Theory and Applications of Topological Representations in Hierarchical Neural Networks

Pitoyo Hartono

School of Engineering, Chukyo University,
Nagoya, Japan

Institute for Human Robot Co-Creation, Waseda University,
Tokyo, Japan

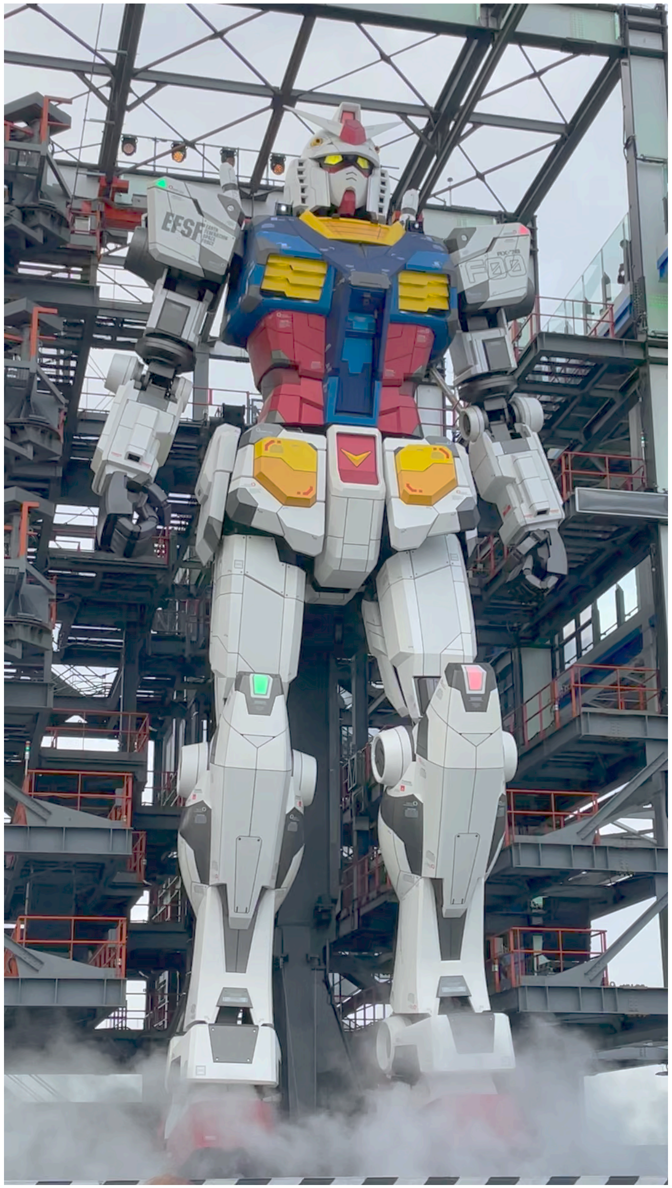
Where do I work ?

CHUKYO UNIVERSITY

Gundam Global Challenge

CHUKYO UNIVERSITY

Gundam Global Challenge

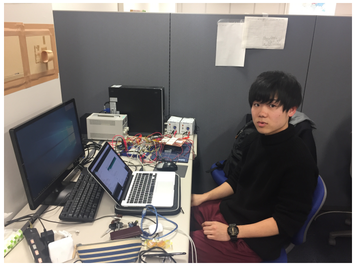


GUNDAM
FACTORY
YOKOHAMA

CHUKYO UNIVERSITY

Smart Interface

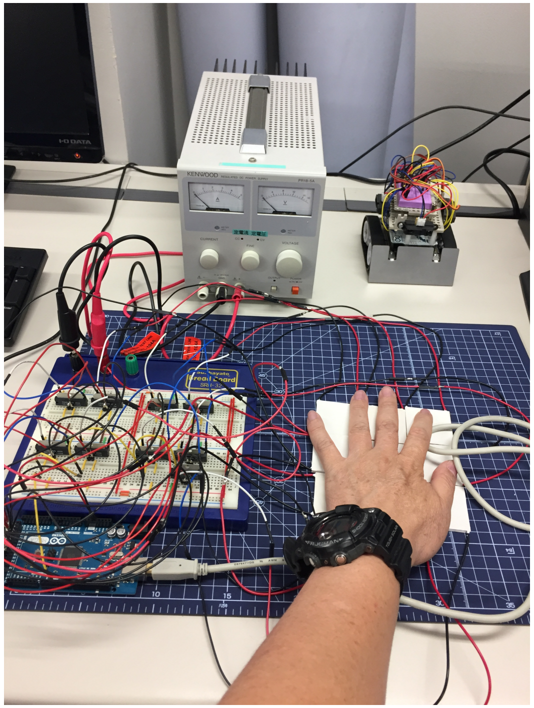
Pain illusion

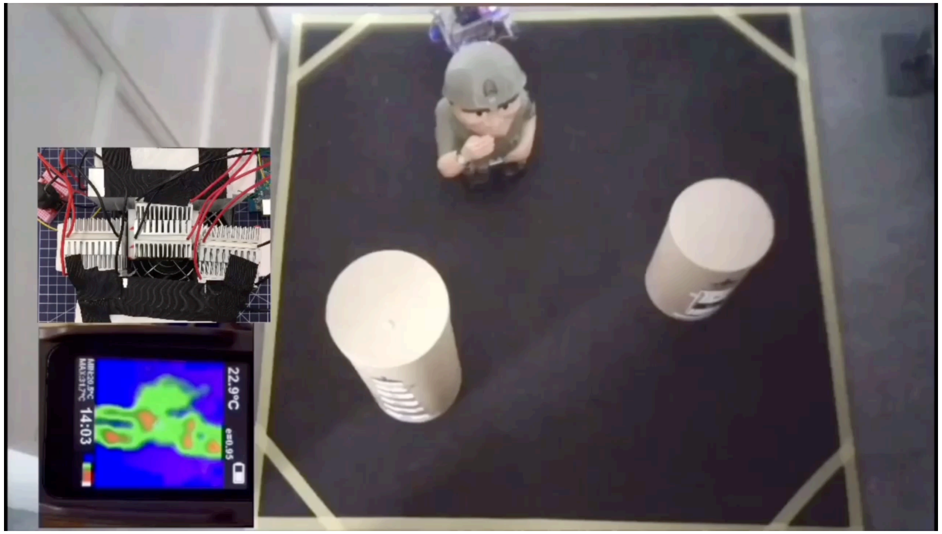


Shota Kato

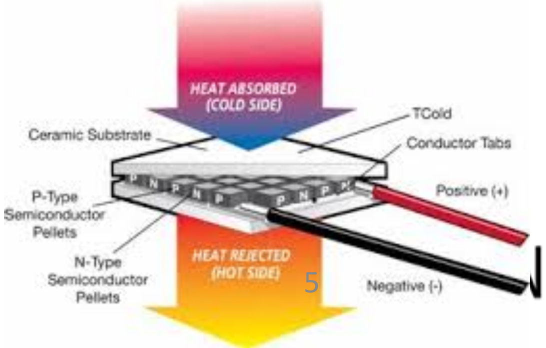
Hiroki Kishi

Thermal Grill Illusion



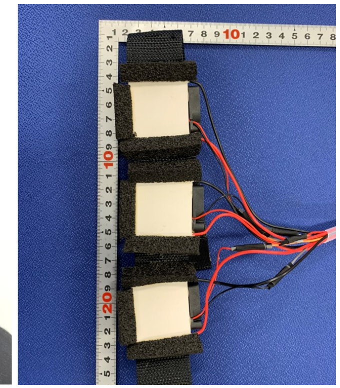


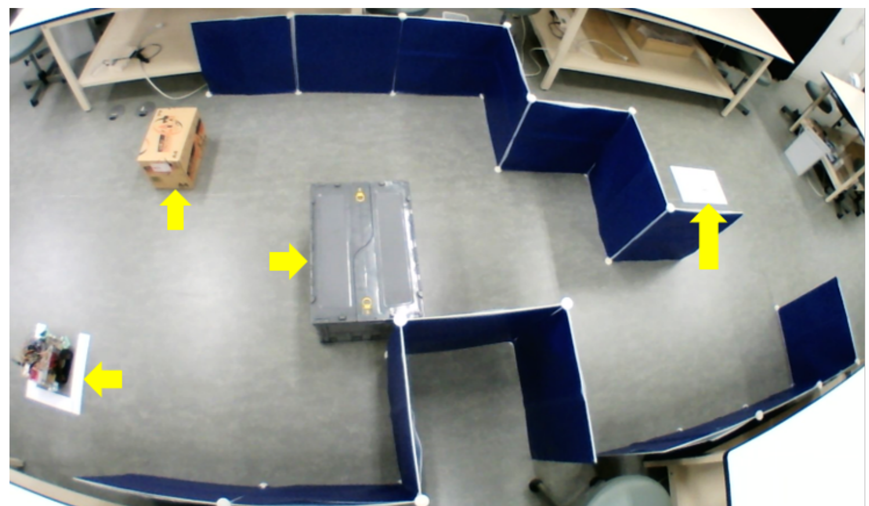
Peltier Device



Smart Interface

Pain illusion

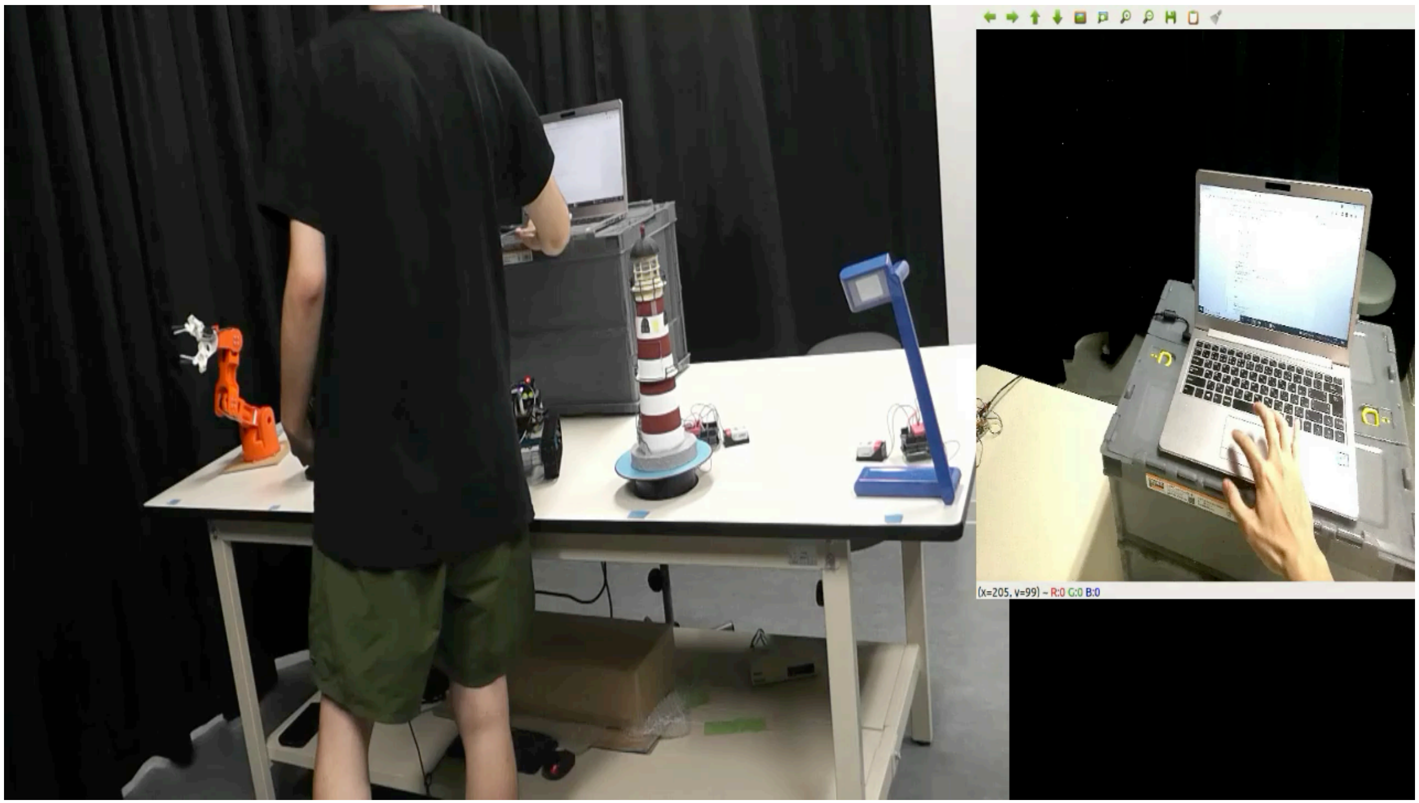




CHUKYO UNIVERSITY

Smart Interface

Gaze Interface



Yoshikazu Murase

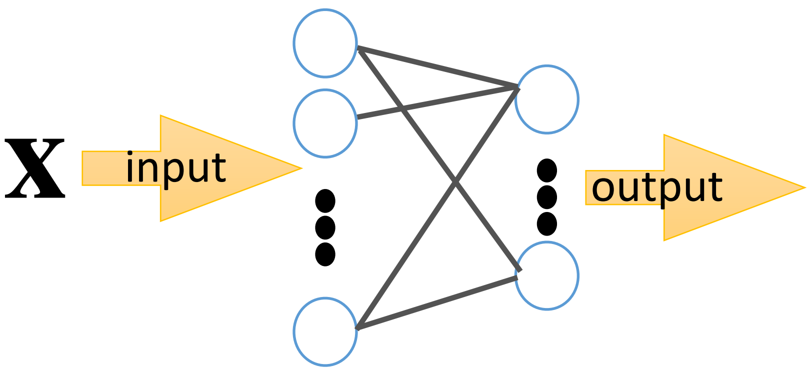
Shusuke Kobayashi

CHUKYO UNIVERSITY

Internal Representations of Neural Nets

Linear Problem

$$Y = f(X)$$



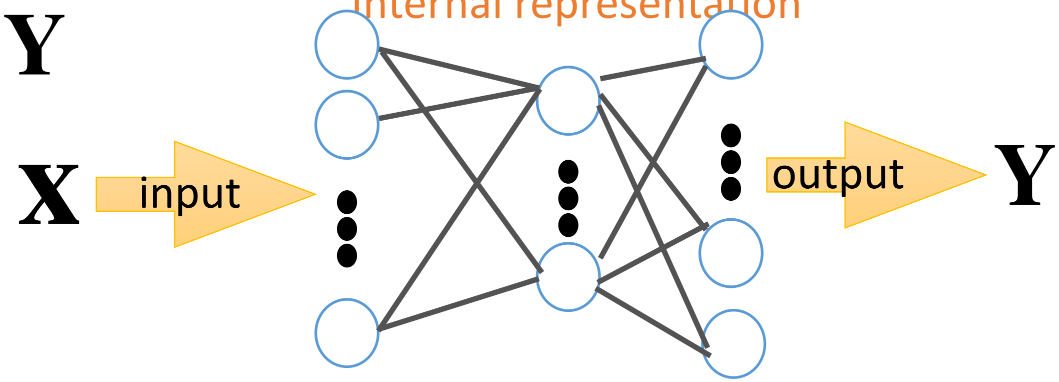
Perceptron

(Frank Ronsenblatt, 1957)

Nonlinear Problem

$$Y = f(g(X))$$

Internal representation



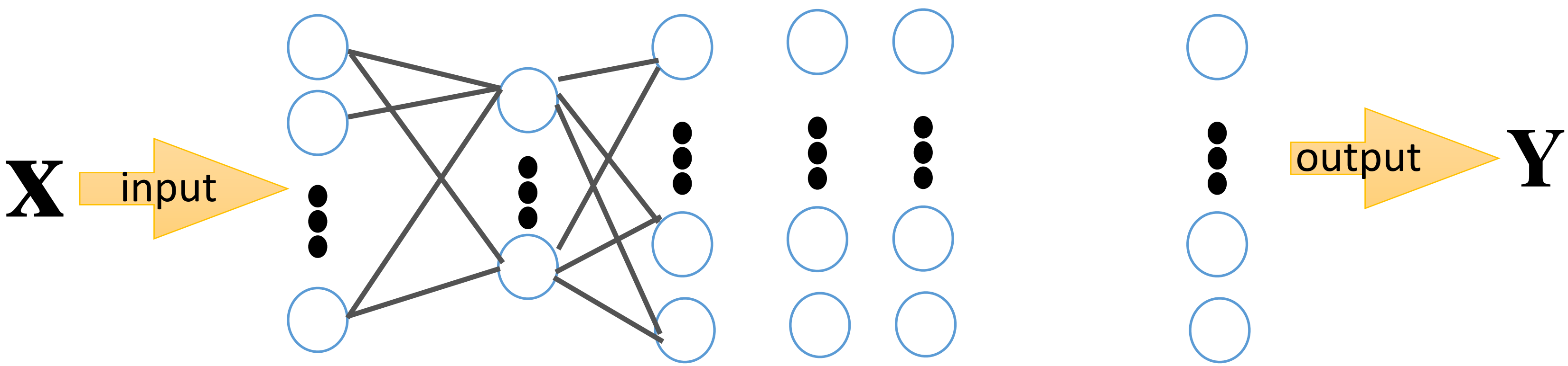
Multilayered Perceptron (MLP)

(Rumelhart, Hinton, Williams, 1986)

Internal Representations of Neural Nets

(Difficult) Nonlinear Problem

$$\mathbf{Y} = \mathbf{f}(\mathbf{g}_1(\mathbf{g}_2(\cdots\mathbf{g}_n(\mathbf{x}))\cdots)$$



Deep Neural Network (DNN)

Low abstraction level High

Physical features

dog

Concept

CHUKYO UNIVERSITY

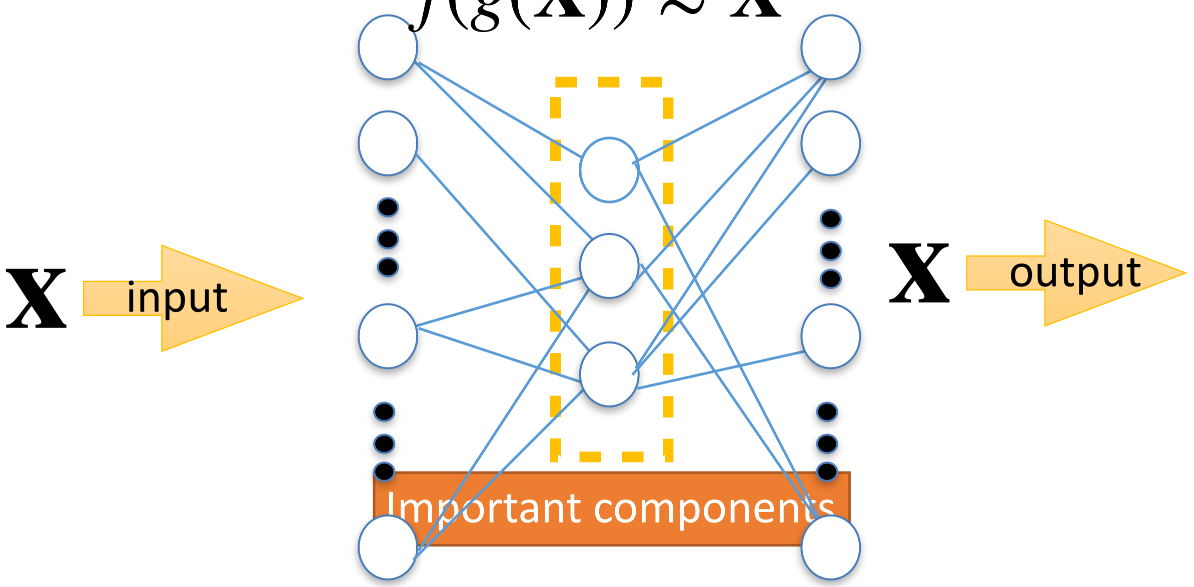
Internal Representations of Neural Nets

Generating Representation

Stacked Autoencoders

Autoencoder

$$f(g(\mathbf{X})) \approx \mathbf{X}$$



Principal Components Analysis (PCA)

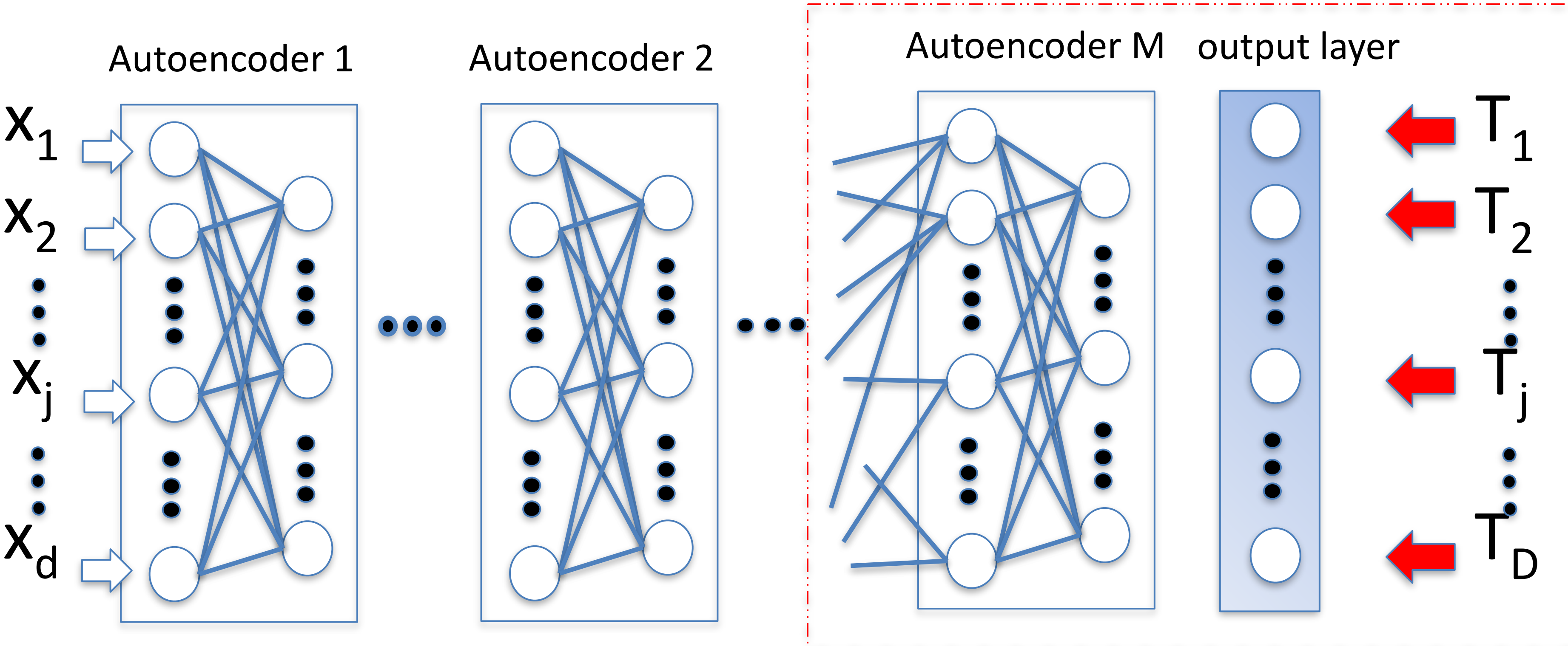
Nonlinear PCA

CHUKYO UNIVERSITY

Internal Representations of Neural Nets

Generating Representation

Stacked Autoencoders



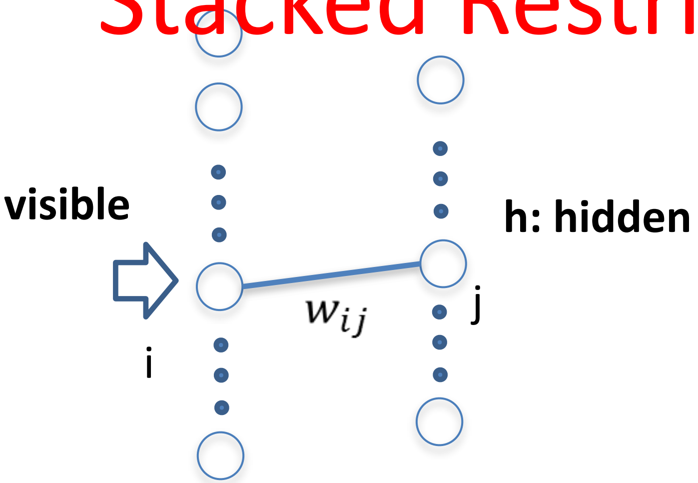
Backprop !

Avoiding Vanishing Gradient !

Internal Representations of Neural Nets

Generating Representation

Stacked Restricted Boltzmann Machines



contrastive divergence

$$\Delta w_{ij} = \eta(< v_i h_j >_{data} - < v_i h_j >_{model})$$

A kind of Factor Analysis

$$p(h_j = 1|v) = f(b_j + \sum_i v_i w_{ij}) \quad p(v_i = 1|h) = f(a_i + \sum_j h_j w_{ij})$$

$$E(v,h) = -\sum a_i v_i - \sum b_j h_j - \sum v_i h_j w_{ij}$$

$$p(v,h) = \frac{1}{Z} e^{-E(v,h)}$$

$p(v) = \sum_h p(v,h)$ **Maximize!**

Internal Representations of Neural Nets

So far so good! So what is the problem?

International Conference on Learning Representation (ICLR)

Comprehensibility!

What representations are good? What are bad?

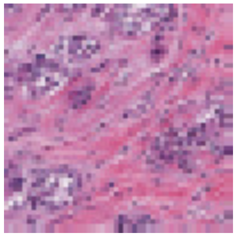
How can human understand neural networks?

How can knowledge be transfered?

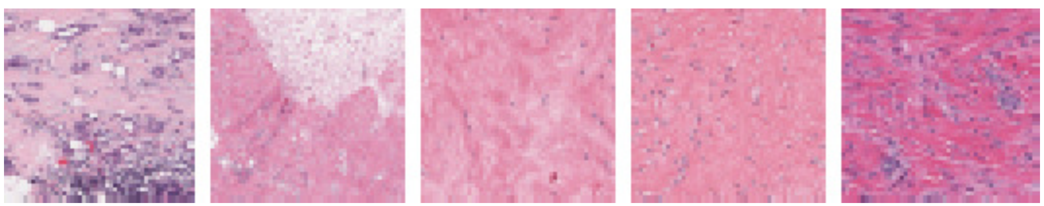
Comprehensibility

Explainable AI for histopathological diagnosis

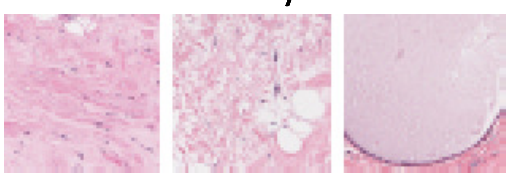
your sample is Invasive Ductal Carcinoma (IDC)



”Your image is IDC. However, there is high possibility that it could be misclassified as non-IDC, because it has medium similarity with some clusters of the non-IDC samples.”

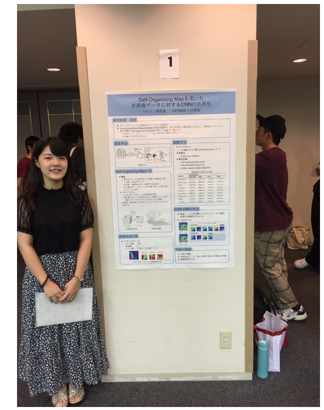


low-similarity non-IDC



non-similar non-IDC

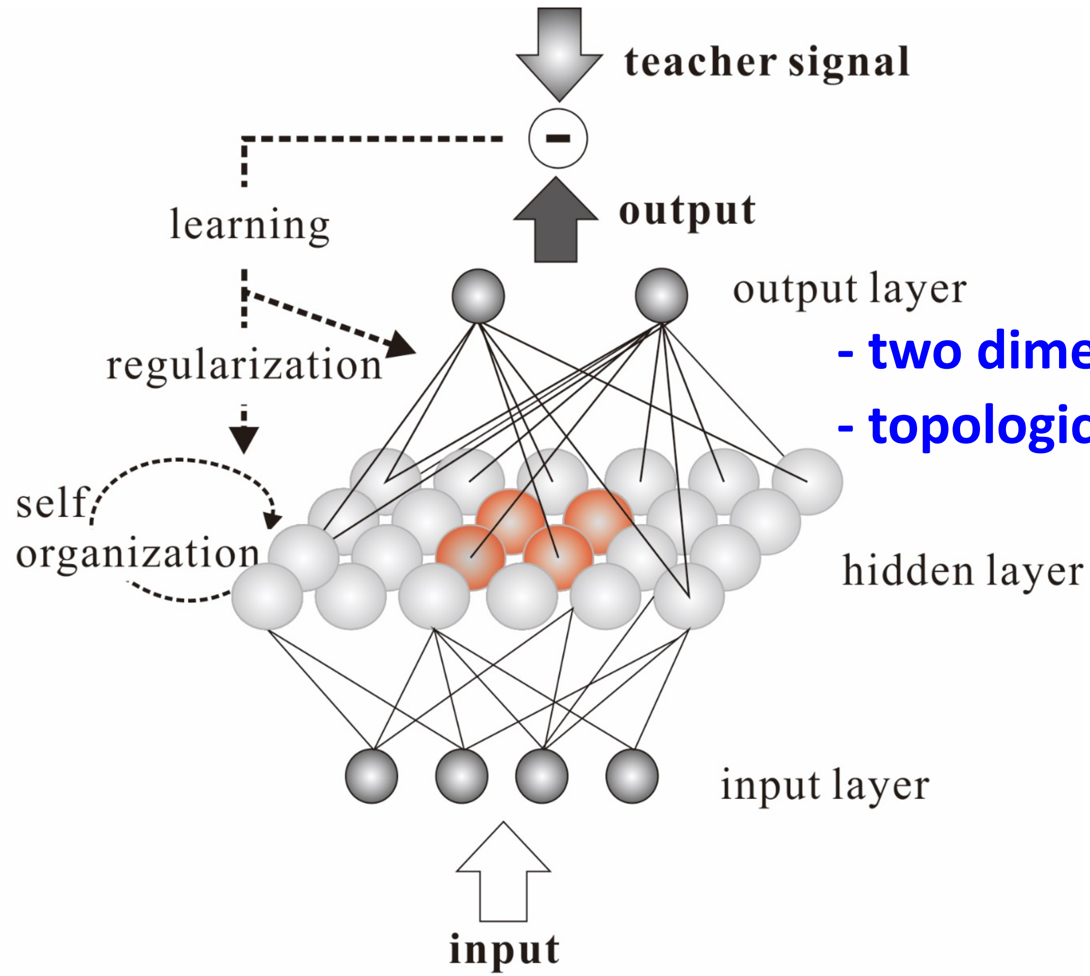
Patrik Sabol
TU Kosice, Slovakia



Kana Ogawa

P. Sabol, P. Sincak, P. Hartono, et al. , Explainable Classifier Supporting Decision-making for Colorectal Cancer Diagnosis from Histopathological Images, Journal of Biomedical Informatics, Vol. 109, 103523 (2020)

Restricted Radial Basis Function Net (rRBF)

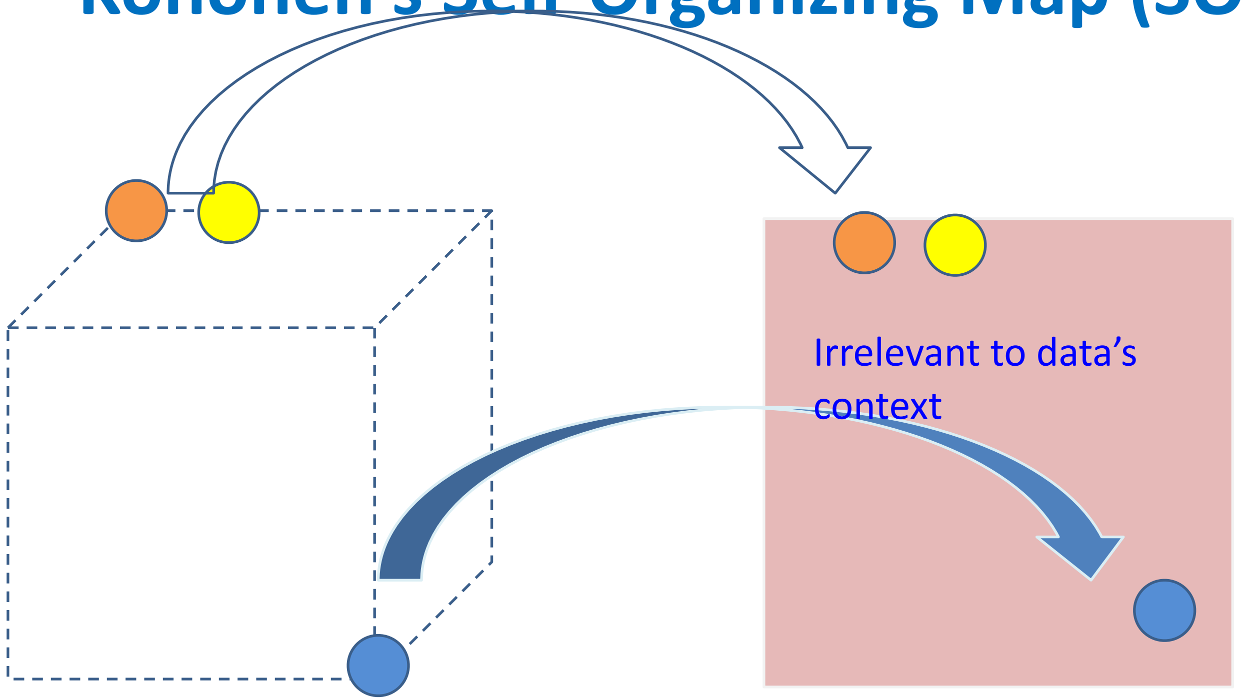


- two dimensional hidden layer
- topologically restricted activation function

P. Hartono, T. Trappenberg, Learning-Regulated Context Relevant Topographical Maps, IEEE Trans. on NNLS, Vol.26, No. 10, pp. 2323-2335 (2015).

Internal Representations of Neural Nets

Kohonen's Self-Organizing Map (SOM)



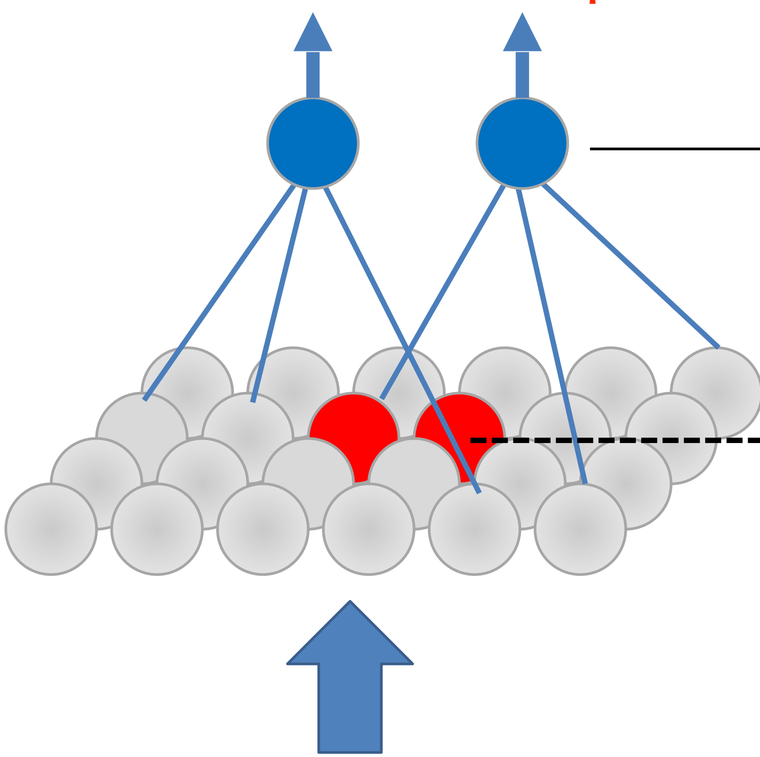
Original Dimension Low Dimension

- preserving the topological order of the data
- context-irrelevant

Kohonen, T. (1982). Self-organized Formation of Topologically Correct Feature Maps, Biological Cybernetics Vol. 43, 59-69.

Restricted Radial Basis Function Net (rRBF)

output



input X(t)

$$p(class = i|X(t)) = \frac{e^{I_i^{out}}}{\sum_l e^{I_l^{out}}}$$

probability density function of predicted output

$$I_i^{out}(t) = \mathbf{V}_i^T(t)\mathbf{H}(t)$$

potential of i-th output neuron

$$H_j(t) = e^{-I_j(t)}\, \sigma(\mathrm{win}, \mathrm{j}, \mathrm{t})$$

output of *j*-th hidden neuron

$$I_j(t) = \frac{1}{2}\left\| X(t) - W_j(t) \right\|^2$$

$$win = \arg\min_j I_j(\mathrm{t})$$

Restricted Radial Basis Function Net (rRBF)

Loss Function: Cross Entropy

$$E(t) = -\log p(class = K | X(t))$$ (correct class =K)

Weight Modification

$$\mathbf{V}_i(t+1) = \mathbf{V}_i(t) - \eta \frac{\partial E(t)}{\partial \mathbf{V}_i(t)}$$

Hidden-Output Weight Modification

$$\mathbf{W}_j(t+1) = \mathbf{W}_j(t) - \eta \frac{\partial E(t)}{\partial \mathbf{W}_j(t)}$$

Reference Vector Modification

Restricted Radial Basis Function Net (rRBF)

Learning Process

$$\frac{\partial E(t)}{\partial \mathbf{W}_j(t)} = -\delta_j(t)\sigma(win, j, t)(\mathbf{X}(\mathbf{t}) - \mathbf{W}_j(t))$$

$$\delta_j(t) = (V_{Kj} - \tilde{V}_j)e^{I_j(t)}$$

Regulation signal from the higher layer

$$\mathbf{W}_j(t+1) = \mathbf{W}_j(t) + \delta_j(t)\sigma(win, j, t)(\mathbf{X}(t) - \mathbf{W}_j(t))$$

Restricted Radial Basis Function Net (rRBF)

Learning Process

$$\mathbf{W}_j(t+1) = \mathbf{W}_j(t) + \delta_j(t)\sigma(win, j, t)(\mathbf{X}(t) - \mathbf{W}_j(t))$$

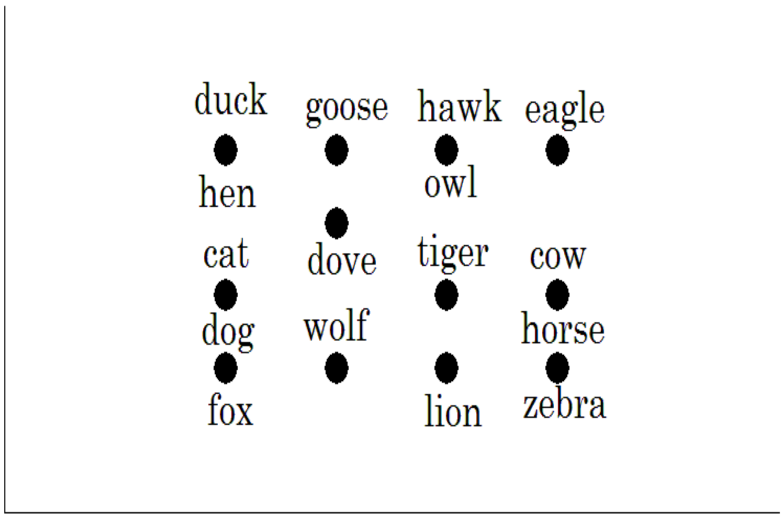
$\delta_i > 0$: SOM Learning

i-th hidden neuron decreases the learning error

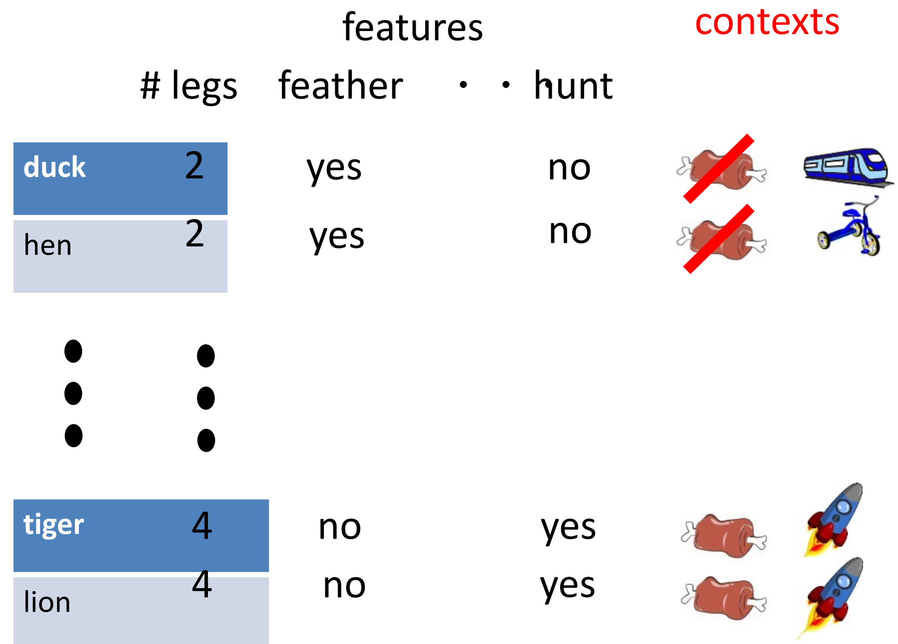
$\delta_i < 0$: repulsive SOM

i-th hidden neuron increases the learning error

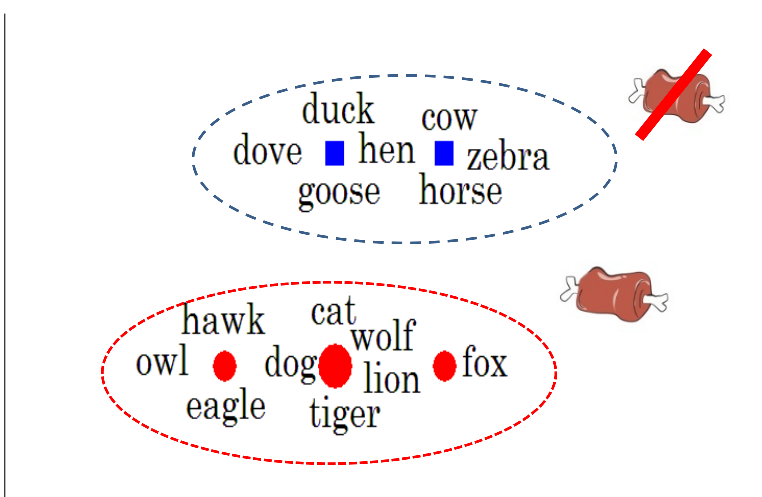
Experiment



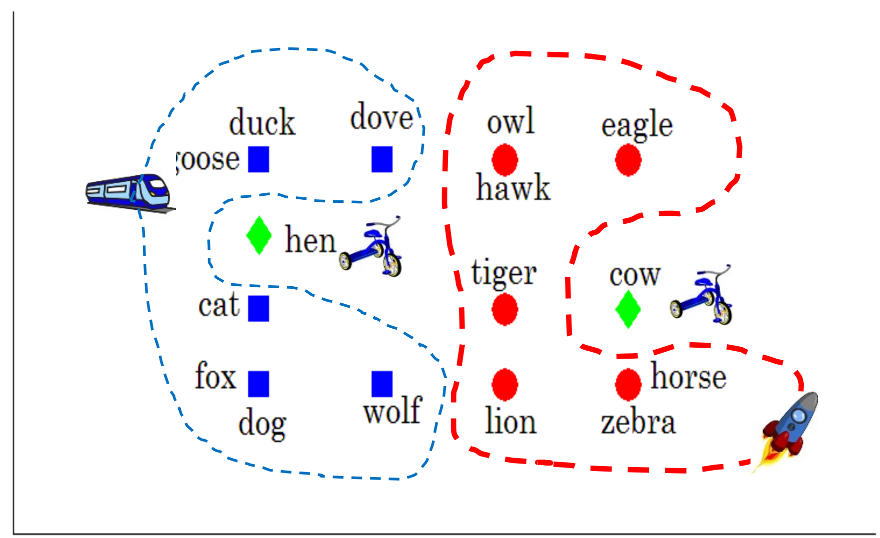
Simple SOM



Ritter, H., and Kohonen, T. (1989). Self-Organizing Semantic Maps. Biological Cybernetics, 61, 241-254.



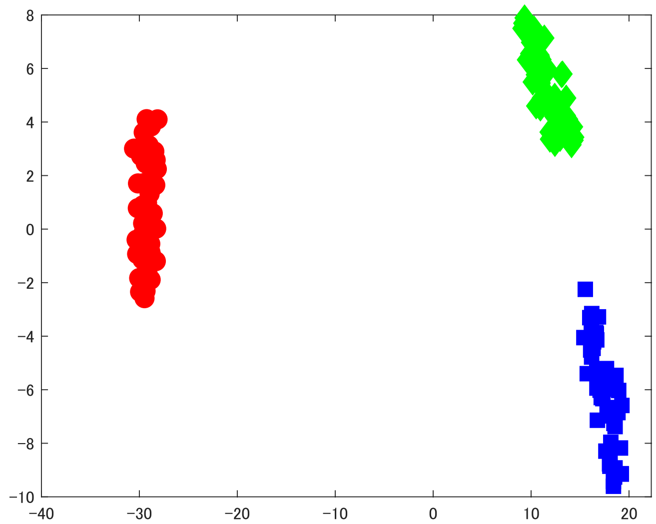
rRBF (herbivore/carnivore)



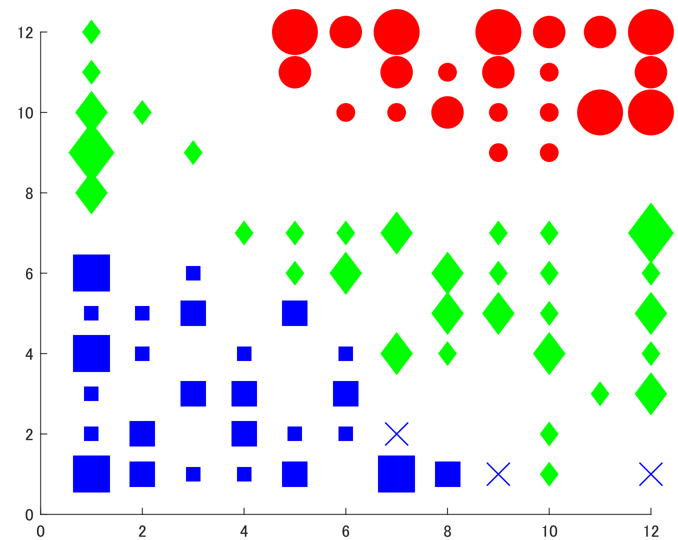
rRBF (speed)

Restricted Radial Basis Function Net (rRBF)

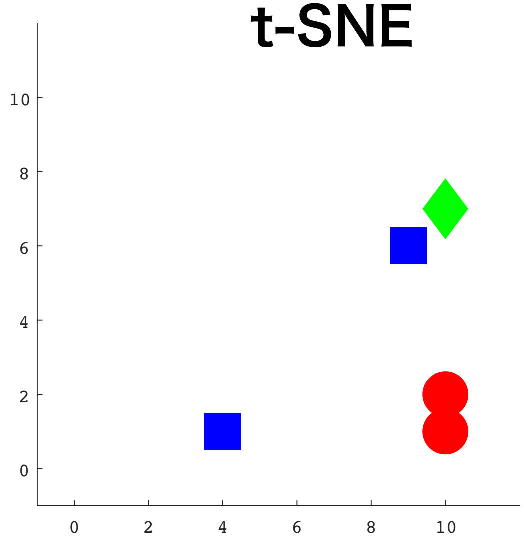
Experiment **Iris Data (dim:4, class:3)**



t-SNE



SOM

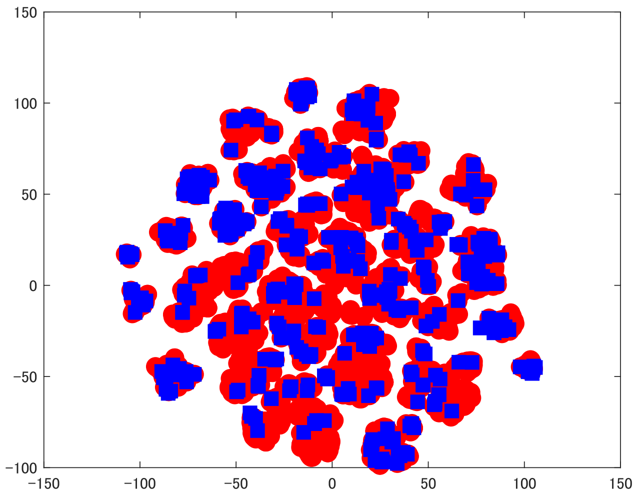


rRBF

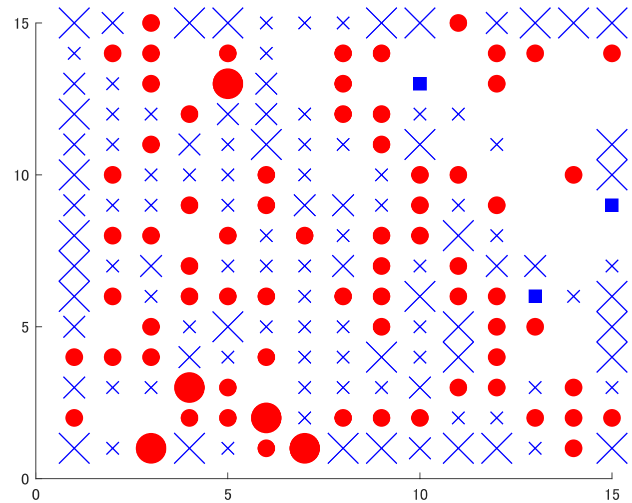
CHUKYO UNIVERSITY

Restricted Radial Basis Function Net (rRBF)

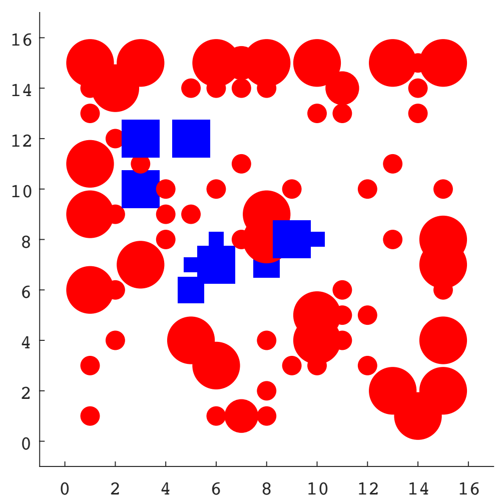
Experiment Bank Credit Data (dim:48, class:2)



t-SNE



SOM

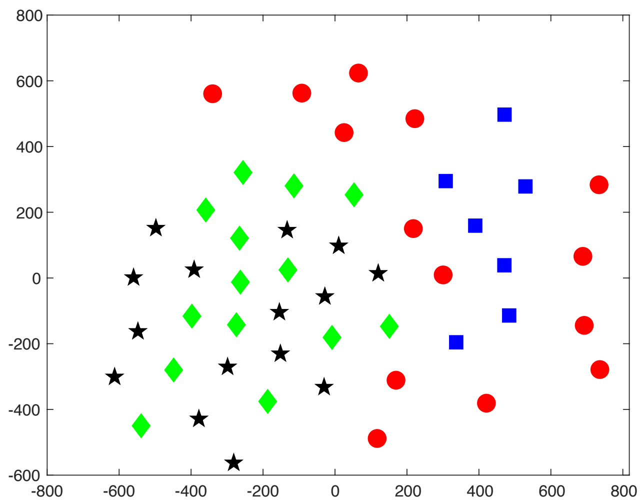


rRBF

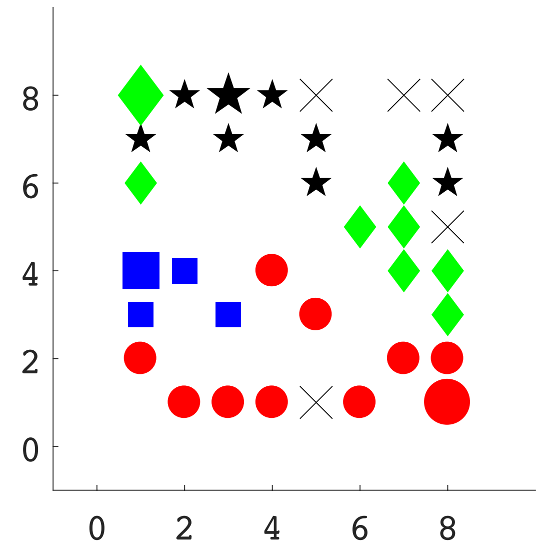
CHUKYO UNIVERSITY

Restricted Radial Basis Function Net (rRBF)

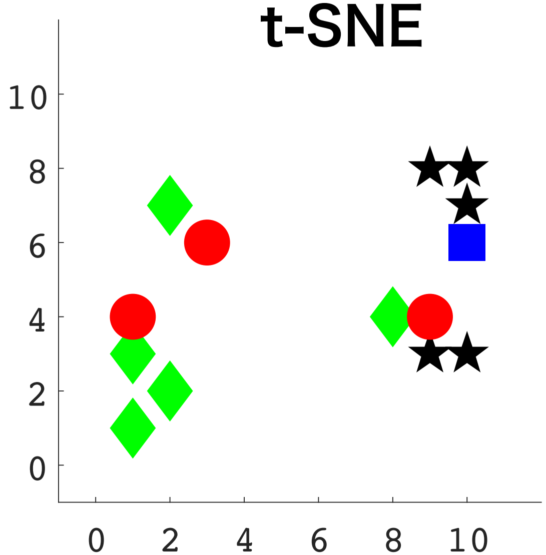
Experiment **Brain Tumor Data (dim:10367, class:4)**



t-SNE



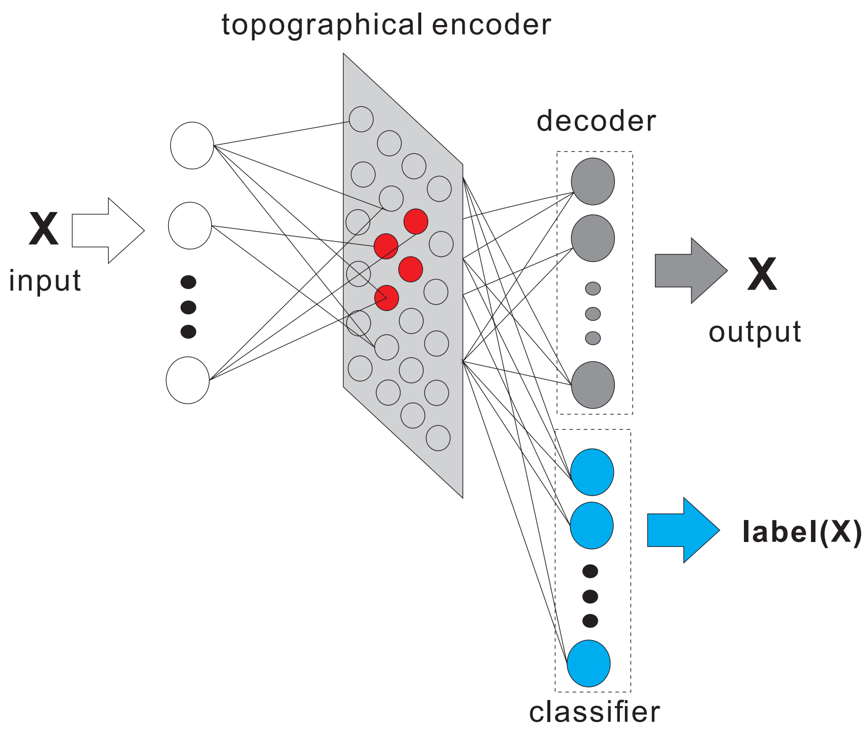
SOM



rRBF

CHUKYO UNIVERSITY

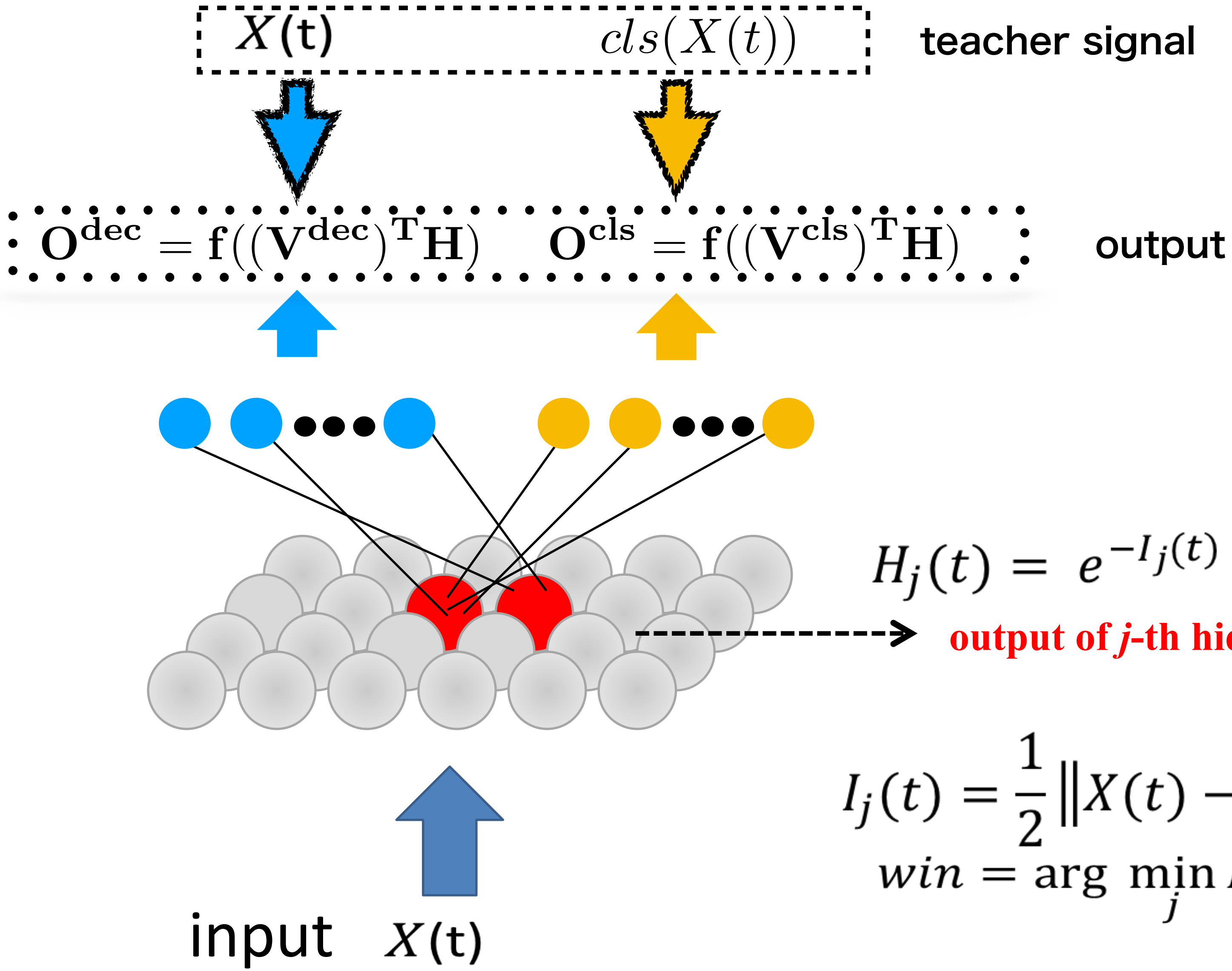
Soft-supervised topological autoencoder(STA)



- two dimensional hidden layer
- topologically restricted activation function

P. Hartono, Mixing autoencoder with classifier: conceptual data visualization, IEEE Access, Vol. 8, pp.105301 -105310 (2020) DOI: 10.1109/ACCESS.2020.2999155

Soft-supervised topological autoencoder(STA)



$$H_j(t) = e^{-I_j(t)}\,\sigma(\mathrm{win}, \mathrm{j}, \mathrm{t})$$

output of *j*-th hidden neuron

$$I_j(t) = \frac{1}{2}\left\|X(t) - W_j(t)\right\|^2$$

$$win = \arg\min_j I_j(\mathrm{t})$$

Soft-supervised topological autoencoder(STA)

Loss Function:

reconstruction error — classification error

$$L = \frac{(1-\kappa)}{2d}\sum_{k}(O_k^{dec} - X_k)^2 + \frac{\kappa}{2d_{cls}}\sum_{l}(O_l^{cls} - T_l)^2$$

Weight Modification

$$\Delta \mathbf{V}_k^{dec} = \frac{\partial L}{\partial \mathbf{V}_k^{dec}}$$

$$= \frac{(1-\kappa)}{d}(O_k^{dec} - X_k)O_k^{dec}(1 - O_k^{dec})\mathbf{H}$$

$$\Delta \mathbf{V}_l^{cls} = \frac{\partial L}{\partial \mathbf{V}_l^{cls}}$$

$$= \frac{\kappa}{d_{cls}}(O_l^{cls} - T_l)O_l^{cls}(1 - O_l^{cls})$$

CHUKYO UNIVERSITY

Soft-supervised topological autoencoder(STA)

Reference Vectors Modification

$$\frac{\partial L}{\partial \mathbf{W}_j} = \frac{\partial L}{\partial O_k^{dec}} \frac{\partial O_k^{dec}}{\partial \mathbf{W}_j} + \frac{\partial L}{\partial O_k^{cls}} \frac{\partial O_k^{cls}}{\partial \mathbf{W}_j}$$

$$= \delta_j^{hid} H_j (\mathbf{X} - \mathbf{W}_j)$$

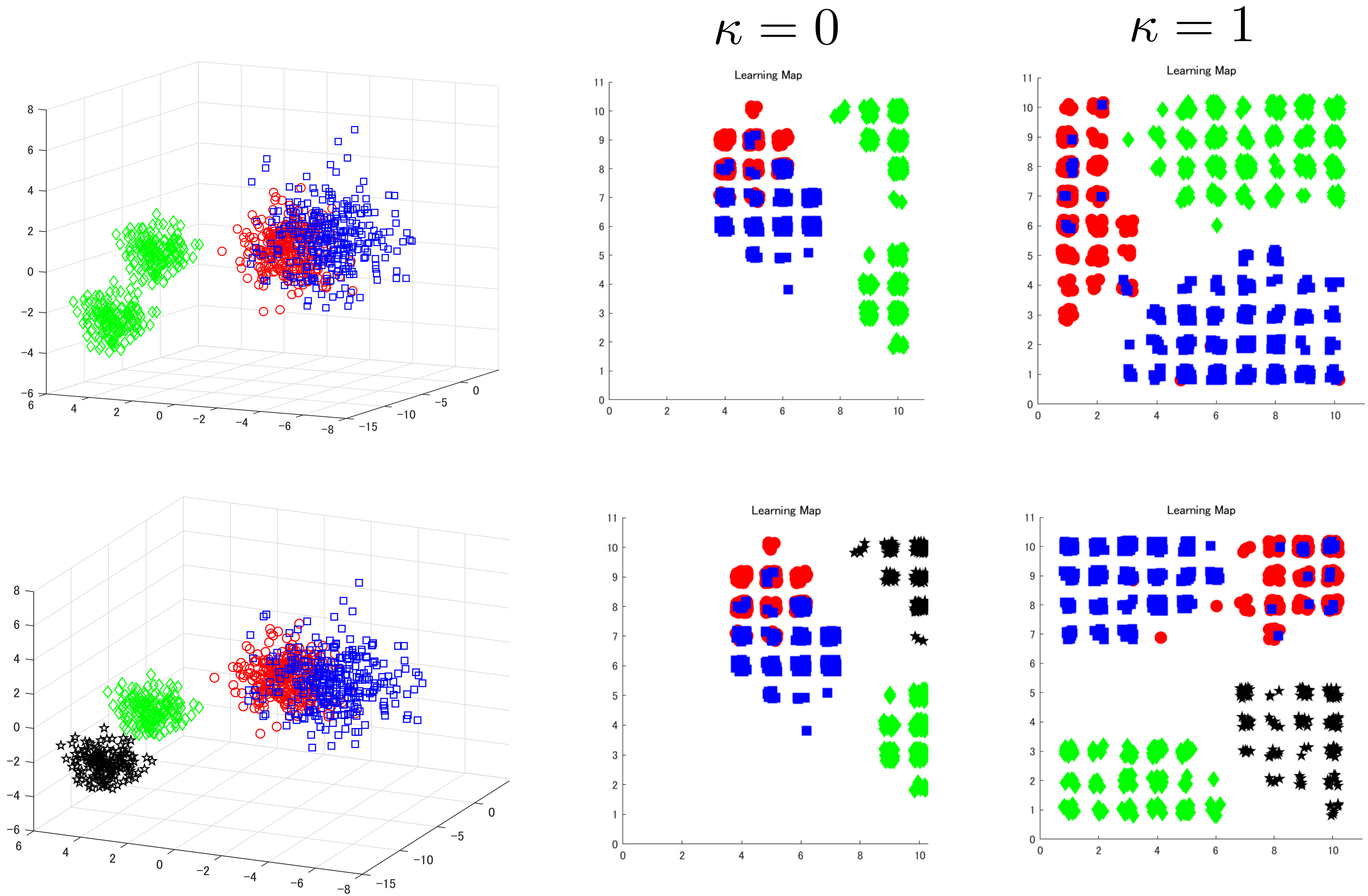
$$\delta_j^{hid} = \frac{1}{\sigma^2} \{ (1-\kappa) \sum_k \delta_k^{dec} v_{jk}^{dec} + \kappa \sum_l \delta_l^{cls} v_{jl}^{cls} \}$$

positive: SOM

negative: repulsive SOM

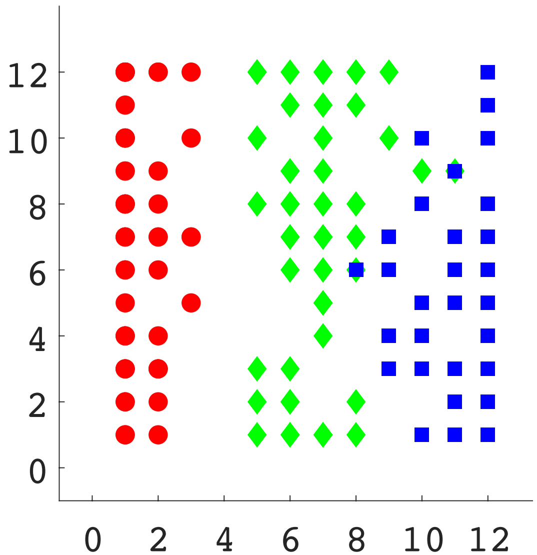
CHUKYO UNIVERSITY

Soft-supervised topological autoencoder(STA)

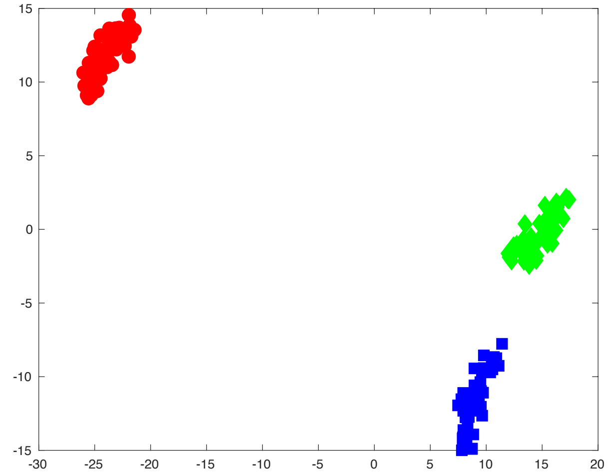


CHUKYO UNIVERSITY

Soft-supervised topological autoencoder(STA)



SOM

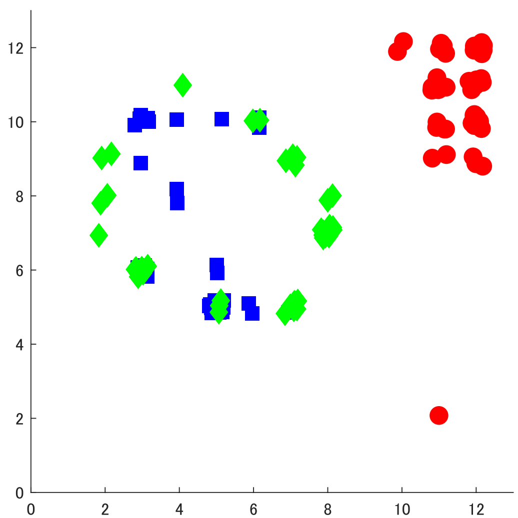


t-SNE

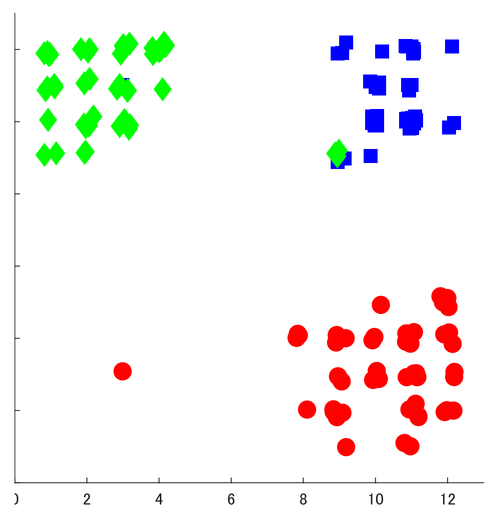
Iris

dim: 4

class: 3



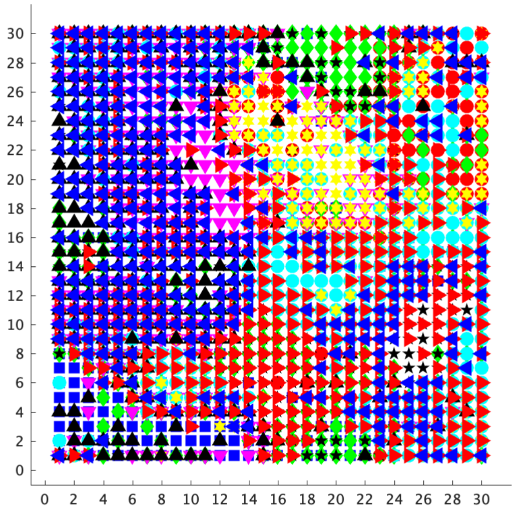
$\kappa = 0$



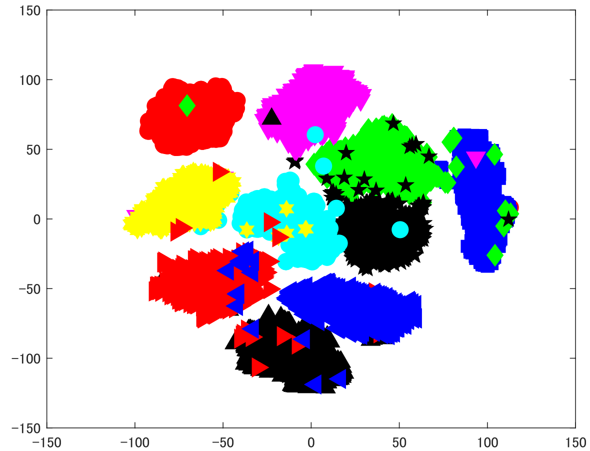
$\kappa = 1$

CHUKYO UNIVERSITY

Soft-supervised topological autoencoder(STA)



SOM

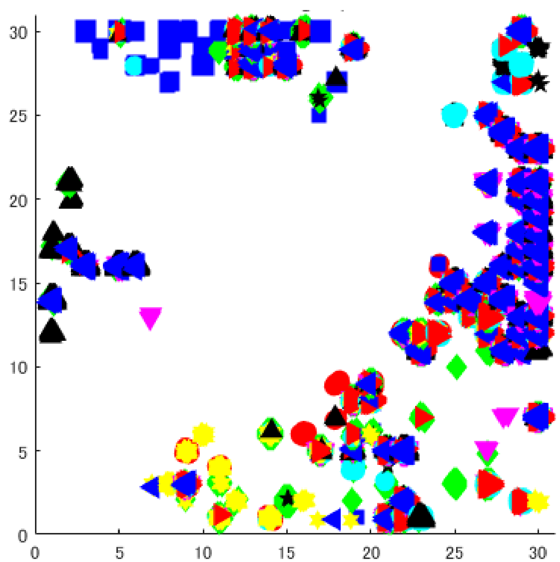


t-SNE

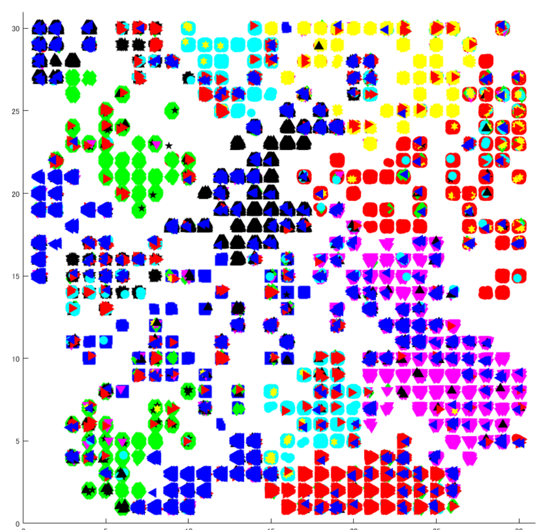
MNIST

dim 28 x 28

class: 10



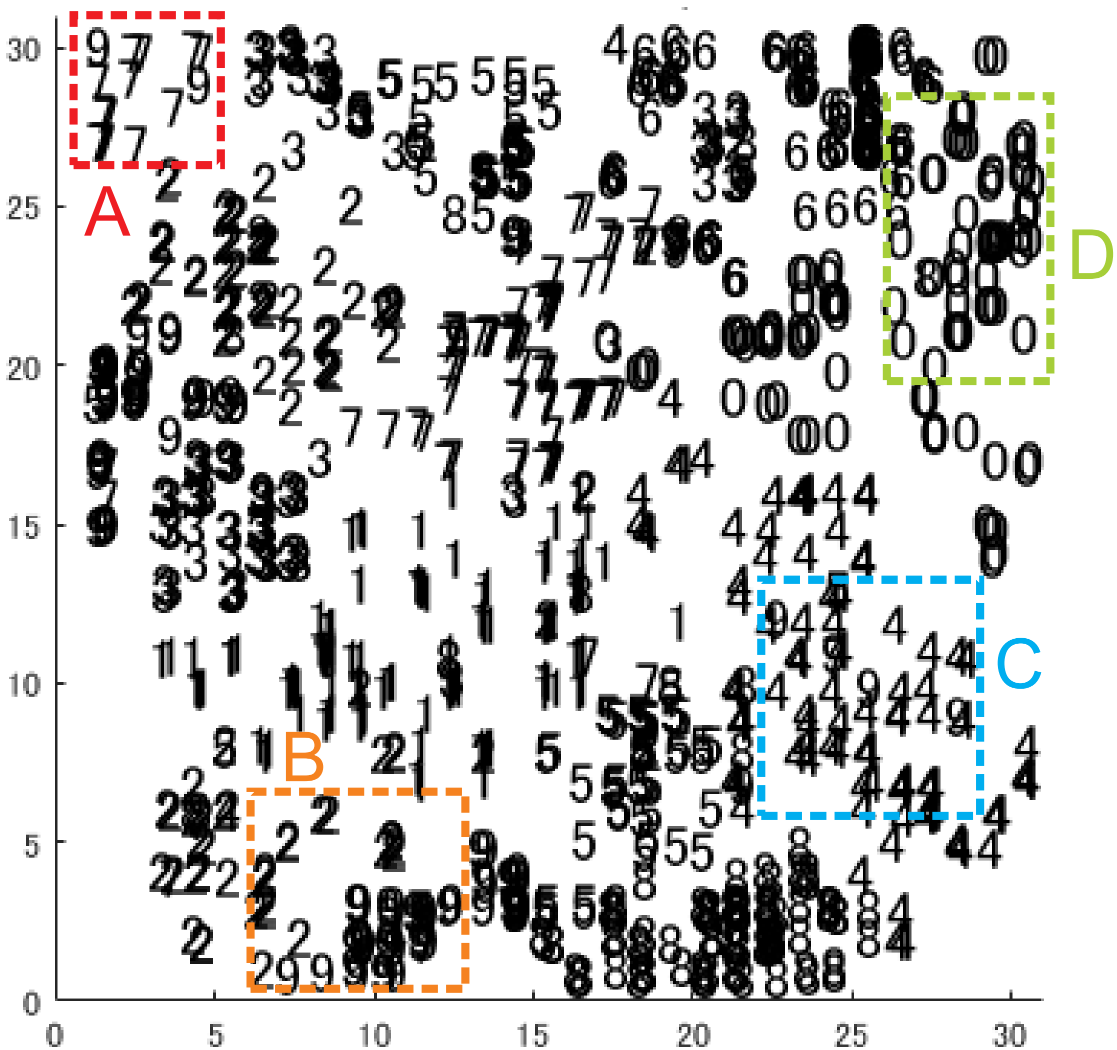
$\kappa = 0$



$\kappa = 1$

CHUKYO UNIVERSITY

Soft-supervised topological autoencoder(STA)



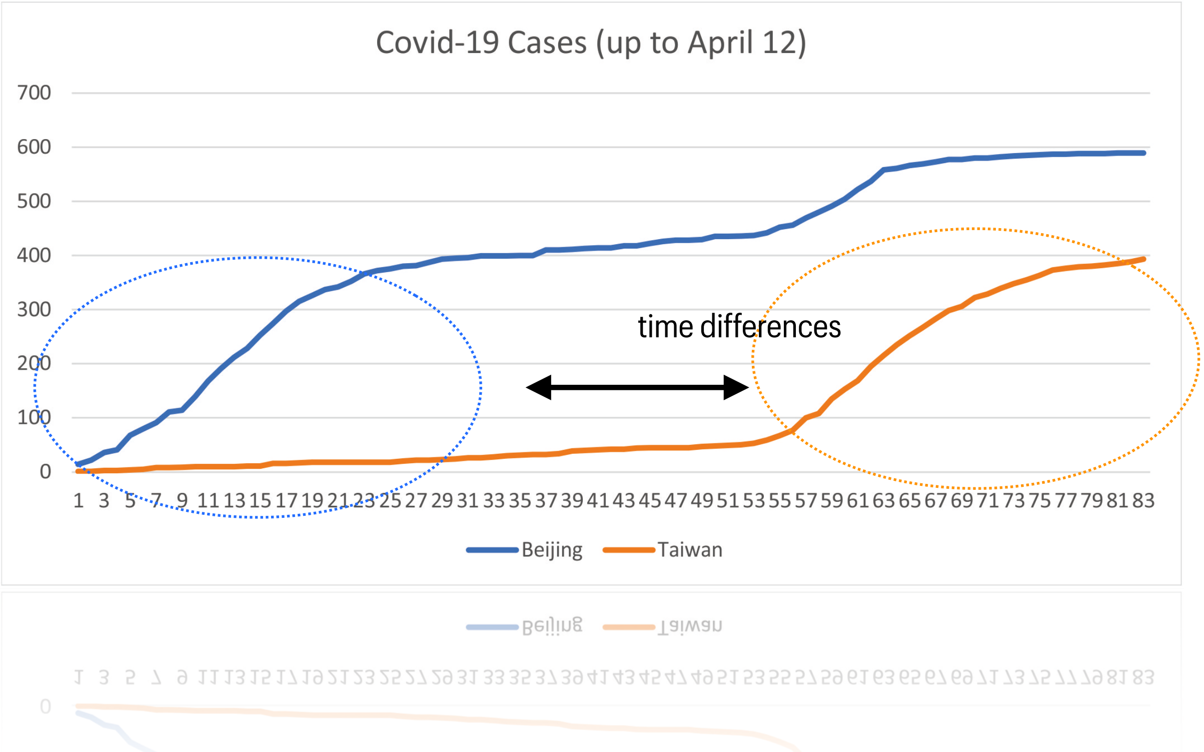
Applications

Pairwise Prediction of COVID-19 patients number

Data are short: started on January 22

Data are abundant: over 250 countries

Time Differences in outbreak

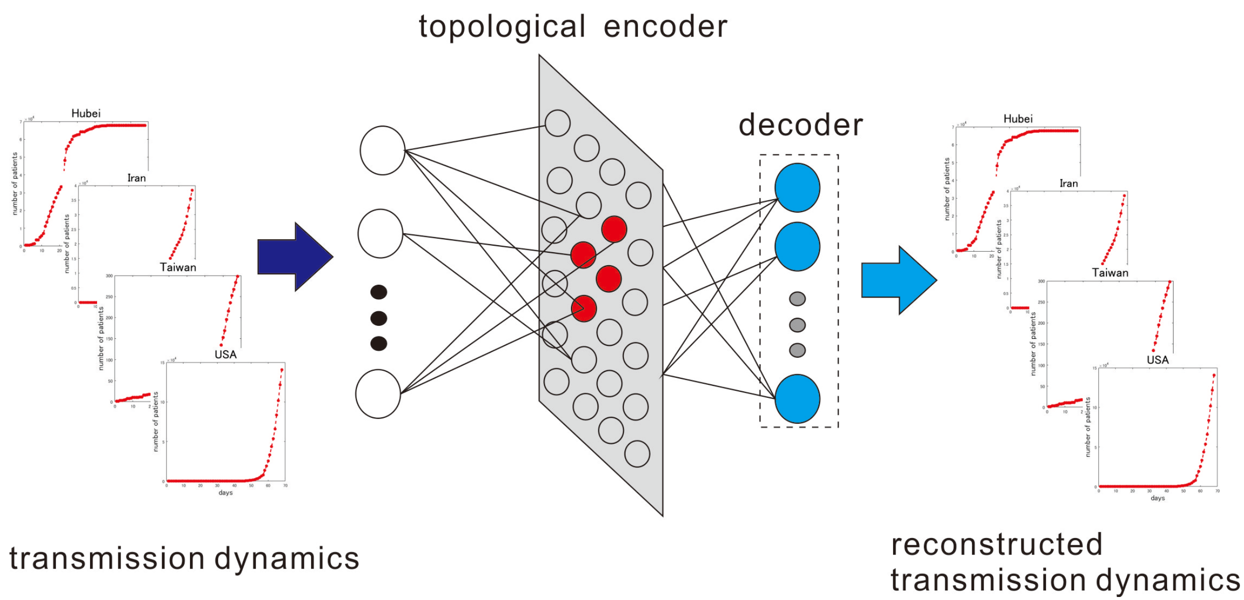


Utilizing Beijing's time series to train a neural network for predicting Taiwan's

Applications

Pairwise Prediction of COVID-19 patients number

Problem: How to find a pair for a target country ?

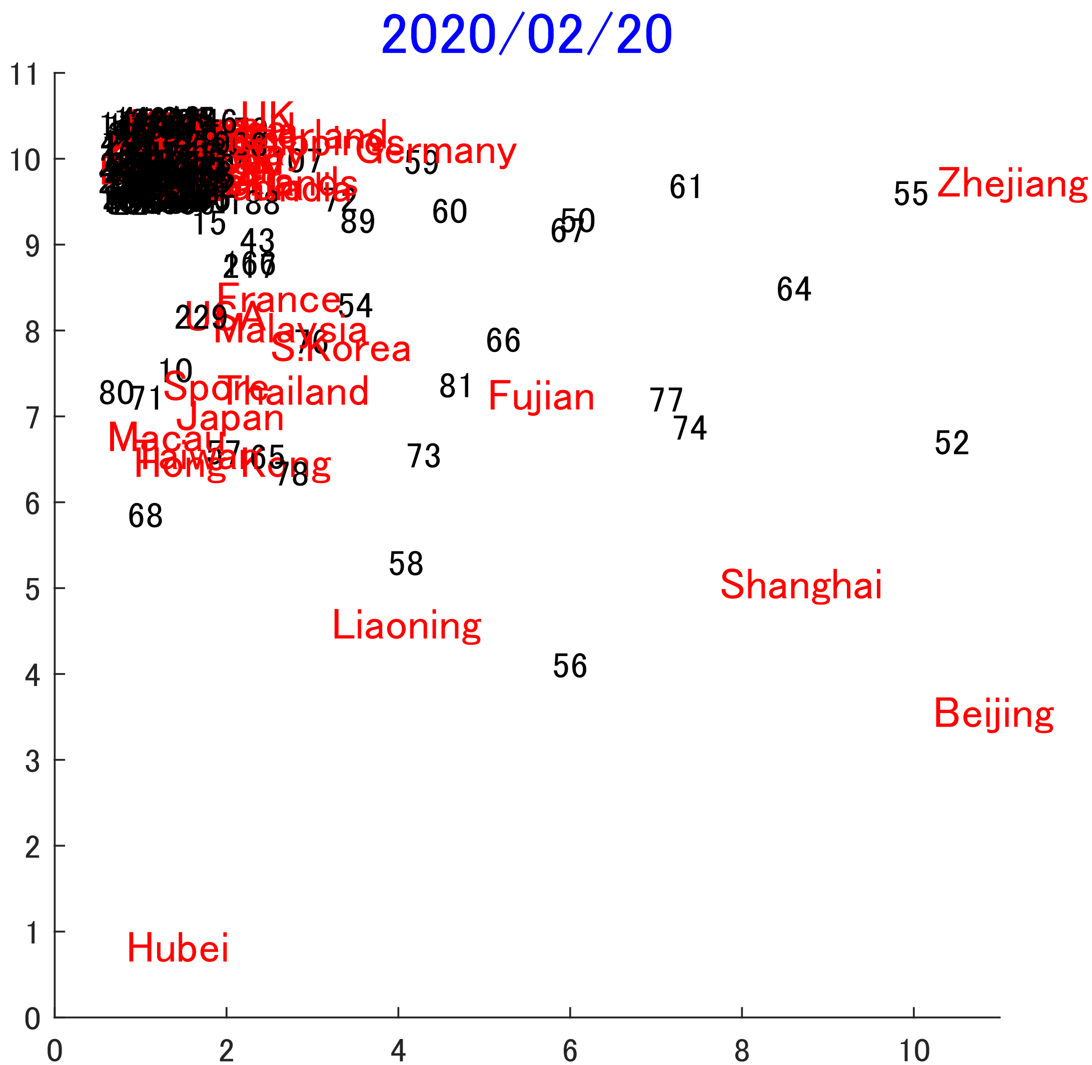


P. Hartono, Similarity Maps and Pairwise Prediction for Transmission Dynamics of COVID-19 with Neural Networks, Informatics in Medicine Unlocked Vol. 20, 100386 (2020)

CHUKYO UNIVERSITY

Applications

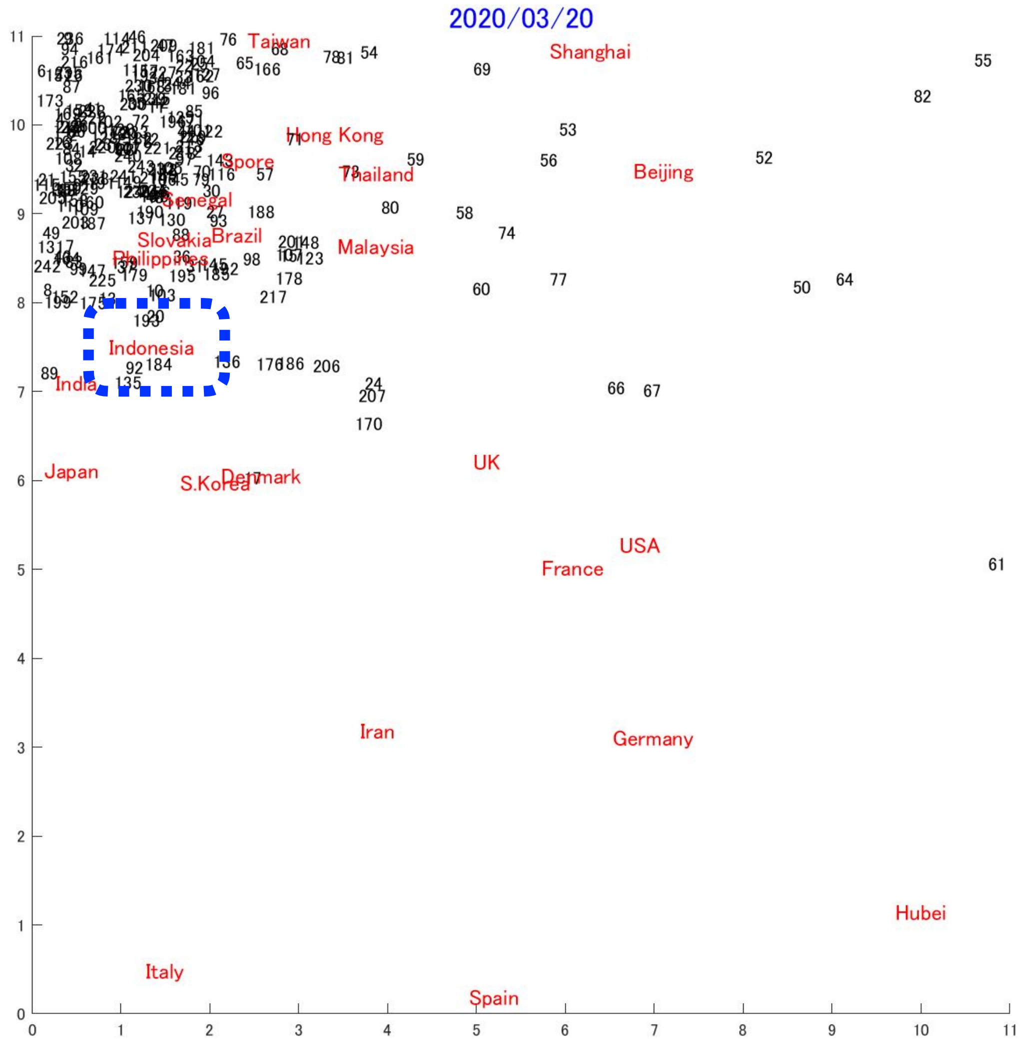
Pairwise Prediction of COVID-19 patients number



CHUKYO UNIVERSITY

Applications

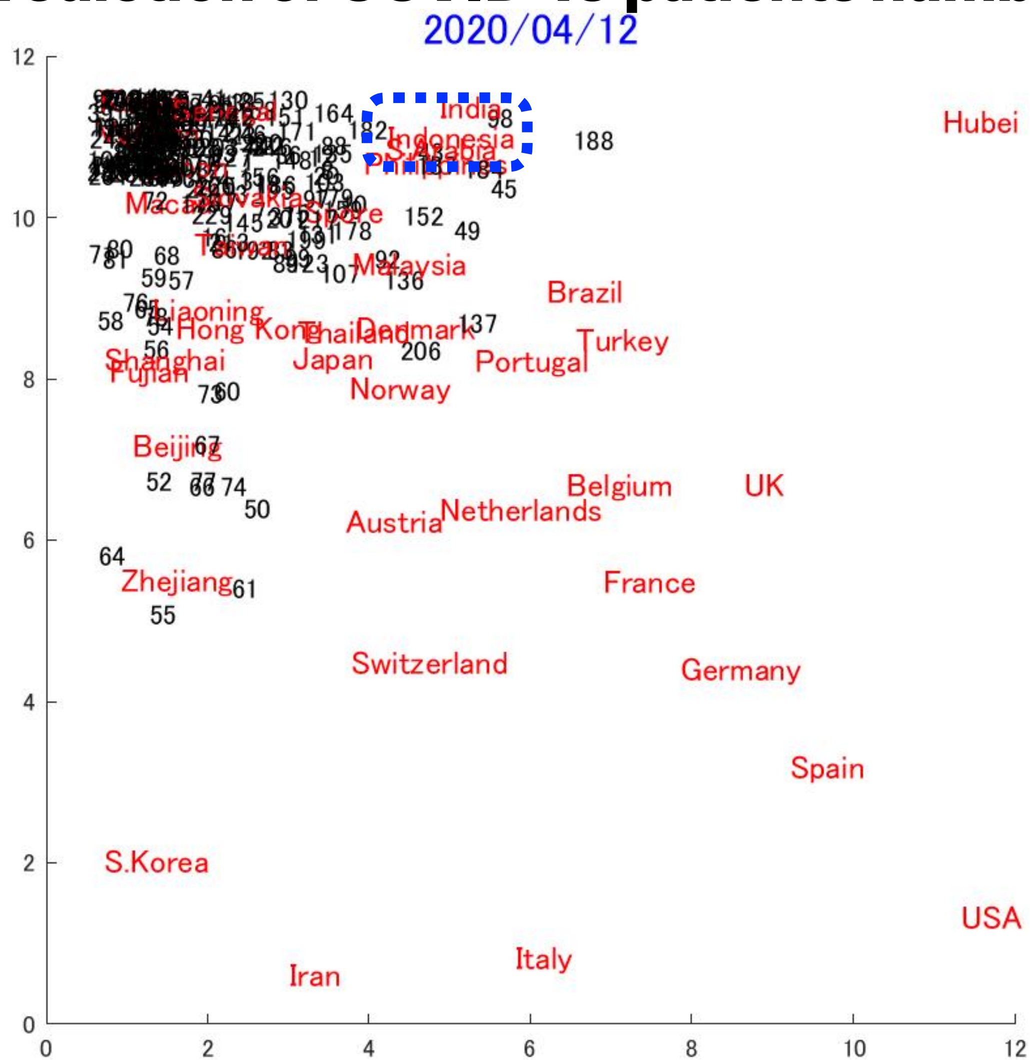
Pairwise Prediction of COVID-19 patients number



CHUKYO UNIVERSITY

Applications

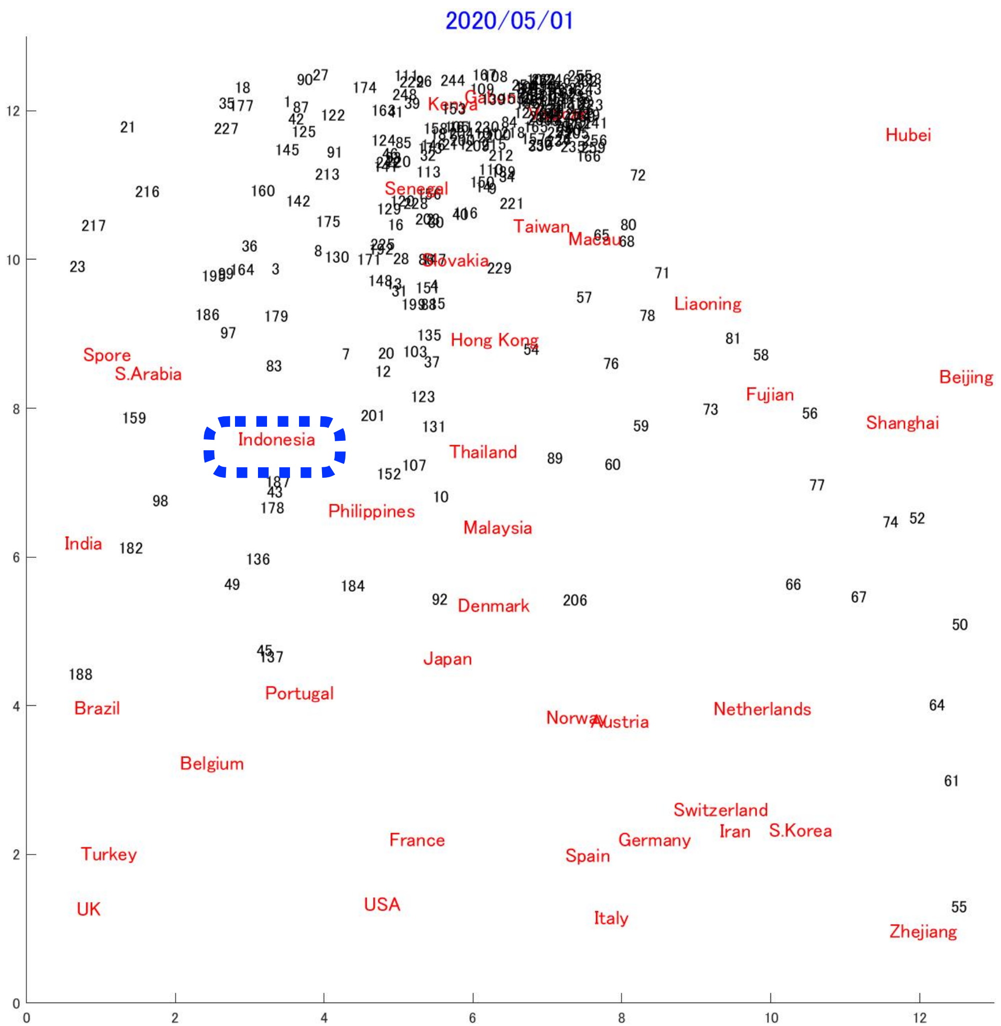
Pairwise Prediction of COVID-19 patients number



CHUKYO UNIVERSITY

Applications

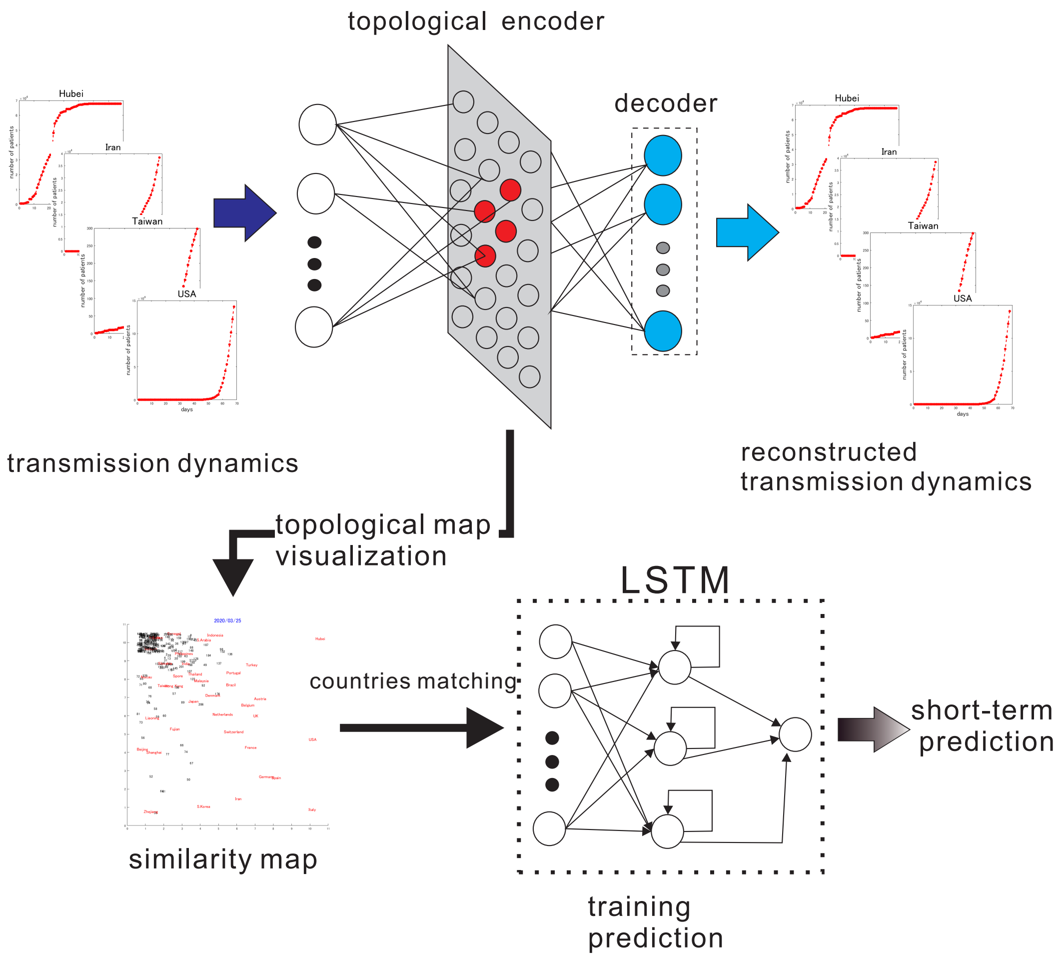
Pairwise Prediction of COVID-19 patients number



CHUKYO UNIVERSITY

Applications

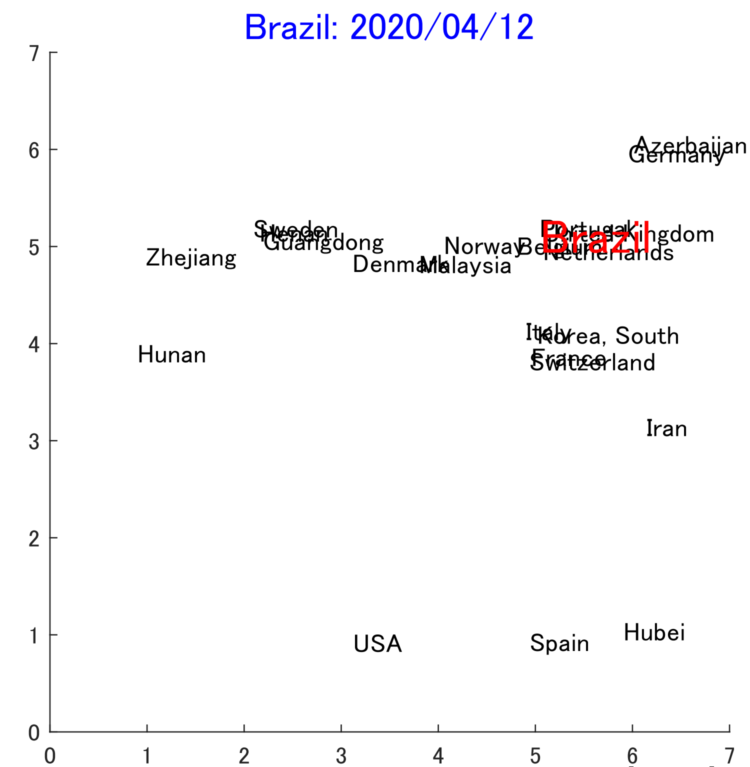
Pairwise Prediction of COVID-19 patients number

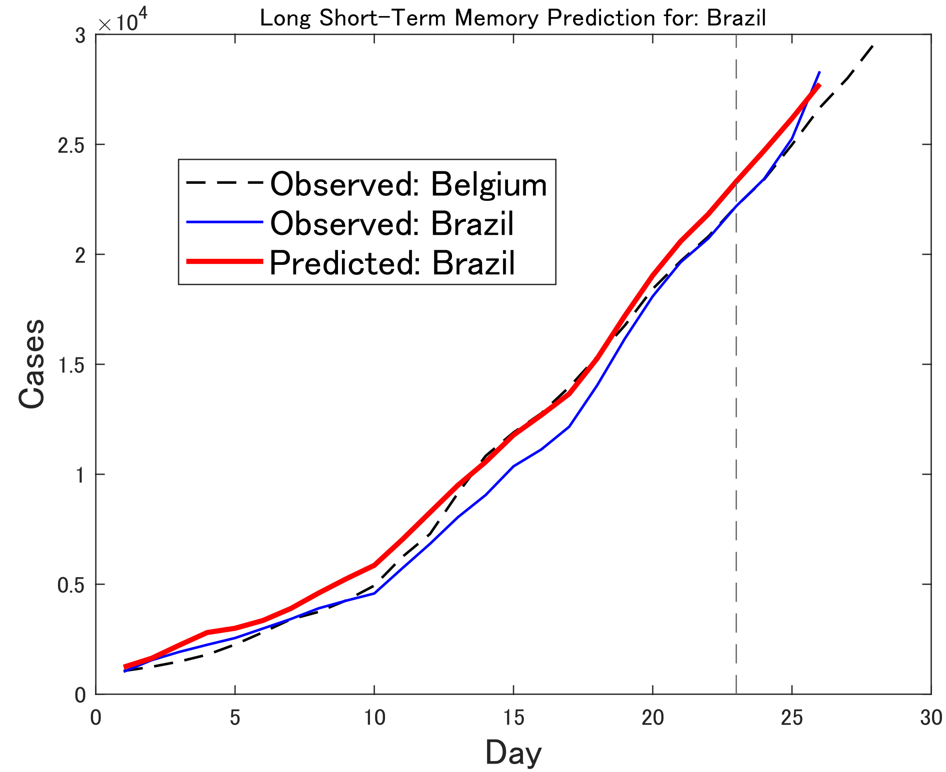


CHUKYO UNIVERSITY

Applications

Pairwise Prediction of COVID-19 patients number



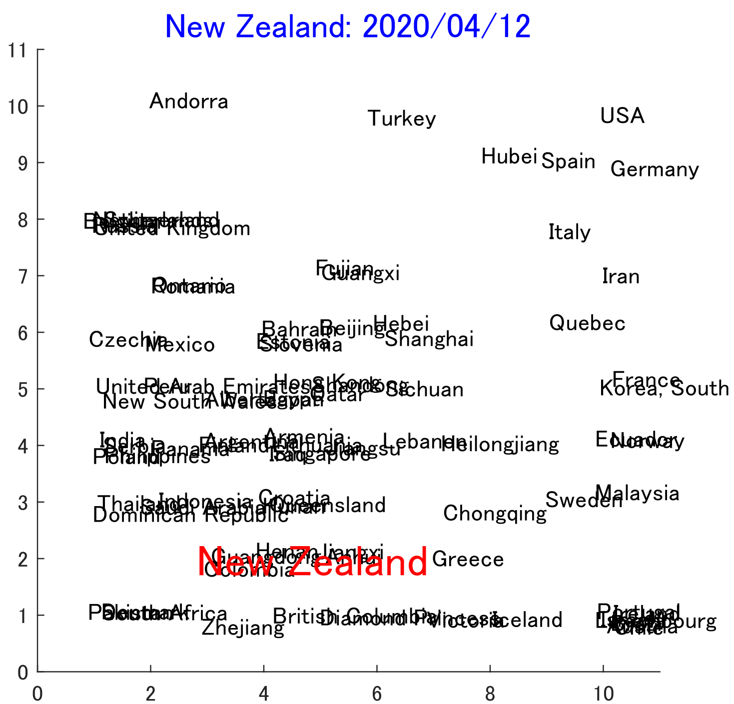


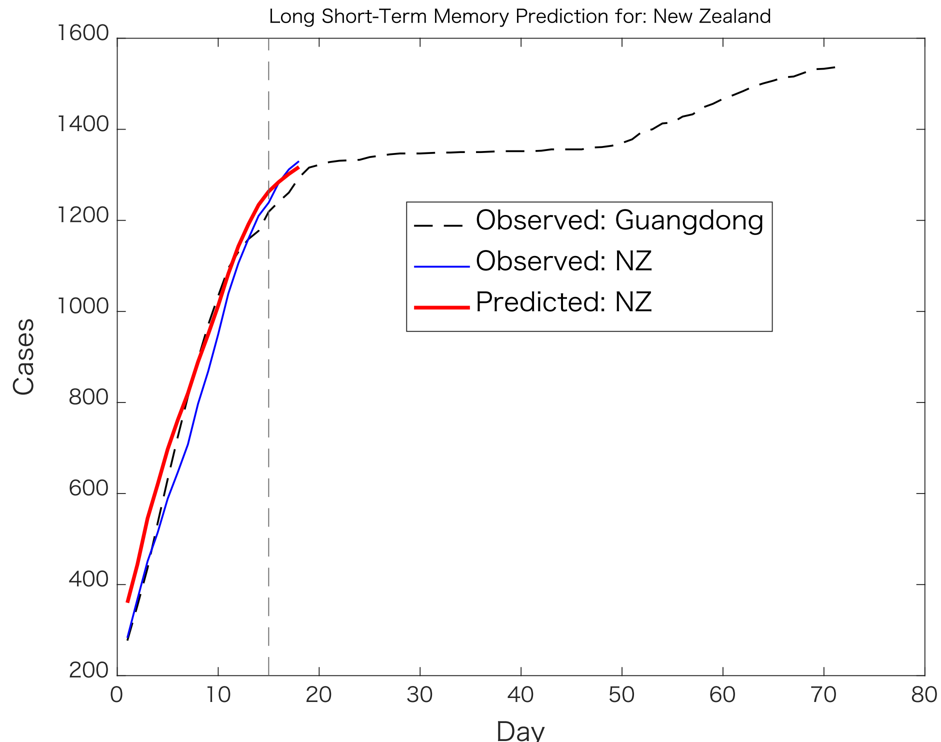
Brazil-Belgium, post 1000 cases

pairwise_error: 3.66%, single: 7.91%

Applications

Pairwise Prediction of COVID-19 patients number





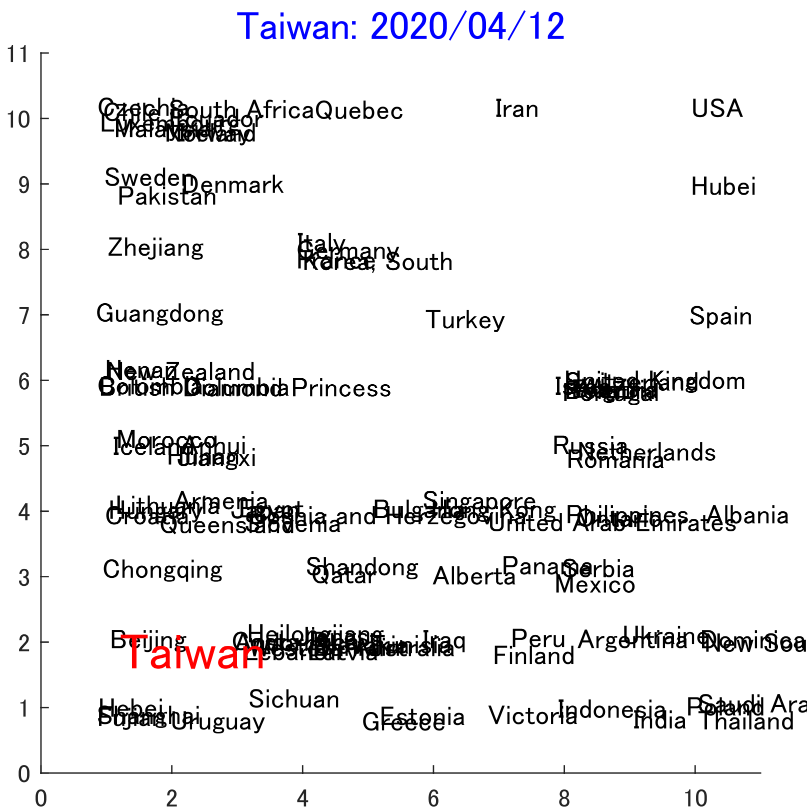
NZ-Guangdong, post 300 cases

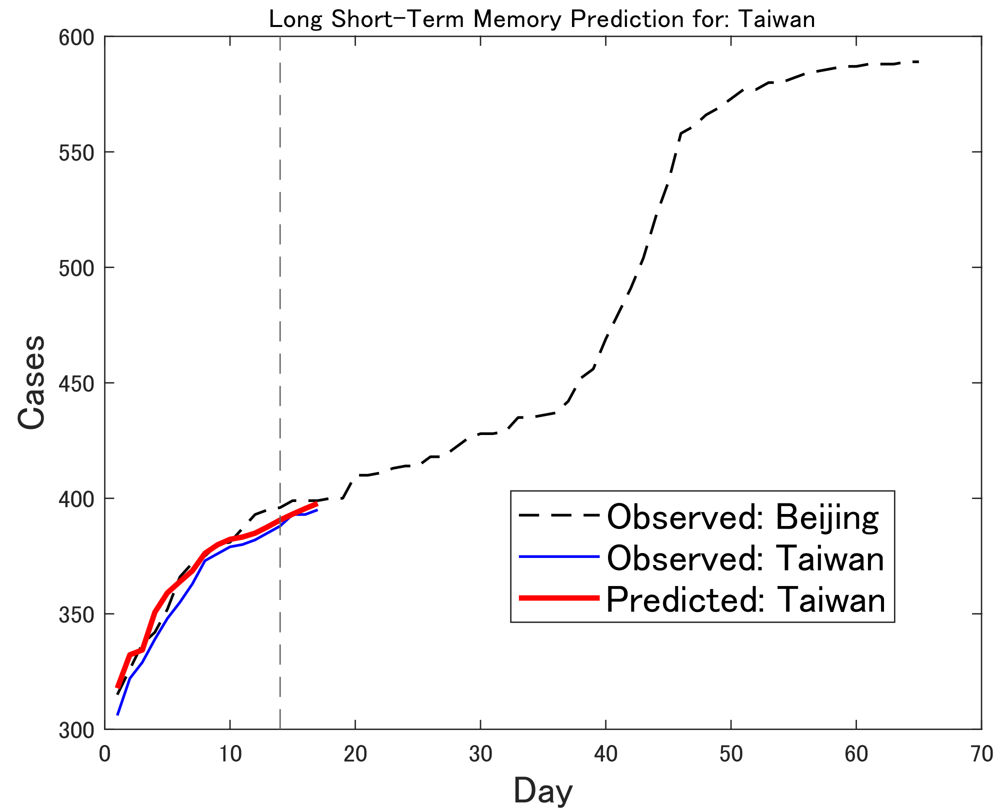
pairwise_error: 0.74%, single: 3.72%

CHUKYO UNIVERSITY

Applications

Pairwise Prediction of COVID-19 patients number



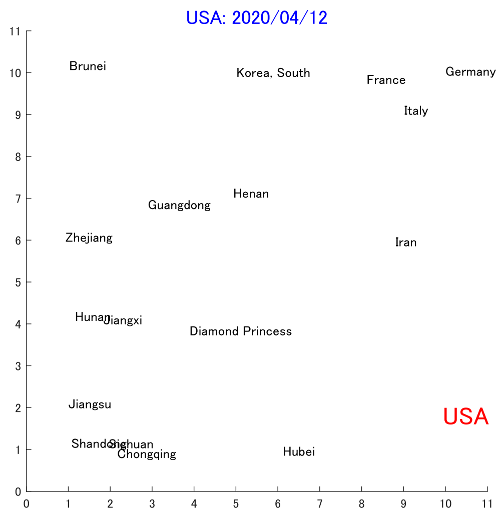


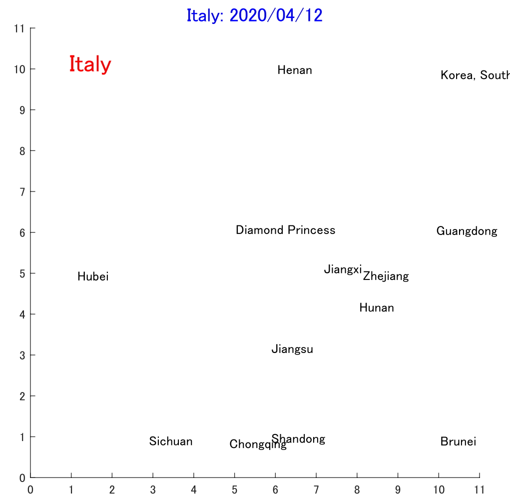
Taiwan-Beijing, post 300 cases

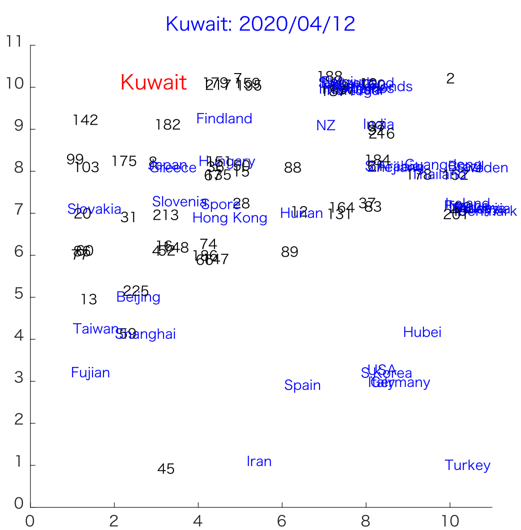
pairwise_error: 0.40%, single: 1.32%

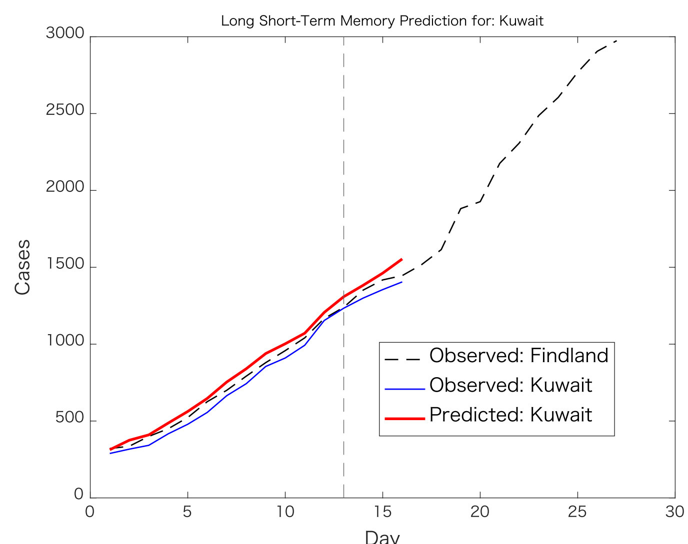
CHUKYO UNIVERSITY

Applications









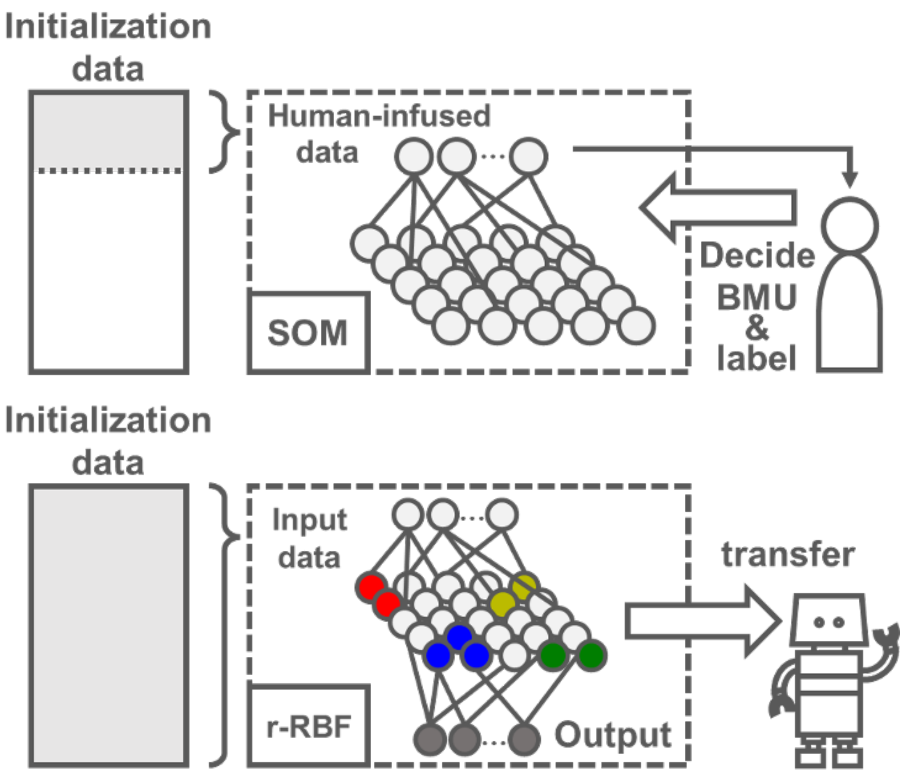
Kuwait-Findland, post 300 cases

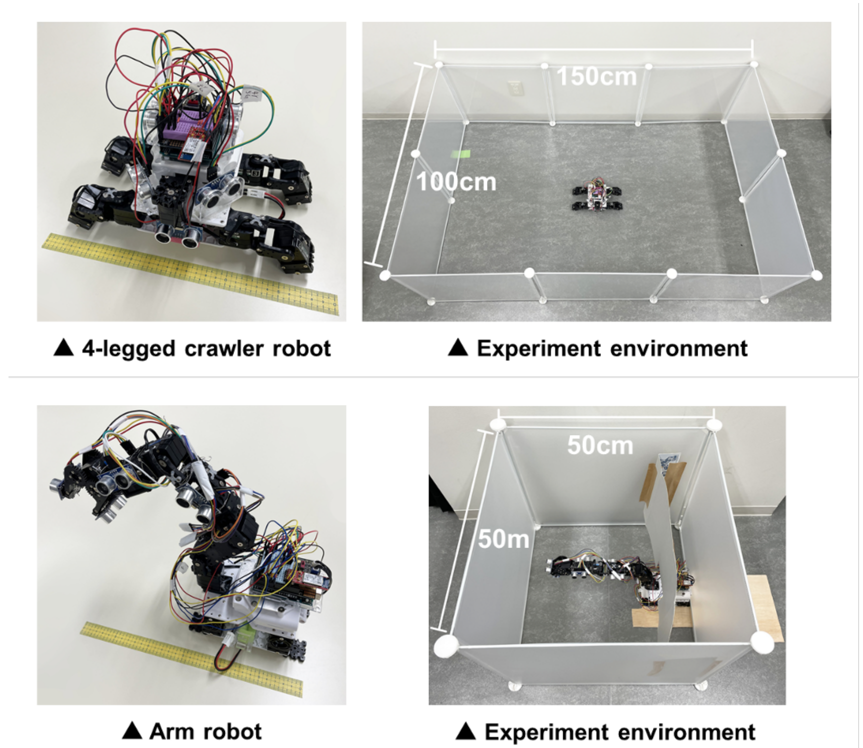
pairwise_error: 10.70%, single: 15.48%

CHUKYO UNIVERSITY

Applications

Transferability and Comprehensibility



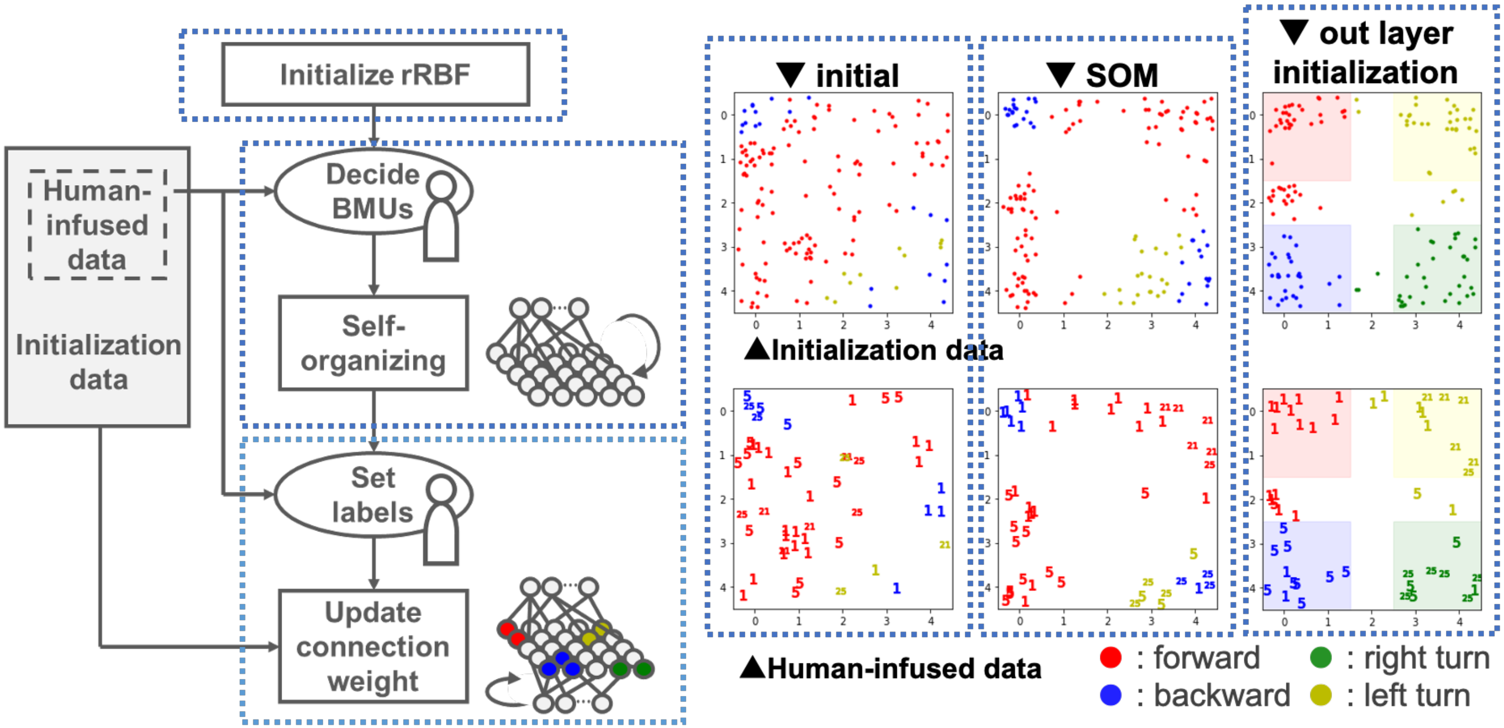


K. Ogawa, P. Hartono, Infusing prior knowledge into topological representations of learning robots, Proc. 27th Int. Symposium on Artificial Life and Robotics, pp. 347-352 (2022) (Young Author Award).

CHUKYO UNIVERSITY

Applications

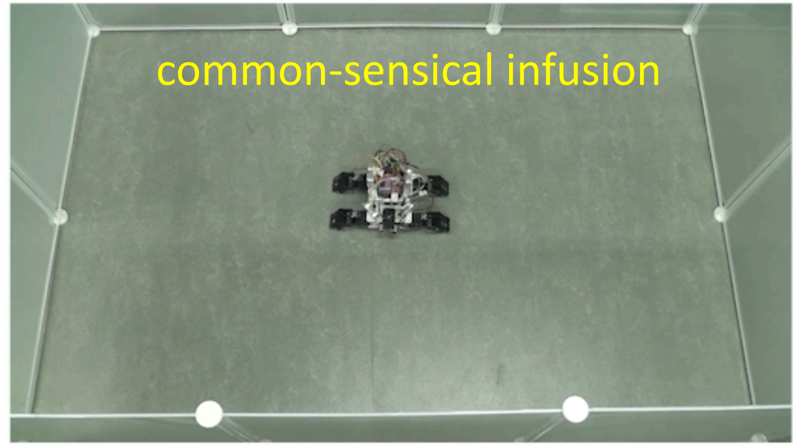
Transferability and Comprehensibility

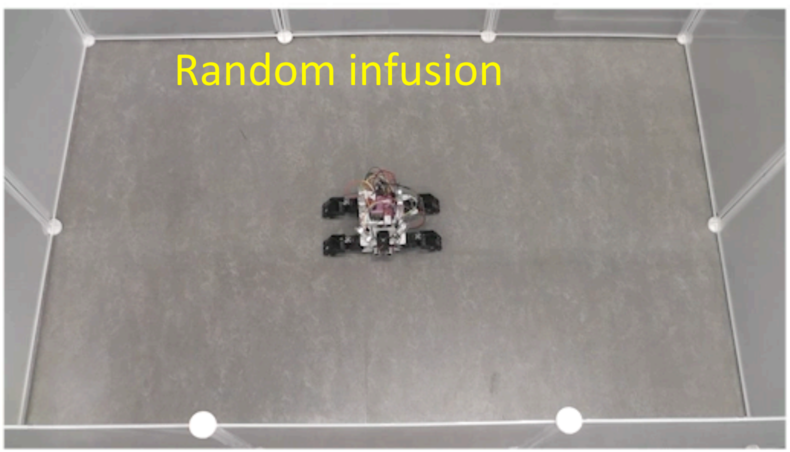


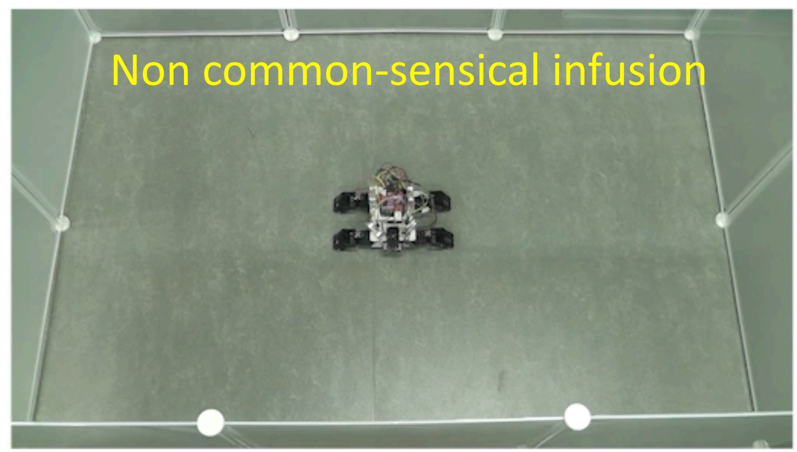
K. Ogawa and P. Hartono, Infusing common-sensical prior knowledge into topologica
representations of learning robots, (under review)

Applications

Transferability and Comprehensibility







K. Ogawa and P. Hartono, Infusing common-sensical prior knowledge into topological representations of learning robots, (under review)

Conclusions

Internal Representations define the performance of NN

Understanding Internal Representations -> Understanding neural networks

Knowing more but constraining: distributed representation vs topological representations

Explanability, Transferability

CHUKYO UNIVERSITY

Thanks to

Kana Ogawa
Hiroki Kishi
Shusuke Kobayashi

Thomas Trappenberg
Peter Sincak
Patrik Sabol

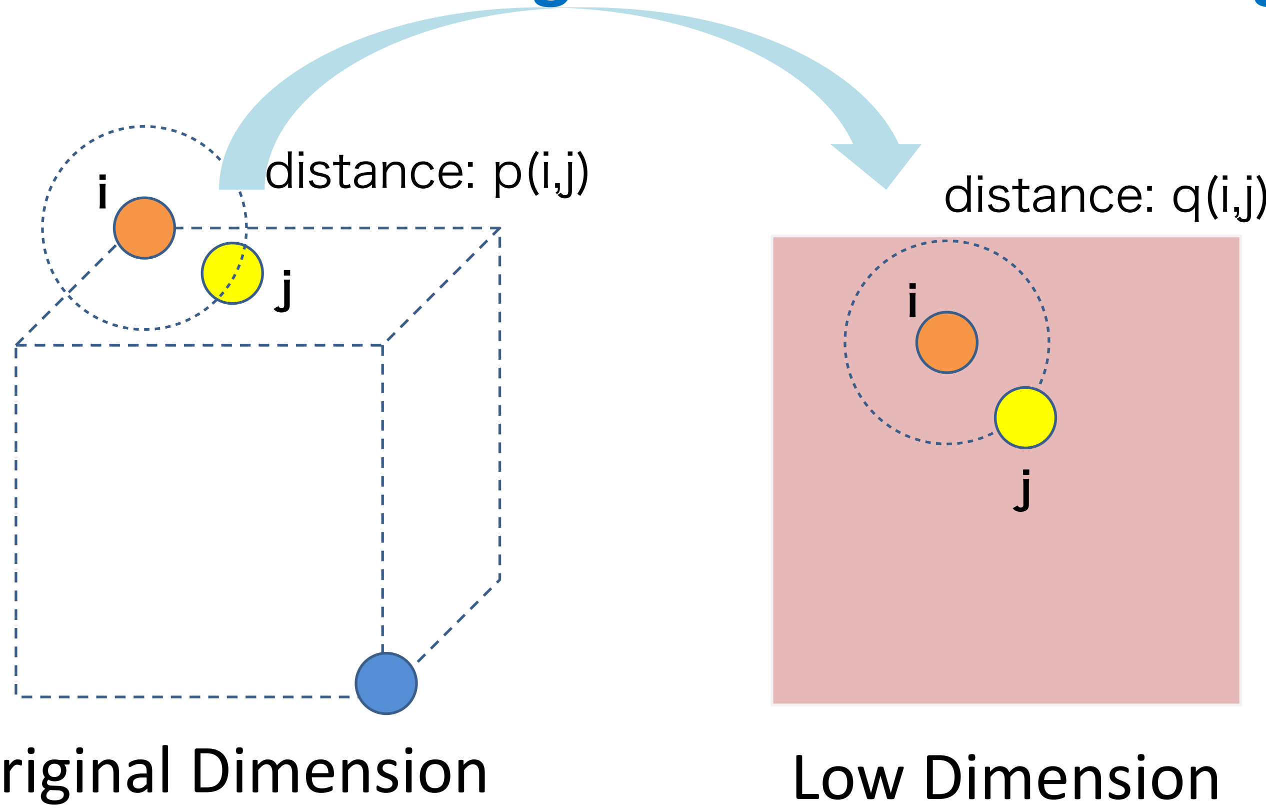
Comments & questions:

hartono@ieee.org

CHUKYO UNIVERSITY

Internal Representations of Neural Nets

t-Stochastic Neighborhood Embedding (t-SNE)



$$min \sum_i \sum_j p(i,j) log \frac{p(i,j)}{q(i,j)}$$

Kullback-Liebler divergence

L. van der Maaten, G. Hinton. Visualizing Data using t-SNE, Journal of Machine Learning Research, Vol. 9, pp. 2579-2605 (2008).