April 28, 2025

Dempsey Rogers

AI/ML Data Scientist

Hyperspectral Imaging (HSI) Anomaly Detection

What is the HSI Model?

The HSI model is a **statistical** approach to **anomaly** detection. HSI does not learn weights through training for later use during inference.

The model was selected for its use of edge and vertex weighted **graphs** to obtain relations between data points at different topological and **temporal scales** by evolving the affinity matrix.

HSI was adapted from the paper: *Graph Evolution-Based Vertex Extraction for Hyperspectral Anomaly Detection by Xianchang Yang et al.*

Model Structure

The HSI model is broken down into two main components, weights generation and loss function minimization.

Pixel Density:

$$\vec{\gamma} := \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \end{bmatrix}$$

Vertex Weights:

$$\mathbf{\Gamma} := \begin{bmatrix} \gamma_0 & & & & \\ & \gamma_1 & & & \\ & & \gamma_2 & & \\ & & & \ddots & \\ & & & & \gamma_n \end{bmatrix}$$

Data $\in \Re_{n,m}$

Separation:

$$\mathbf{D} = \sqrt{\sum_m (\vec{x_i} - \vec{x_j})^2}$$

Affinity Matrix:

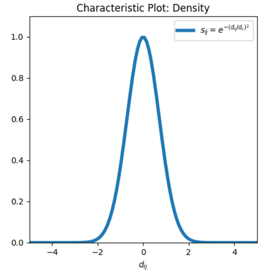
$$\mathbf{A} := \mathbf{\Gamma}\tilde{\mathbf{S}}\mathbf{\Gamma} = \begin{bmatrix} a_{0,0} & a_{0,1} & a_{0,2} & \cdots & a_{0,n} \\ a_{1,0} & a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,0} & a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \vdots & \ddots & \\ a_{n,0} & a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{bmatrix}$$

Similarity Matrix:

$$\tilde{\mathbf{S}} := \begin{bmatrix} s_{0,0} & s_{0,1} & s_{0,2} & \cdots & s_{0,n} \\ s_{1,0} & s_{1,1} & s_{1,2} & \cdots & s_{1,n} \\ s_{2,0} & s_{2,1} & s_{2,2} & \cdots & s_{2,n} \\ \vdots & \vdots & \vdots & \ddots & \\ s_{n,0} & s_{n,1} & s_{n,2} & \cdots & s_{n,n} \end{bmatrix}$$

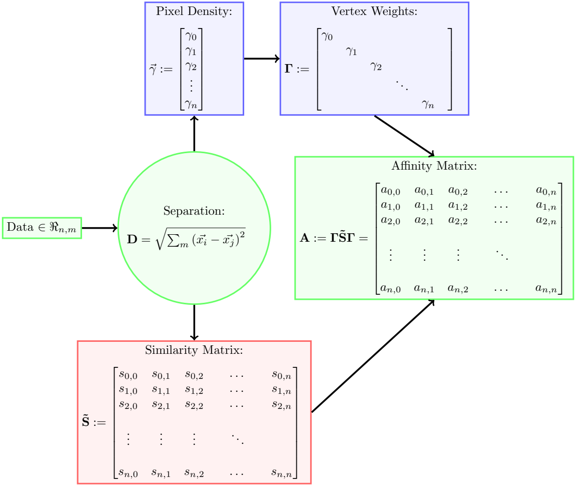
Density Calculations:

$$\gamma_i = \sum_{i \neq j} e^{-\left(d_{ij}/d_c\right)^2}$$



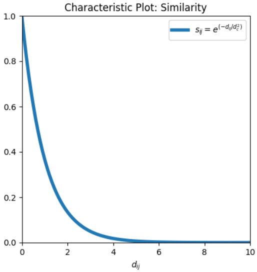
Model Structure

The HSI model is broken down into two main components, weights generation and loss function minimization.



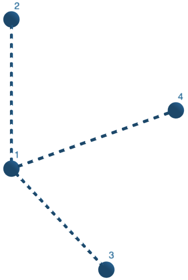
Similarity Calculations:

$$s_{ij} = e^{\left(-d_{ij}/d_c^2\right)}$$



Matrix Evolution (aside)

Adjacency matrices are used in graph theory to encode a graph's structure.

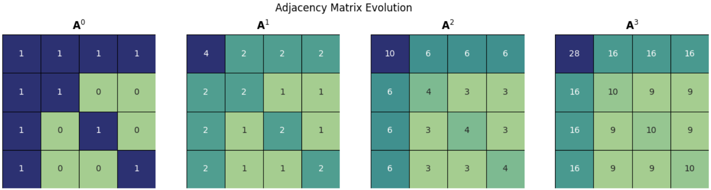


Adjacency $\Rightarrow$ Discrete

- Distance: Integer number of Hops
- Connectivity: Integer Number of Paths

Affinity $\Rightarrow$ $\sim$Continuous

- Distance: Function Space Measurement
- Connectivity: Encoded in $\mathbf{A}$ and $\vec{\gamma}$



Model Structure

The HSI model is broken down into two main components, weights generation and loss function minimization.

Affinity Matrix Set: $\mathbf{A^k}$

Distance Matrix Set: $\mathbf{D^k}$

Pixel Density: $\vec{\gamma}$

Penalized Objective Function:

$$\Phi(\vec{m}) :=$$

$$\vec{m}^T \mathbf{A}^k \vec{m} + \left(\gamma^T \mathbf{D}^k\right)^T \vec{m}$$

Such that: $0 \leq m_i \leq 1$

Optimize POF($\vec{m}$)

Assumption: Anomalous data has a fragile relationship with background data. This relationship will not minimize the POF as well as common points.

HSI Pipeline:

- A set of "evolved" matrices are kept $\vec{A}^k$
- Anomaly scores, $\vec{m}$, are used to minimize POF
- Minimized anomaly scores are used as starting point for POF at $k+1$
- Once anomaly score converge with in a tolerance, optimization ends
- $|Z_{score}|$ is calculated for anomaly scores

Data points with $|Z_{score}| >$ Anomaly Threshold value are predicted.

Model Inference

$\mathbf{n} \times \mathbf{n} \Rightarrow$ Computationally Expensive / Reduced *n* statistics

Multifilter: iterate on suspected points

- A data frame of anomalous predictions is made from all batches predictions
- Upon completion of the initial pass non-anomalous data points are randomly selected from entire data set.
- Anomalous data points from each batch are compared to each other, as well as, background points from the entire data set
- A count of times each point is predicted anomalous is kept as a batch score

Datapoints with batch score $<$ batch threshold are less likely to be false positives.

Model Parameter Summary

Parameters that affect the HSI Model's predictions:

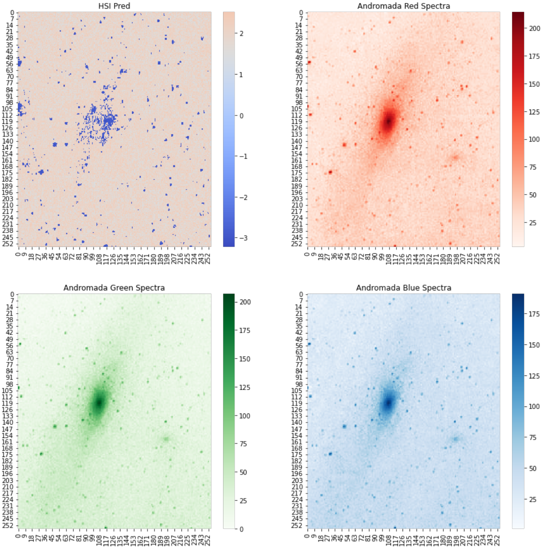
Parameter	Name	Effect
c_d	Cut-off Distance	Scales $\vec{\gamma}$ and **A**
$(P_r)^k$	Penalty Ratio $0 < P_r < 1$	Decays loss contributions at greater topological scales
l_r	Learning Rate	Scales steps along gradient decent of POF
batch_size$_d$	Batch Size	Specifies volume of data available for statistics.
a_t	Anomaly Threshold	Selects points for Multifilter based on $Z_{score}(\vec{m})$
m_t	Multifilter Threshold	Selects points for prediction based on multifilter count

Exmple Usage

Single pass of HSI:

$p \times q$ RGB channels ranging from 0-256:

- Ravel to $(p \times q, 3)$
- Scale
- HSI Inferences
- Heat map $\vec{m}$ as $p \times q$



Exmple Usage

Single pass of HSI: $p \times q$ Internet Protocol Data:

- Time series parsing of the Carbon-Black Logs
- Preprocessing: Encoding, Scaling, and PCA
- HSI Inferences
- Multi-filtered
- Multi Filter Count

Figure: Preprocessed input data relevant to Cyber Security.

Exmple Usage

Single pass of HSI: $p \times q$ Internet Protocol Data:

- Time series parsing of the Carbon-Black Logs
- Preprocessing: Encoding, Scaling, and PCA
- HSI Inferences
- Multi-filtered
- Multi Filter Count

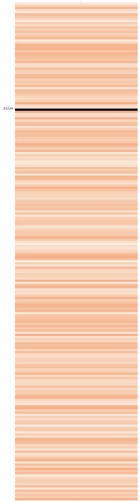


Figure: Multi-filter count after 15 filters. Max value of 13.

Exmple Usage

Multi-filtered HSI: $p \times q$ Internet Protocol Data:

- Time series parsing of the Carbon-Black Logs
- Preprocessing: Encoding, Scaling, and PCA
- HSI Inferences
- Multi-filtered
- Multi Filter Count

Figure: Anomalous data point predicted by the HSI model.

Closing

Observations:

- Generalizable to data types
- Effective in Cyber Tasks
- Computation time scales batch_size^2

Future Work:

- Test model on new data sources. . . Audio, .etc

Conclusion:

The HSI model has been implemented on image and cyber security related data. The model has been used to identify actionable malicious behaviors in IP data.

Questions?

Idaho National Laboratory