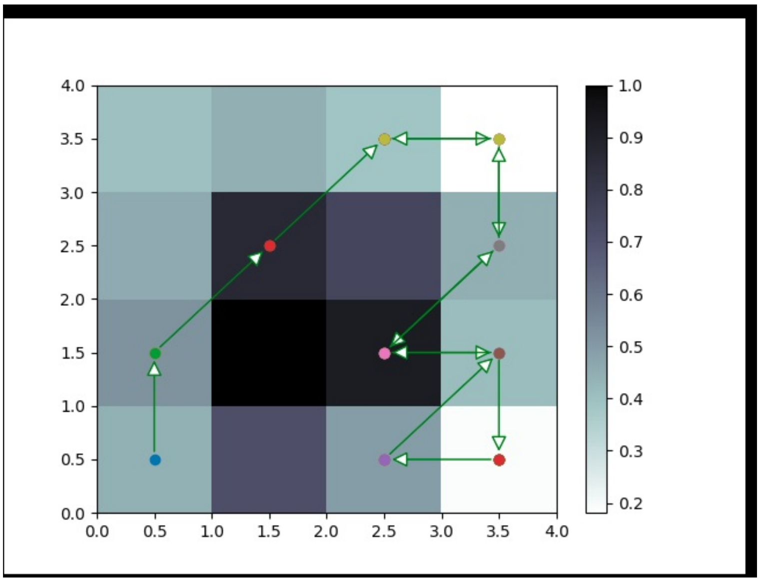
AE/VAE-SOM

Goal for the project:

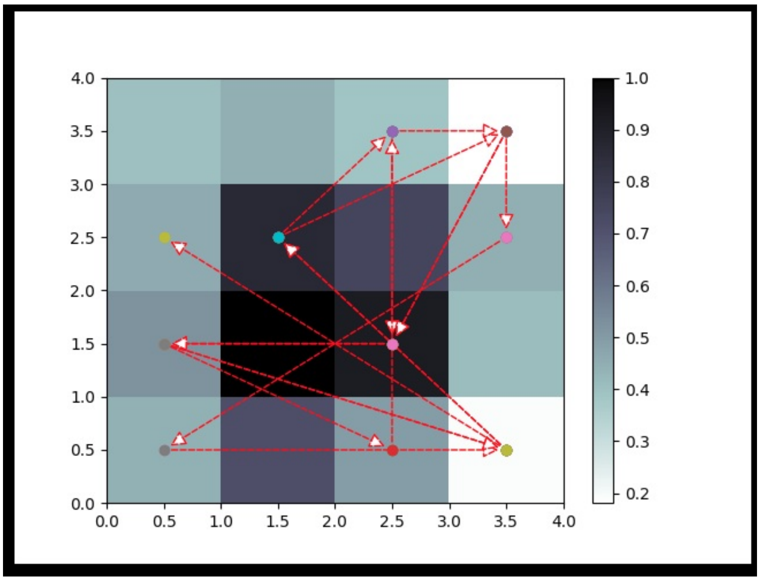
1. Latent space exploration of SOM.
2. Representation learning of AE/VAE/GAN.
3. Anomaly detection for time-serials data.
   - 1) Spatial anomaly detection.
   - 2) Temporal anomaly detection.
   - 3) Generalization and robustness of the Algorithms.
4. Quantization of uncertainty.
   - 1) Spatial uncertainty.
   - 2) Temporal uncertainty.

Updated May 27th:

- Common Trajectory:



- Faulty Trajectory:



To do list after May 27th update:

1. Investigate correlation between the common input vs. faulty input
2. How to process the biodirectional arrow in the Common trajectroy?
3. ~~How to down sample the~~ z_e ?
4. ~~Markov Model~~

Updated June 4th:

- Fixed up the Markov Chain. Need to be validated.
  - *Note: The Markov Chain is not the only method to optimize the temporal model.*
- Set up a new block for spatial anomaly detection.
  - $P_{anomaly}(v) = 1 - exp(-(error_q v/\sigma))(1)$.
    v refers to input data.
    σ refers to $\frac{1}{2} \cdot ||w_{bmu}(v) - w'_{bmu}(v)||$
    $error_q v$ refers to $||v - bmu(u_i)||$
- ~~σ means the divergence of the data. sigma is larger, the results are more dispersed.~~ σ ~~should be around~~ $0.5 \to 1.5$. $Learning\ rate$ ~~should be around~~ $0.5 \to 2.5$. ~~The map size should not more than~~ $(10, 10)$.
- The detail about hyper parameters can be found here [A simple demo for Hyper parameters]

- An example:

- Loss

To do list after June 4th's update

1. ~~Tidy the source code.~~
2. Change the dataset.
   - ~~*Socketsense*~~
   - *Itrust*
3. ~~Add a new block for anomaly detecion in Spatial space. It needs to test more to avoid FP.~~
4. ~~The hyper-parameters needs to be clarify in order to get a better results in SOM.~~
5. Temporal anomaly detection.
   - Investigation in Markov Chain
   - ~~Investigation in Affinity Matrix.(Continue to abstract for the SOM neuron)~~.
6. Confirm the Spatial anomaly detection and V&V results.

Updated June 5th

1. The mapsize should be calibrated with the σ. In other words, when map size is larger, the sigma should be larger. **Question: How we should interpret these two hyper parameters?**
2. The Sin function has been replaced by SocketSense.
3. Function update:
   - Add a new block to check the FP.
   - Add a new block to generate the affinity(similarity) matrx, which could be treated as *Truth table*.

To do list after June 5th's update

1. ~~Normaliztion the Affinity matrix, which should either indicate the probability or similarity.~~
2. Continue to work on the termporal anomaly detection.

Updated June 8th

1. Get the Affinity probability of each *Path state.*
2. Visualization each Affinity Probability according to
   - $P_{affinity} = exp(-distance(w_i, w_j)/\delta)(2)$
     w_i refers to the path state, w_j refer to the other state
     I set a threshold δ, to select the affinity. When $P_{affinity} > \delta$, we treat w_j as *affinity state.*
3. Map the input to SOM.
   - For example, in our case, we have four pressures, these four pressure's sum will map with the Som. When the Pressure is larger, the color is deeper.

4. Overview results.
   - Temporal path-state trajectory.

   - Visualization of affinity and path states.

To do list after June 8th's update

1. ~~Reschematic Markov Chain.~~

   - ~~The Markov Chain should save as a table (dataframe) such as following:~~

	S1	S2	S3	....
S1	$\times$	P_{12}	P_{13}	....
S2	P_{21}	$\times$	P_{23}	....
S3	P_{31}	P_{32}	$\times$	....

2. Add text (probability) in the temporal trajectory figures.

3. Write a Problog program for the figures.

Discussion

1. How can we prove that *Formula (2)* is reliable?
   - Usually, we should use $(w_i - w_j)^2$ to compute the affinity matrix. But in this case, if we use square, the amount of affinity state will increase a lot. Is *Formula (2)* reasonable ?
   - How should we understand δ?
2. How can we improve/calibrate the method used to map the input data to SOM?
3. Validate our data.....(This may meet a lot of problems)

Updated June 12h

1. The sampling problem can be checked from the following link (the classical methods):
   - [Engineering statistics handbook.]()
2. Fix up the datatype of Markov Transition Matrix.
3. Fix up some bugs.

Discussion

1. How can we understand **Average Sample Number** (ASN)?
   How we understand the $\beta = 0.029$ and which value should be p_2 in [Double Sampling Plan]()?
2. In the [Double Sampling Plan](), they should define the P_1 .
   For example, As I understand, $P_1 = 1 - \alpha + \beta$, although α and β belong to different threshold $< n_1, c_1 >$ and $< n_2, c_2 >$.
3. Should Markov Chain have a sample step(frequency) as sampling data? Otherwise, the tranistion probability would equal to 0 if we use the upsampling states, which shows as below:

Updated June 15h

1. Add a flowgraph to illustrate the workbench.

Note: the sampling problem has not added into the flowgraph. It is a critical problem we should consider now.

Discussion

1. ~~How should we sample the data?~~
2. ~~How can we merge the slide windows in our code~~ ?

Updated June 22rd

1. Fulfill the simlarity and temporal graph to show the dependency.
   !
2. Some summary for the potential publication.
   - 1) What's the purpose of SOM used in latent space?

     SOM can be understood as the measurement of latent space's Euclidean distance.
   - 2) Spatial modeling

$$
\begin{gathered}
ELBO: \mathbb{E}_{PD(x)}[\log p_\theta(x)] \geq \mathbb{E}_{PD(x)}[\mathbb{E}_{q\phi(z|x)}[\log p_\theta(x|z)] - \mathbb{KL}(q_\phi(z|x)||p(z))] \\
Threshold: H_0 \ and \ H_1 \\
Similarity: g(z_i, z_j) = exp(-distance(w_i, w_j)/\delta \\
Spatialanaomly: g(z_i, z_j) \ compare \ to \ H_0 \ and \ H_1 \\
VAE: \mathcal{L}_{vae} = \mathbb{E}_{PD(x)}[\mathbb{E}_{q\phi(z|x)}[\log p_\theta(x|z)] - \mathbb{KL}(q_\phi(z|x)||p(z))] \\
Loss \ for \ the \ observation \ and \ latent \ space: \mathcal{L}_{vae-som}(\phi, \theta, x_i, x_j) = \mathcal{L}_{vae}(\phi, \theta, x_i, x_j) + g(z_i, z_j)
\end{gathered}
$$

   - 3) Temporal modeling

     How we get the threshold of temporal modeling?
3. Add a function for sliding window.