Requisite Variety

Robert M. Haralick

Computer Science, Graduate Center
City University of New York

Requisite Variety

Requisite Variety was a term introduced by William Ross Ashby to understand how complex a control mechanism must have enough states to return the system (an organism or environment) to a place of homeostasis.

The idea was this. If the control mechanism did not have enough effective states, it would not be able to bring the system to a state in the subset of the desired states. He was arguing that the control mechanism had to have a greater requisite variety than the variety of the disturbances the system has to handle.

Memorization and Generalization

- Memorization
  - The number of observations from the classes are too small
  - The requisite variety of the Machine Learning Algorithm is too large
  - In effect the Machine Learning Algorithm just memorizes the training set
- Generalization
  - The number of observations from the classes is large
  - The requisite variety of the Machine Learning Algorithm is small
  - The Machine Learning Algorithm captures the structural and defining aspect of the classses

Measurement Space Size and Training Set Size

- The size of measurement space is M
- The size of the training sets for each class is N
- $M >> N$ Requisite Variety of Measurement space is too high
  - Every tuple in measurement space, either has observed 0 or 1 training set tuple
  - The likelihood is that a substantial fraction of the testing set tuples will not have been observed
  - Performance on the test set will be poor
  - Memorization
- $M << N$ Requisite variety of Measurement Space smaller than Requisite Variety of Training Set
  - Every tuple in measurement space has observed multiple instances of every class tuple from the training
  - Performance on the test set will be good
  - Generalization

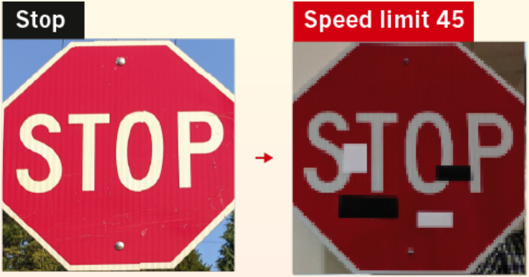
Requisite Variety of Training Sets Too Small

“Deep learning algorithms are well-known to have a propensity for fitting the training data very well and often fit even outliers and mislabeled data points. Such fitting requires memorization of training data labels, a phenomenon that has attracted significant research interest but has not been given a compelling explanation so far.”

“Further, it is now well-known that standard deep learning algorithms achieve high training accuracy even on large and randomly labeled datasets”

Vitaly Feldman and Chiyan Zhang, “*What Neural Networks Memorize and Why: Discovering the Long Tail via Influence Estimation*”, https://arxiv.org/pdf/2008.03703.pdf

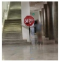
Mistakes Understood By Requisite Variety



Douglas Heaven, "*Why Deep-Learning AIs Are So Easy To Fool*", Nature, October 9, 2019, pp. 1-15.
https://www.nature.com/articles/d41586-019-03013-5

Mistakes Understood By Requisite Variety

Table 1: Sample of physical adversarial examples against LISA-CNN and GTSRB-CNN.

Distance/Angle	Subtle Poster	Subtle Poster Right Turn	Camouflage Graffiti	Camouflage Art (LISA-CNN)	Camouflage Art (GTSRB-CNN)
5′ 0°					
5′ 15°					
10′ 0°					
10′ 30°					
40′ 0°					
Targeted-Attack Success	100%	73.33%	66.67%	100%	80%

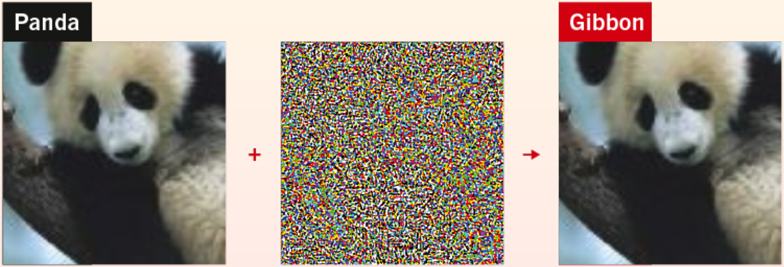
K. Eykholt et. al, “*Robust Physical-World Attacks on Deep Learning Visual Classification*”, IEEE/CVF Conference Computer Vision Pattern Recognition, 2018, pp. 1625-1634

Understanding Requisite Variety

- Similarity
  - The training samples from the Stop Sign class were similar
  - And did not include any stop signs with rectangular blobs
  - The CNN memorized the similar stop sign class
  - And could not generalize to the “perturbed stop sign” class
- The population included a wider variety of patterns not represented in the training set
- The requisite variety of the stop sign class was too low

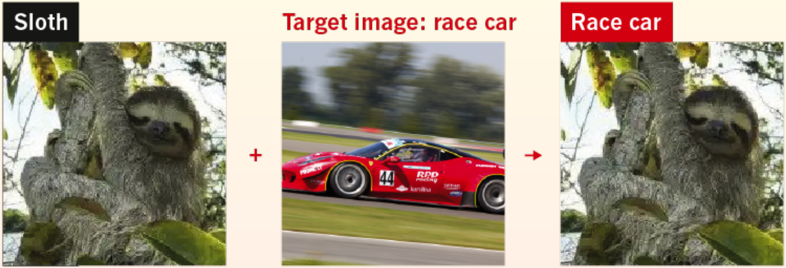
Panda Gibbon Mistake

Gibbon



Add a little colored noise to the Panda image and the CNN identifies it a Gibbon

Race Car Mistake



Add a little of the Race Car Image to the Sloth Image and the CNN identifies the Sloth as a race car.

Douglas Heaven, "*Why Deep-Learning AIs Are So Easy To Fool*", Nature, October 9, 2019, pp. 1-15. https://www.nature.com/articles/d41586-019-03013-5

Adversarial Programming

There is now a subfield called Adversarial Programming. Check out

Gamaleldin Elsayed, Ian Goodfellow, Jascha Sohl-Dickstein,“Adversarial Reprogramming of Neural Networks” https://arxiv.org/pdf/1806.11146.pdf

Requisite Variety of the Decision Rule

- Number of words of memory for tables M_{tables}
- Number of words of memory containing the computations M_c
- The computational complexity of assigning a tuple to a class C
- The Requisite Variety of Decision Rule: $R_{decision} = M_{tables} + M_c + \alpha C$

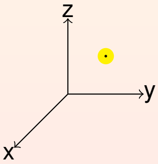
Requisite Variety of the Training Set

- Number of words required to store the Training Set M_{train}
  - $M_{train} > 10R_{decision}$

Requisite Variety of the Class Population

Consider a class c

- Having $Q = Q_1 \cup Q_2 \cup Q_3$ disjoint subclasses
- Q_1 is the subclass for which all its tuples are close to the class mean (Mahalanobis Distance, Euclidean Distance)
- Q_2 is the subclass for which all its tuples are close to subspace A
- Q_3 is the subclass for which all its tuples are far from subspace B



- The Training Set For Each Class
- The Testing Set For Each Class
- Must Be Randomly Sampled For Each Subclass

The Classification Table

1 - Alfalfa
2 - Barley
3 - Safflour
4 - Sugar Beet
5 - Lettuce
6 - Onion
7 - Pasture
8 - Bare Soil

DECISION RULE BOUNDARIES
used by: Discrete Bayes Rule

DATA: Raw, Equal Space Quantized

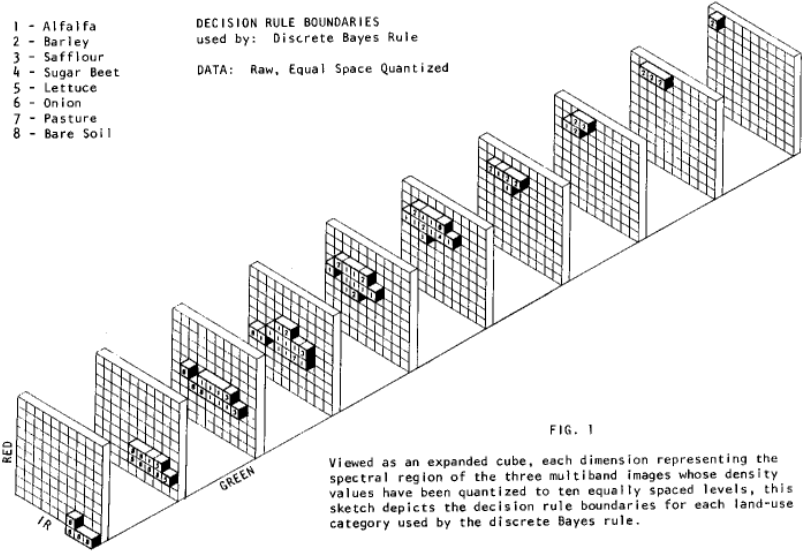


FIG. 1

Viewed as an expanded cube, each dimension representing the spectral region of the three multiband images whose density values have been quantized to ten equally spaced levels, this sketch depicts the decision rule boundaries for each land-use category used by the discrete Bayes rule.

Memorization and Generalization: Requisite Variety

- Discrete Bayes Decision Rule
  - V features each having B values
  - Size of measurement space is B^V
  - Size of training set for each class must be $10B^V$
  - If $B = 10$ and $V = 20$ $B^V = 10^{20}$
  - There is not enough memory to store the classification table
  - There is not enough memory to store the training sets for each class
  - Even on 10 TB disks
- We cannot use a Discrete Bayes Rule

Using Subspaces

- Suppose that we partition the V-dimensional space into M V/M-dimensional subspaces
- Suppose that $M = 4$ and M divides V
- Size of each subspace is $B^{V/M} = 10^5$
- Training set for each class needs to be at least $10 \times B^{V/M} = 10^6$ tuples
- Each tuple having 20 components
- Size of Training set for each class is $20 \times 10^6 = 2 \times 10^7$
- For Each Class
  - Total size of all class conditional probability tables is 4×10^5
- No need to store the Classification Table
  - Any needed entry of the Classification Table is computed on the fly with 4 table look-ups and K class comparisons (trivial)
- Requisite Variety of Training Set for each class is significantly greater than Requisite Variety of Decision Rule