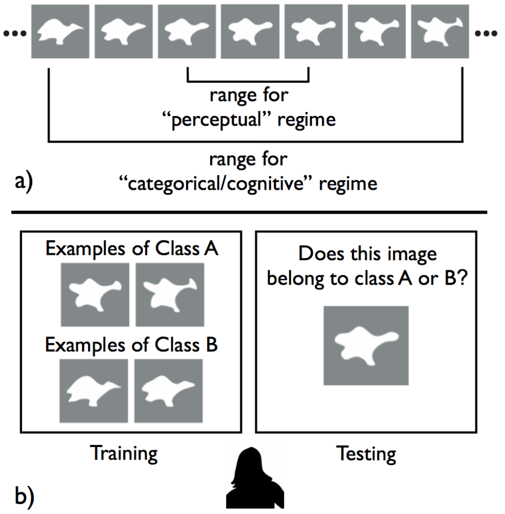


Fig. 4. Example of images from training classes A and B, and a testing image. Courtesy of Boult.

For the experiment itself, we used a program called PsychoPy, version 1.80 [2]. PsychoPy is open source psychophysics software, developed by Piece [10].

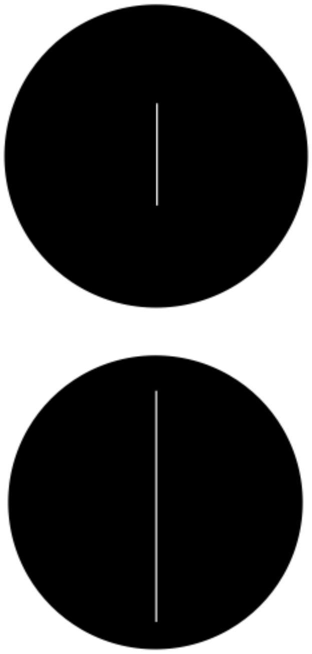


Fig. 5. Example of images duplicated from Hsu and Griffiths' [3].

There were 8 participants total, some of whom took the experiment on multiple occasions. From them, we gathered 30 trials for each of the 3 conditons. Our participants were asked to categorize a series of images into one of two groups, Group 0 or Group 1, and told that there was to be a training component where they would be shown the images and their respective categories, and a testing component where they would label the images themselves. The groups had a separate training set or training style, and testing set. For every group, the participants were trained on the 10 shapes at extrema, 5 from each tail. They were then tested on those same shapes, along with the shapes from the middle of the set, some of which were repeated to make up a total of 20 shapes per training. All trials were repeated, for a total of 20 trial blocks.

IV. Data Analysis

We recorded which middle stimuli were categorized into Group 0 or 1, and the frequency to which these stimuli were placed in these groups (see Fig. 9). The EVT-based model was fitted to the data. It reflected the biases the participants have in categorization, as it did in the pilot study.

A. Factor 1: The Generative Condition

In the generative condition, participants were shown the training stimuli. The training set consisted of lines that were in high variance and low variance categories. The low variance lines were 110, 120, 130, 140, and 150 pixels in length, while

[2] Accessed here: http://www.psychopy.org

the lines in the high variance category were 300, 375, 450, 525, and 600 pixels in length. During training, a box appeared 0.5 sec before before the stimuli, indicating which group the image belonged to. After the stimulus appeared, both the box and the image remained on the screen for 1.5 sec. This was repeated for all 10 stimuli in the training set.

The testing set was comprised of the training set, as well as a set of "middle" stimuli with line lengths of 167, 183, 200, 216, 233, 250, 267, and 283 pixels. The probability that each of these lines would be categorized into the high variance category was 0.17, 0.20, 0.33, 0.4, 0.53, 0.70, 0.77, 0.90, respectfully. The data fit our model well, with the main deviation being at line length 267 (see Fig. 6). Out of the three conditions, this condition was the closest fit to our EVT-based model.

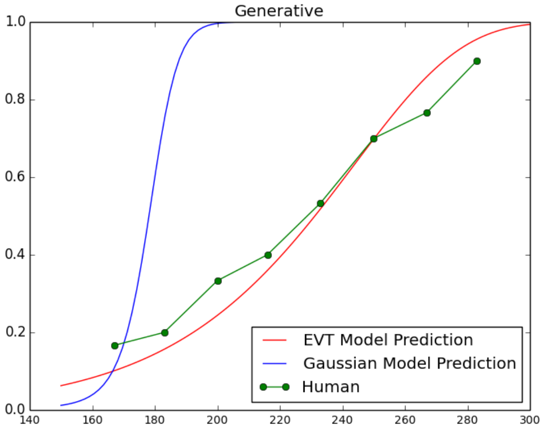


Fig. 6. Comparison of our EVT model with human data for the generative condition.

B. Factor 2: The Discriminative Condition

In the discriminative condition, participants were shown the same training and testing stimuli as the generative condition. However, the indicator box remained on the screen throughout the training session, with only the text changing. Each image still remained on the screen for 1.5 sec. For this condition, the probability that each of the middle stimuli would be categorized into the high variance category was 0.13,0.23, 0.30, 0.37, 0.63, 0.67, 0.83, and 0.90, respectively. The data for this condition also fit our model well, with the main deviation being at line length 233 (see Fig. 7).

C. Factor 3: The Enhanced Tails Condition

The final training set of lines contained the set of low variance lines as the generative and discriminative conditions, but the high variance lines had an elongated tail, with pixel lengths of 300, 375, 450, 600, and 800. The training set up was identical to that of the generative condition. For this condition, the probability that each of the middle stimuli would be categorized into the high variance category was 0.00,0.10, 0.20, 0.20, 0.43, 0.53, 0.60, 0.77, respectively. This shows a shift towards the low variance category. We trained the model on the same set of training data used for the previous conditions for a better visual representation of the bias towards the low variance category (see Fig. 8).

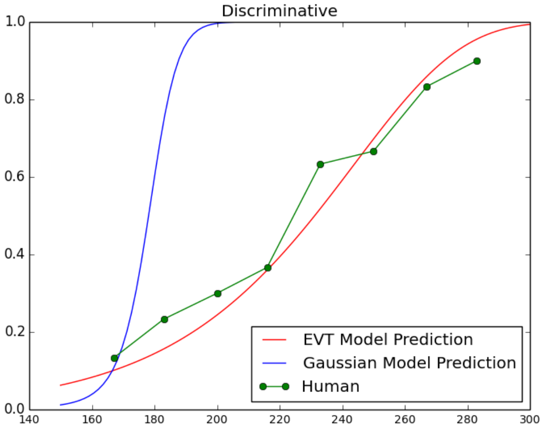


Fig. 7. Comparison of our EVT model with human data for the discriminative condition.

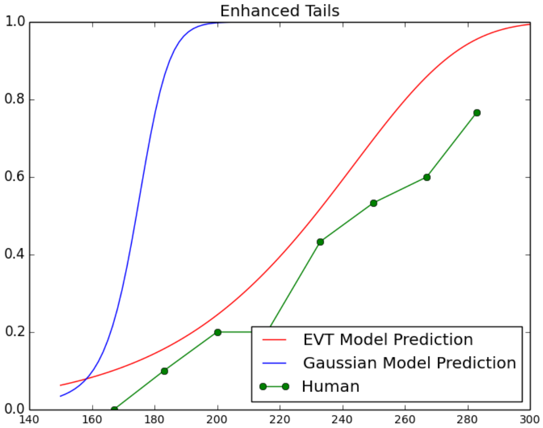


Fig. 8. Comparison of our EVT model with human data for the enhanced tails condition.

D. Comparison

In this section, we will compare the generative, discriminative, and the enhanced tails conditions, and discuss the statistical analysis for the experiment. Fig. 9 is a summary of the probabilities of each condition. The error bars indicate the variance of each line lengths probability. Both the generative and discriminative categories had similar trends. The variance of the generative, discriminative, and enhanced tails conditions were 0.073, 0.083, and 0.072, respectively.

V. Future Work

For our future research, we will incorporate measurements from the NURBS shapes. Because these shapes are mathematically based, the stimuli's dimensions can be easily applied to

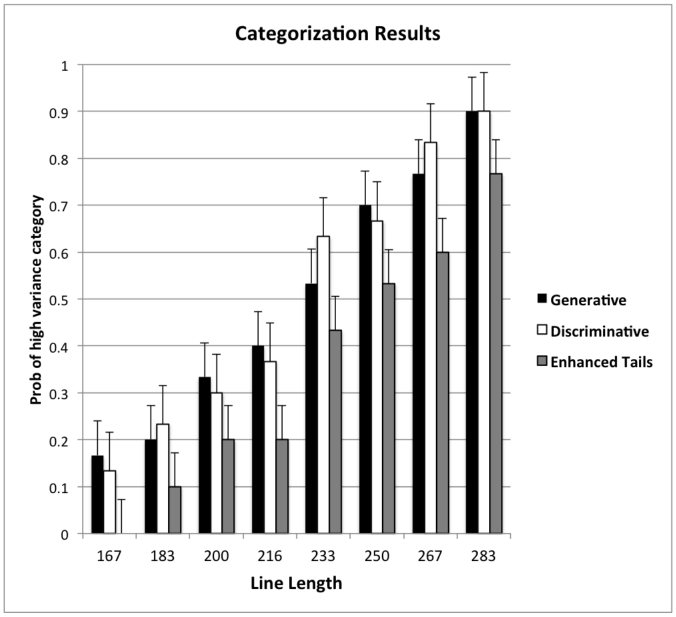


Fig. 9. Probability of categorization of middle stimuli into the high variance category for the generative, discriminative, and enhanced tails conditions.

probability models. One such model, which was mentioned in Section II, is the generalized context model (GCM), which states that ”For the case of two categories A and B, the probability that a given stimulus X is classified in category A is given by

$$P(A|X) = \frac{\beta_a \eta_{XA}^{\alpha}}{\beta_A \eta_{XA}^{\alpha} + (1-\beta_A)\eta_{XB}^{\alpha}} \tag{6}$$

where β_A is a response bias toward category A and η_{XA} and η_{XB} are similarity measures of stimulus X toward all stored exemplars of categories A and B, respectively” [11].

Because our stimuli are so diverse, we plan to make at least one other variation on the current experiment. This may involve changing the task difficulty, the time length, or varying the amount of stimuli in the training sessions.

VI. Conclusion

This paper proposed a new EVT based model for visual recognition. For our purposes, we hope our model will prove to be consistent and accurate in predicting human recognition and categorization. If it is shown to be both of these things, the model could be used to select training sets for machine learning more efficiently, as EVT-based models focus on training data at the extremes, which may cut down on costs of supervised learning. We have seen that EVT-based models can be applied to both generative and discriminative learning situations. We believe that EVT-based models should also be insensitive to the difference between categorical and perceptual learning. With more research, our model may be applied to other human learning tasks, not just visual recognition.

Acknowledgement

I would like to acknowledge Dr. Walter Scheirer, Dr. David Cox, and the team of scientists at Harvard University, who have greatly contributed to this project. I would also like to thank Dr. Terrance Boult, Dr. Lori James, Dr. Kristen Walcott-Justice, and Dr. Jugal Kalita for their invaluable guidance. This project is being supported by NSF REU Grant 1359275.

References

[1] E. Alpaydin, *Introduction to machine learning*. MIT press, 2004.

[2] I. H. Witten, E. Frank, and A. Mark, “Hall (2011).” data mining: Practical machine learning tools and techniques,” 2011.

[3] A. S. Hsu, T. L. Griffiths *et al.*, “Effects of generative and discriminative learning on use of category variability,” in *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, 2010, pp. 242–247.

[4] M. K. Nasution, “The ontology of knowledge based optimization,” *arXiv preprint arXiv:1207.5130*, 2012.

[5] S. Kotz and S. Nadarajah, *Extreme value distributions: Theory and applications*. World Scientific, 2000, vol. 31.

[6] S. Hugueny, D. A. Clifton, and L. Tarassenko, “Probabilistic patient monitoring with multivariate, multimodal extreme value theory,” in *Biomedical Engineering Systems and Technologies*. Springer, 2011, pp. 199–211.

[7] W. Scheirer, A. Rocha, R. Micheals, and T. Boult, “Robust fusion: extreme value theory for recognition score normalization,” in *Computer Vision–ECCV 2010*. Springer, 2010, pp. 481–495.

[8] A. L. Cohen, R. M. Nosofsky, and S. R. Zaki, “Category variability, exemplar similarity, and perceptual classification,” *Memory & Cognition*, vol. 29, no. 8, pp. 1165–1175, 2001.

[9] L. Piegl and W. Tiller, “The nurbs book,” *Monographs in Visual Communication*, 1997.

[10] “Psychopy—psychophysics software in python,” *Journal of Neuroscience Methods*, vol. 162, no. 1–2, pp. 8 – 13, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165027006005772

[11] T. Smits, G. Storms, Y. Rosseel, and P. De Boeck, “Fruits and vegetables categorized: An application of the generalized context model,” *Psychonomic Bulletin & Review*, vol. 9, no. 4, pp. 836–844, 2002.