Which of These Does Not Belong?

Anomaly Detection in Images Using a Deep Neural Network and Isolation Forest

Mark Wilber, Galvanize Hiring Day

Motivation

"... as we know, there are known knowns; ... things we know we know. We also know there are known unknowns; ... we know there are some things we do not know. But there are also unknown unknowns ~ the ones we don't know we don't know."

Donald Rumsfeld, 2002

Much of anomaly detection is done using classifiers, *supervised learning*

- Use existing data sets, label anomalies (fraud, network intrusions, etc.)
- Train on anomalies we know about: *known knowns*

Motivation

Alternatives exist

- Suppose that data can be fit by multi-variate Gaussian
- flag as anomalies those points sufficiently far into tail

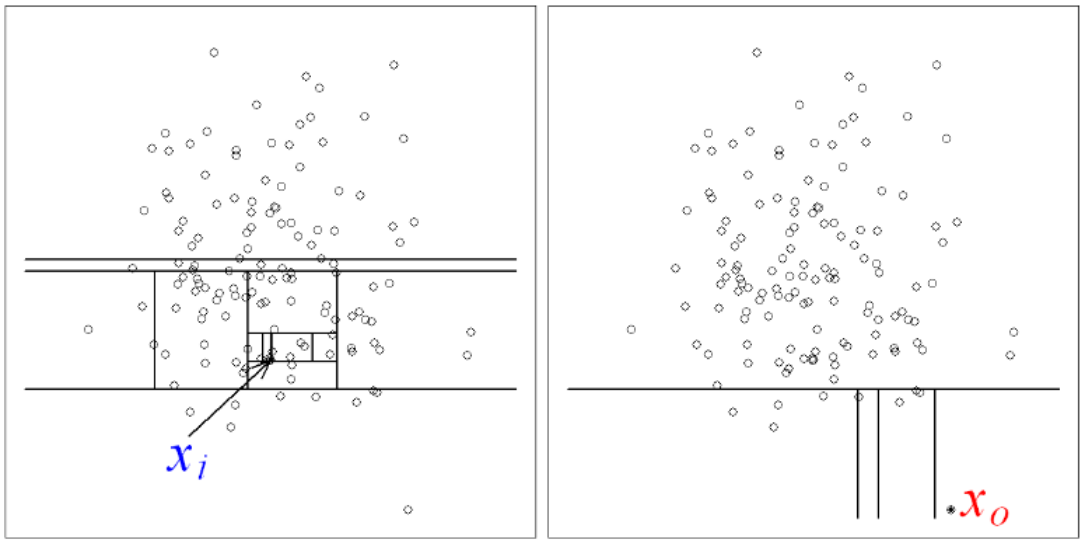
  Will not work if "ordinary" data are in widely-spaced clusters

- Distance-based methods, which become impractical in high-dimensional space, or which require knowledge of how many clusters of normal points there are.

Solution

Isolation Forest[1]

- Un-supervised method of statistically isolating points that have no close neighbors in feature space
- Advantage: no need to compute large numbers of distances to find which points are distant from others.



(a) Isolating x_i (b) Isolating x_o

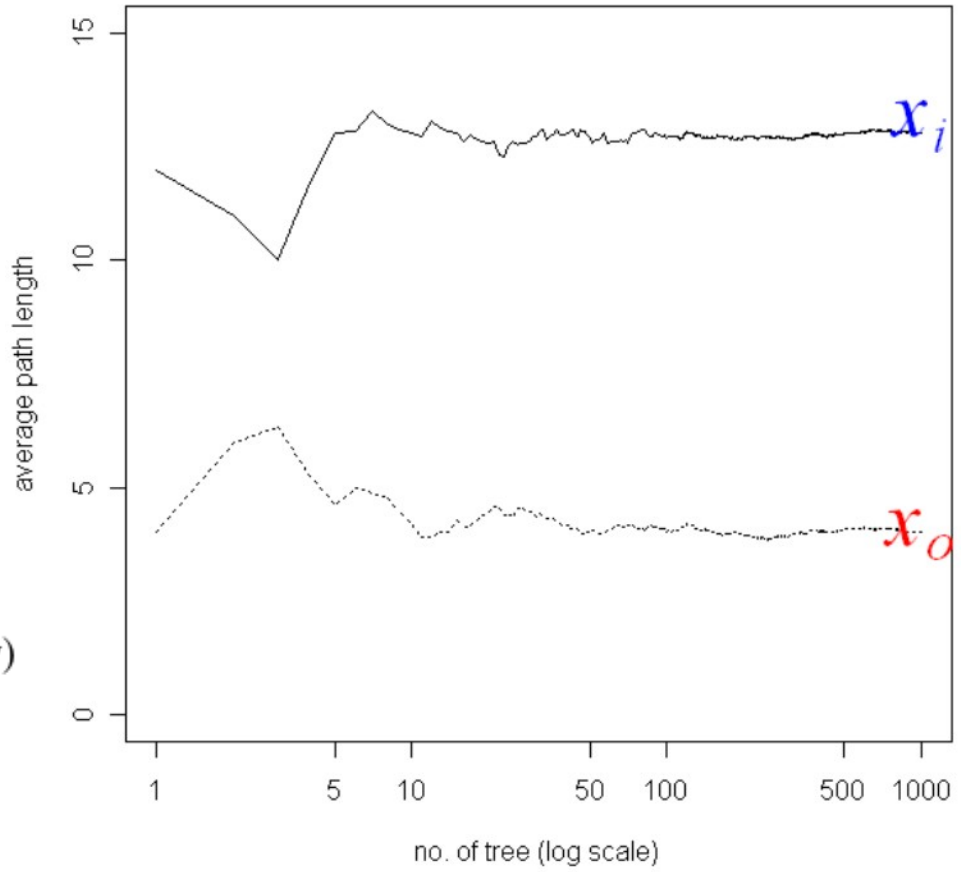
[1]Liu, F.T. et al, ICDM `08b,

Solution

Isolation Forest[1]

Anomaly score, makes use of ensemble averages for trees to isolate each point, with:

$$s(x,n) \longrightarrow \begin{cases} 1.0 & \text{for } \overline{h(x)} \longrightarrow 0 \quad \text{(clear anomaly)} \\ 0.0 & \text{for } \overline{h(x)} \longrightarrow n-1 \quad \text{(normal)} \end{cases}$$



[1]Liu, F.T. et al, ICDM `08b, 2008.

Appropriate Features

When data are in the form of images, could work with raw pixels, *but ...*

- Identical images translated by a few pixels can be located far apart in p-dimensional feature space
- Should take advantage of excellent work done in image feature representation — deep neural nets (DNN)

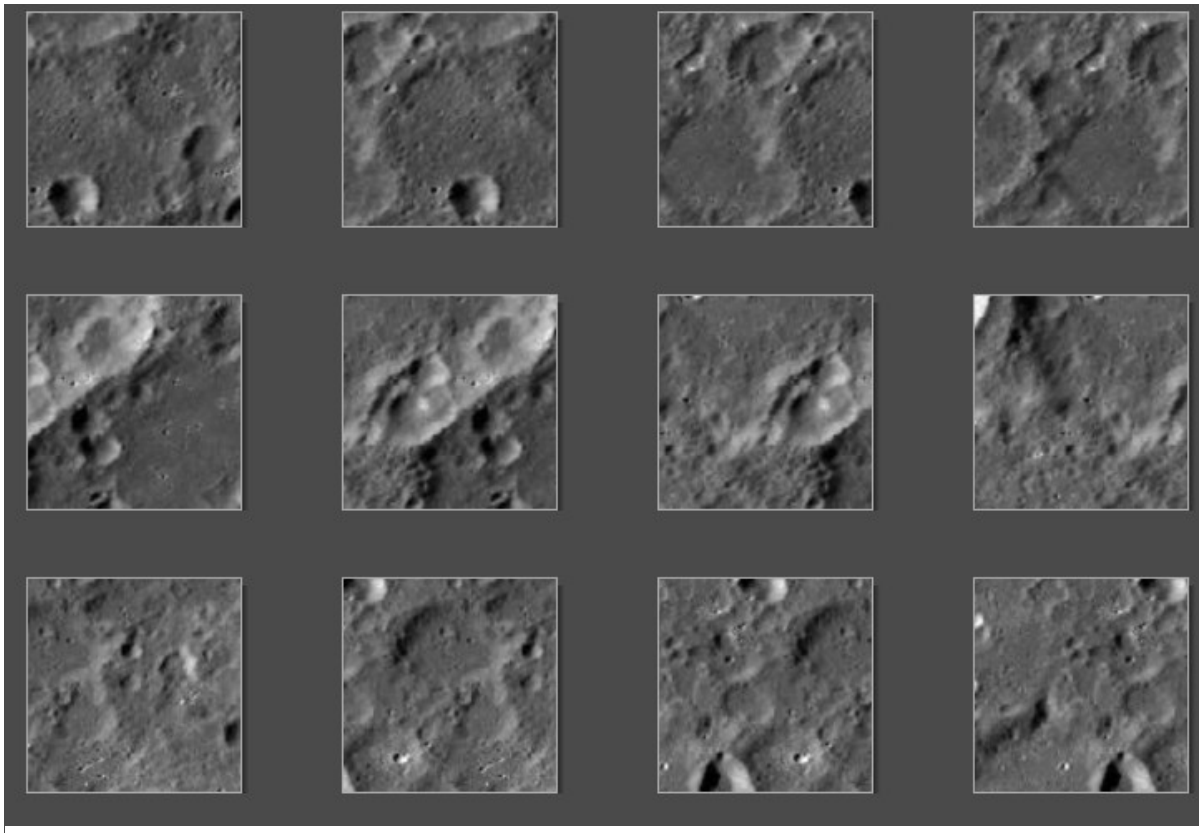
“Featurizing” Images

Introduce available pre-trained deep neural network[2]

- Re-size images, subtract training average, insert into DNN
- Extract (4096) high-level feature weights for each image
- Apply Isolation Forest to these features
- Determine anomaly score threshold:
  - That which matches some significance criterion
  - That which yields true positive rate that can afford to investigate in detail.

[2]Krizhevsky, A, et al, *Imagenet classification with deep convolutional neural networks,* 2012.

Results



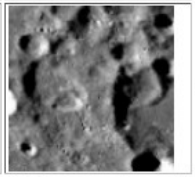
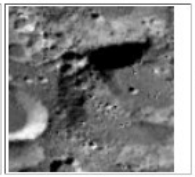
Starting with 2000+ photos of lunar landscapes, with a few sections from a photo of the Golden Gate Bridge.

Results

Out of 2000+ images, the anomaly scores for the Golden Gate Bridge are at the top, separated by all others by 10% or more.

Most Anomalous Files

So much fun! Start all over.

1		0.6237253736
2		0.603379225875
3		0.571449161944
4		0.558636179925
5		0.558527263