



Introduction to Research & Experiences

Younghoon Kim, Ph.D.

Profile



Younghoon Kim

- **Assistant Professor**, Kyung Hee University (2022. 3 ~)
- **Assistant Professor**, Sungshin Women's University (2020. 3 ~ 2022. 2)
- **Adjunct Professor**, Samsung Institute of Technology (2020, 7 ~ 2021, 2)
- **Data Scientist**, SK Innovation (2019. 3 ~ 2020. 2)
- Korea University, Industrial Management & Engineering
 - Ph.D. (2013.3 ~ 2019.2), Advisor: Prof. Seoung Bum Kim

Research Areas



- Robust Deep Learning against Data Variations
- Machine Learning Algorithm with Discrete Optimization
- Smart Manufacturing & Quality
- Bio-Chemical Data Analysis

Research Activities & Grants



- 14 Accepted Papers, 5 Working Papers, 1 Patent
- 15 Industrial Research Projects (Samsung Electronics, SK Hynix, etc.)
- Electrical Fire Prevention System with IoT and Artificial Intelligence, MOLIT, Apr 2021 – Dec 2022, (1.8 billion KRW / 2 Years)
- Recycling Malfunctioned Gas Insulated Substation Controller with Data Analytics, KIAT, Oct 2020 – Dec 2022, (1.1 billion KRW / 3 Years)
- Robust Deep Learning with Test-time Data Augmentation for Autonomous Vision Inspection, NRF, Sep 2020 – Aug 2021, (30 million KRW / 1 Year)

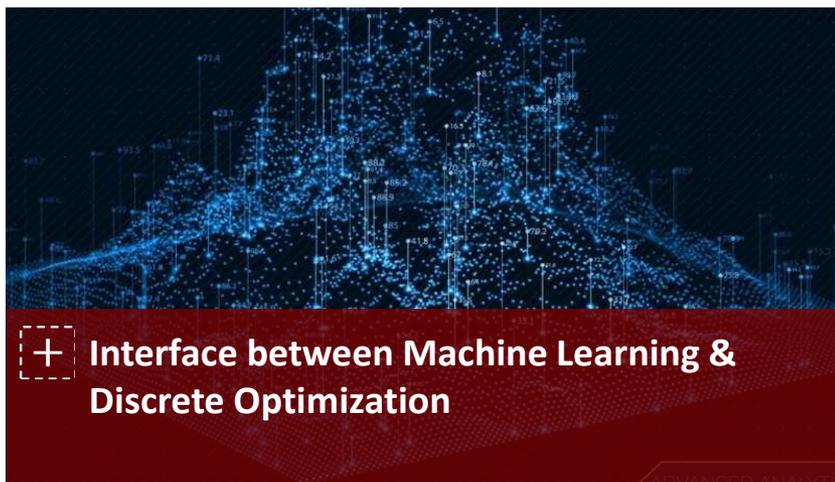
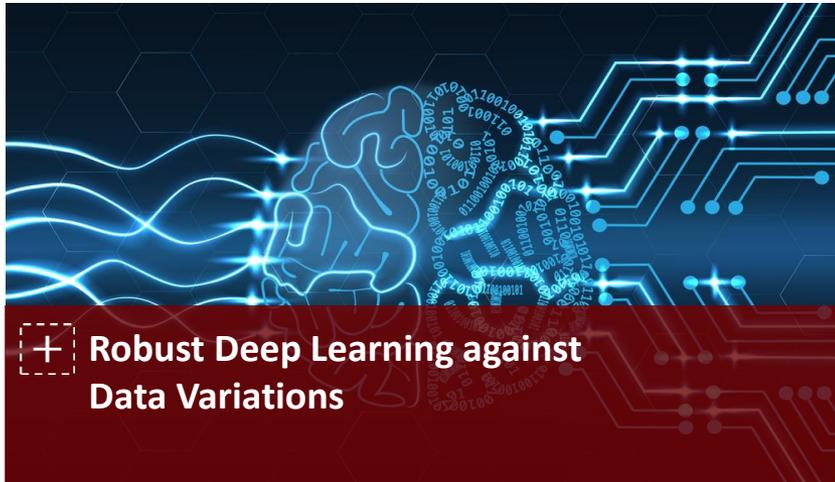


Research Highlights

Research Highlights



Research Area



Details

- Learning Loss for Test-time Data Augmentation, *NeurIPS*
- Preprocessing-free Bacteria Spectra Identification with Explainable Convolutional Neural Network, *Journal of Chemometrics*
- Multiple Instance Neural Network for Cancer Diagnosis with Sparse Attention, **Submitted**
- Robust Graph Convolutional Neural Network with Graph Data Augmentation, *Ongoing*

- Collinear Groupwise Feature Selection via Discrete Fusion Group Regression, *Pattern Recognition*
- *Outer-Points Shaver: Robust Graph-Based Clustering via Node Cutting, Pattern Recognition*
- *Swarm Ascending: Swarm-Intelligence-Based Exemplar Group Detection for Robust Clustering, Applied Soft Computing*

Learning Loss for Test-time Augmentation



$\mathcal{T}_{Identity}$
0.086



$\mathcal{T}_{Rotate: -20^\circ}$
0.049



$\mathcal{T}_{Rotate: +20^\circ}$
0.099



$\mathcal{T}_{Zoom: 0.8}$
0.15



$\mathcal{T}_{Zoom: 1.2}$
0.098



$\mathcal{T}_{AutoContrast}$
0.055



$\mathcal{T}_{Sharpness: 0.2}$
0.139



$\mathcal{T}_{Sharpness: 0.5}$
0.119



$\mathcal{T}_{Sharpness: 2.0}$
0.037



$\mathcal{T}_{Sharpness: 4.0}$
0.038

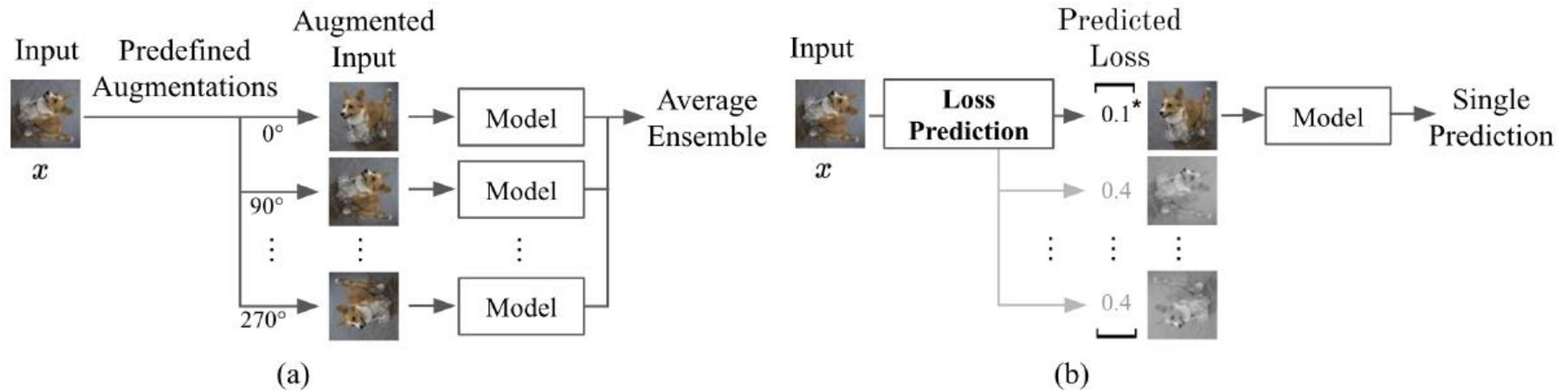


$\mathcal{T}_{Color: 0.5}$
0.091

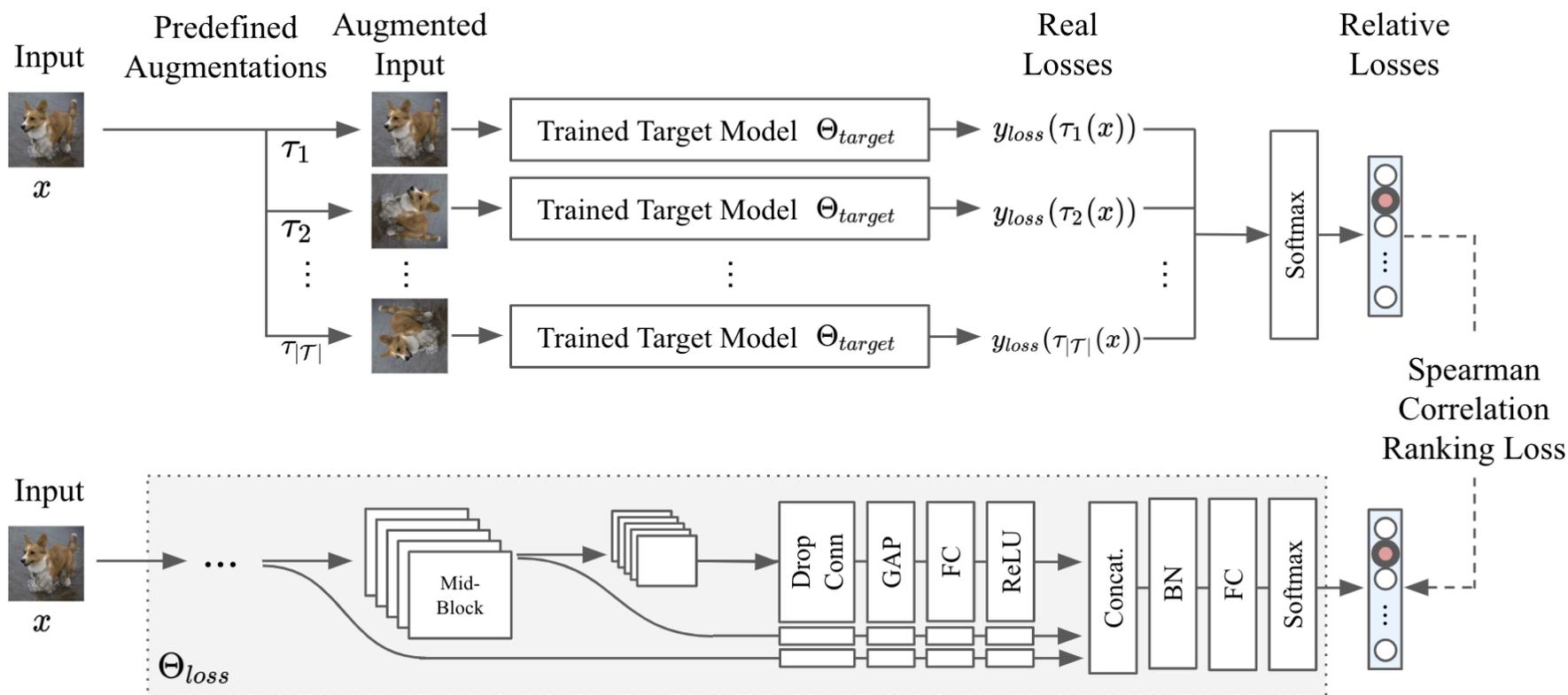


$\mathcal{T}_{Color: 2.0}$
0.039

Learning Loss for Test-time Augmentation



Learning Loss for Test-time Augmentation



Discrete Fusion Group Regression - Formulation

$$\text{Minimize } \|Y - X\beta\|_2^2$$

$$\text{Subject to } \sum_{i=1}^n z_i \leq K_1$$

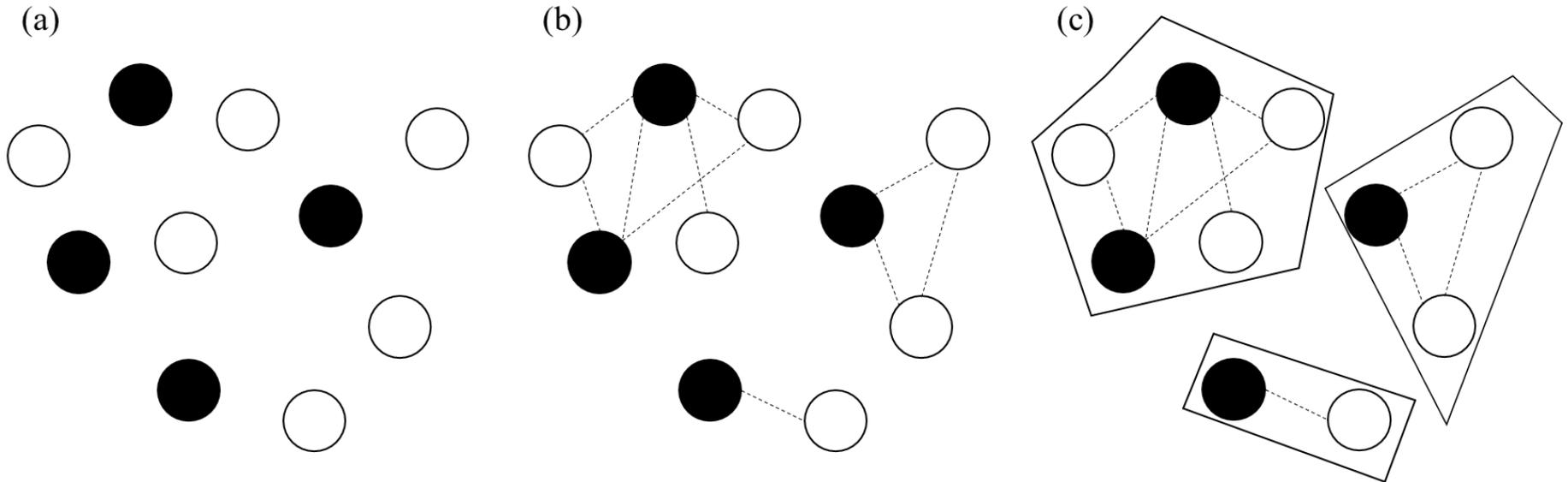
$$\sum_{i \in g_k} z_i \leq K_2, \quad \text{for all } k = 1, 2, \dots, p_g,$$

$$\sum_{i < j, (i, j) \in g_k} (\beta_i - \beta_j)^2 \leq s \quad \text{for all } k = 1, 2, \dots, p_g,$$

$$-Mz_i \leq \beta_i \leq Mz_i \quad \text{for all } i = 1, \dots, n,$$

where $Y \in R^m$, $X \in R^{m \times n}$, $\beta \in R^n$, $z \in B^n$.

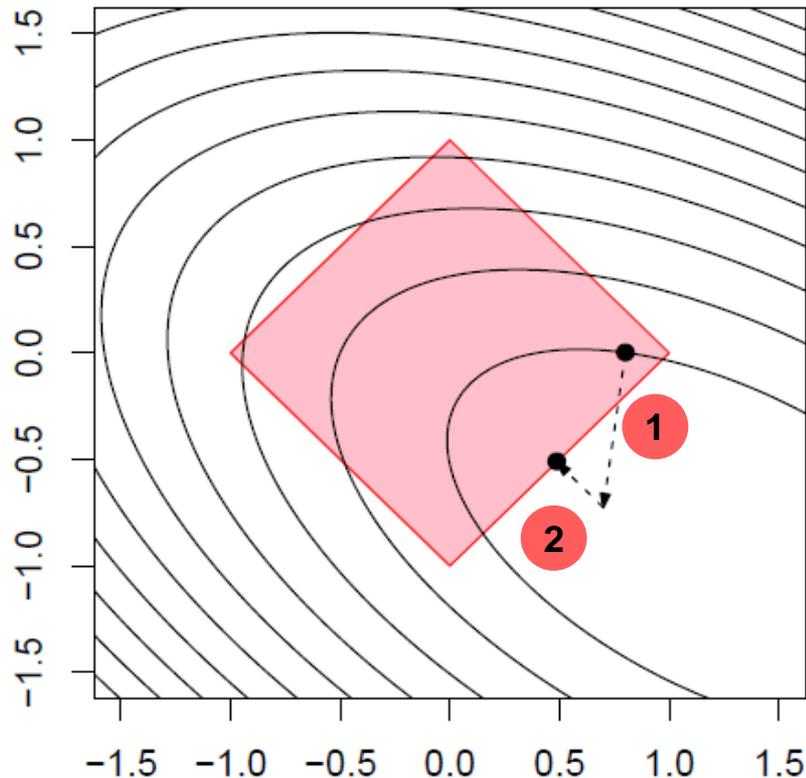
Discrete Fusion Group Regression - Collinear Group Detection



Three different phases of CGD with toy example: (a) identifying highly correlated features with VIF, (b) constructing a collinear feature graph, and (c) defining connected components as groups of collinear features

Discrete Fusion Group Regression - Projected Gradient Descent

$$\beta^+ = P_C(\beta - t\nabla g(\beta))$$



Step 1

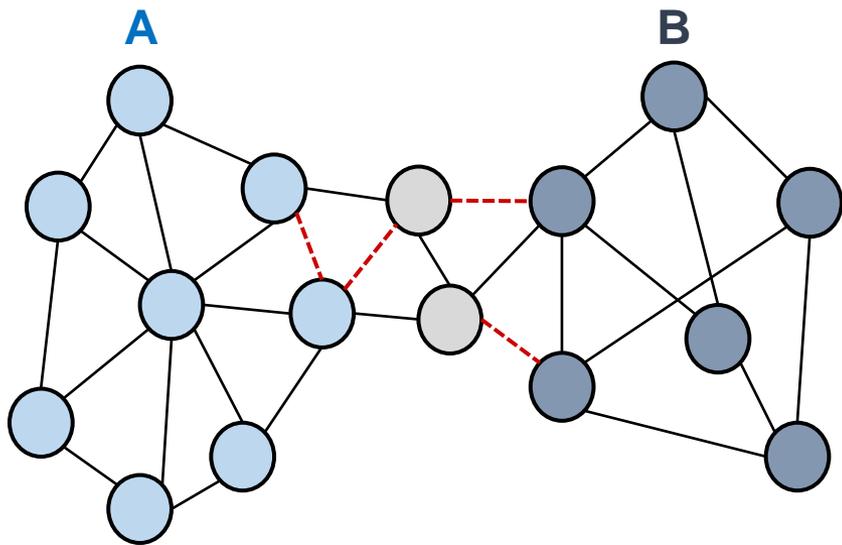
- Gradient descent with differentiable part $g(\beta)$

Step 2

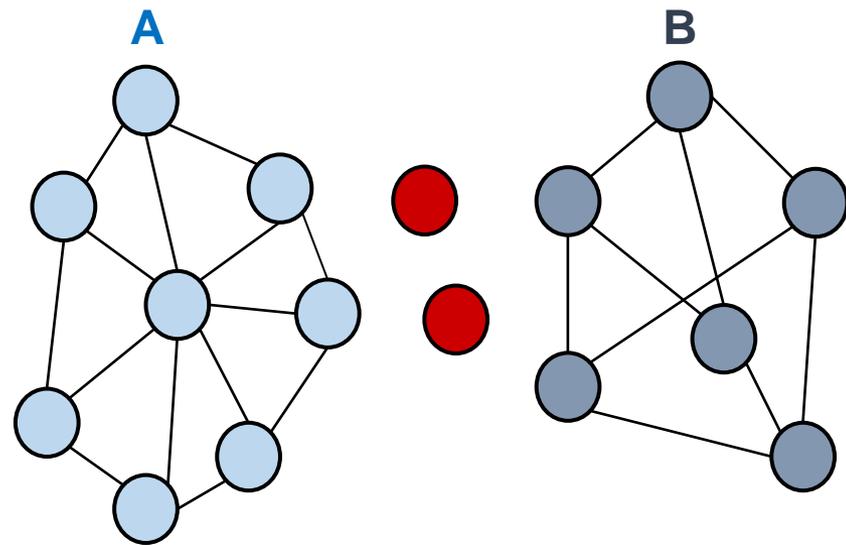
- Projection to the feasible space of non-differentiable part $I_C(\beta)$

Robust Graph-Based Clustering with Outer Points Shaver

Edge Cutting



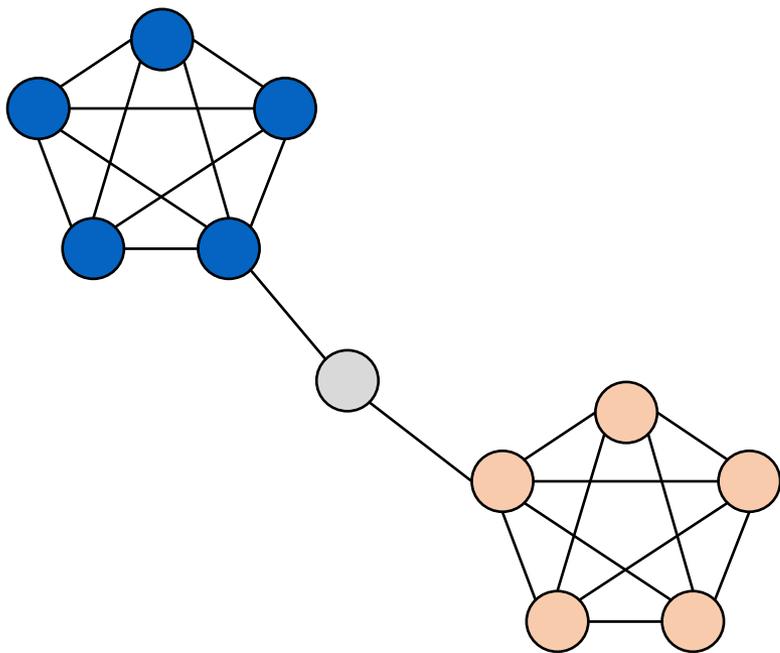
Node Cutting



Robust Graph-Based Clustering with Outer Points Shaver

$$\text{Minimize } \|S_{sum} - S\beta_o\|_2^2 + \lambda_1 \|\beta_o\|_1 + \lambda_2 \|\beta_o\|_2$$

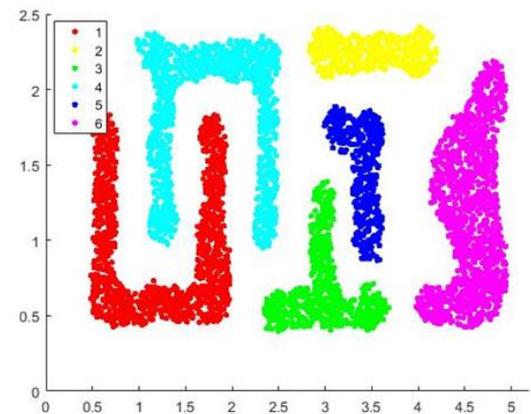
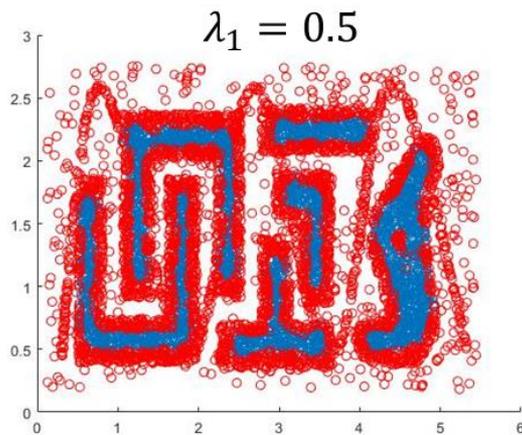
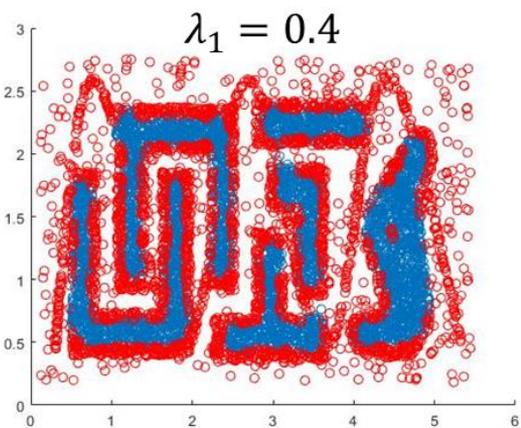
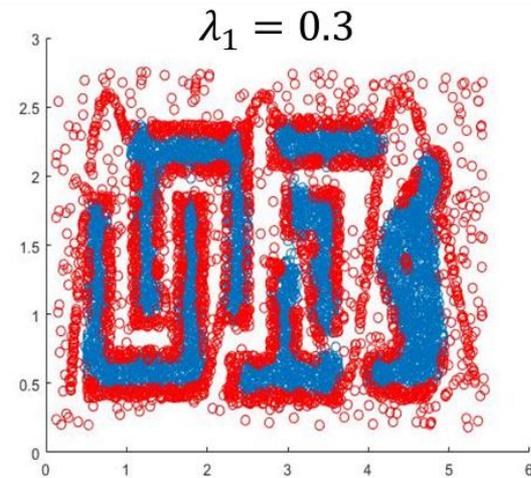
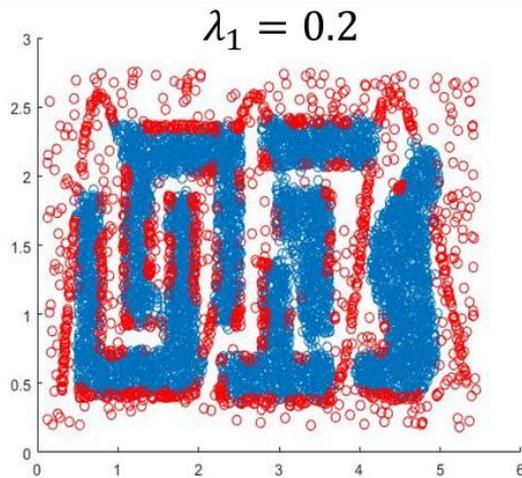
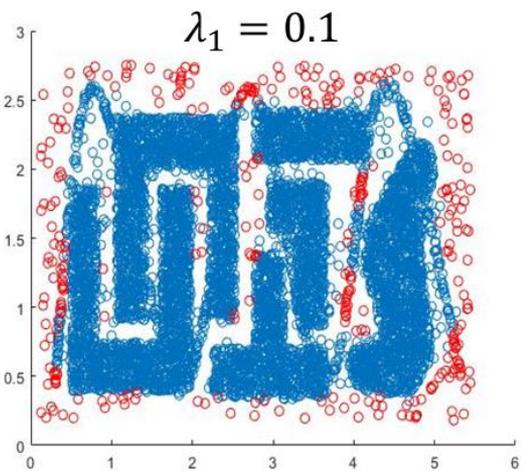
KNN graph



Pseudo Density & Similarity Matrix

4		1	1	1	1									
4	1		1	1	1									
4	1	1		1	1									
4	1	1	1		1									
5	1	1	1	1		1								
2					1		1							
5						1		1	1	1	1	1	1	1
4							1		1	1	1	1	1	1
4								1	1		1	1	1	1
4									1	1	1			1
4										1	1	1	1	

Robust Graph-Based Clustering with Outer Points Shaver



Research Highlights



Research Area



Details

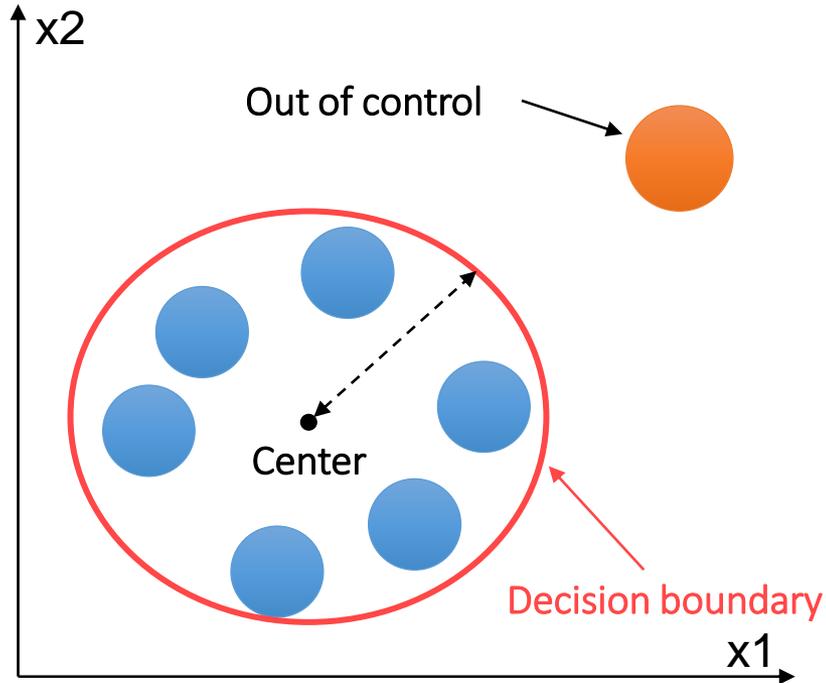
- Optimal False Alarm Controlled Support Vector Data Description for Multivariate Process Monitoring, *Journal of Process Control*
- Multitask Learning for Virtual Metrology in Semiconductor Manufacturing Systems, *Computers & Industrial Engineering*
- Intelligent Traffic Control for Autonomous Vehicle Systems Based on Machine Learning, *Expert Systems with Applications*

- Analysis of Bacterial FAMES Using Gas Chromatography – Vacuum Ultraviolet Spectroscopy for the Identification and Discrimination of Bacteria, *Talanta*
- An Ensemble Regularization Method for Feature Selection in Mass Spectral Fingerprints, *Chemometrics and Intelligent Laboratory Systems*

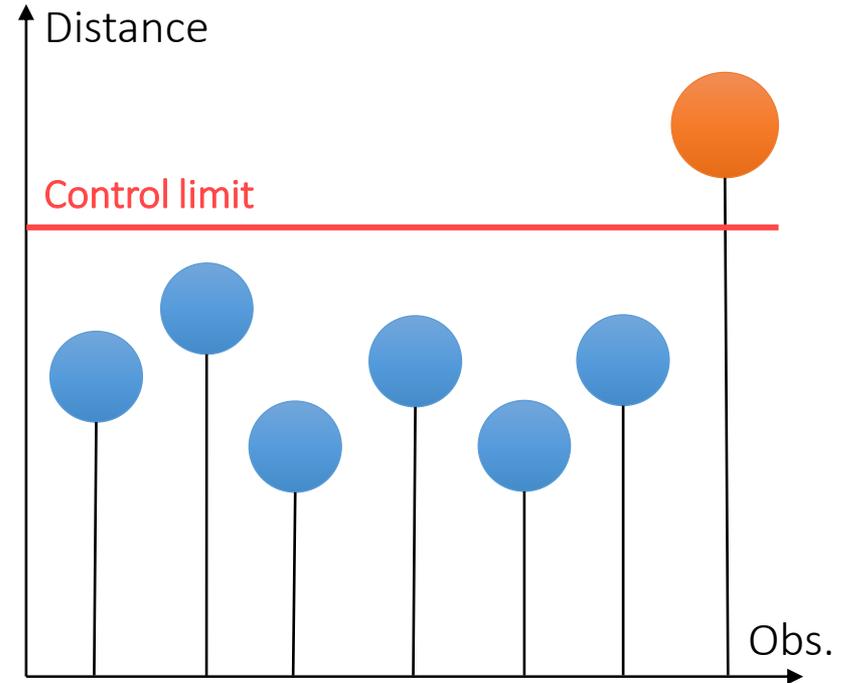


Optimal False Alarm-Controlled Data Description

One-Class Classification Plot



Control Chart



Reference : Sukchotrat, Thuntee, Seung Bum Kim, and Fugee Tsung. (2009). One-class classification-based control charts for multivariate process monitoring, *IIE Transactions*, 42, 107-120.

Optimal False Alarm-Controlled Data Description

Support Vector Data Description

$$\min_{R, a, \xi} R^2 + C \sum_i \xi_i$$

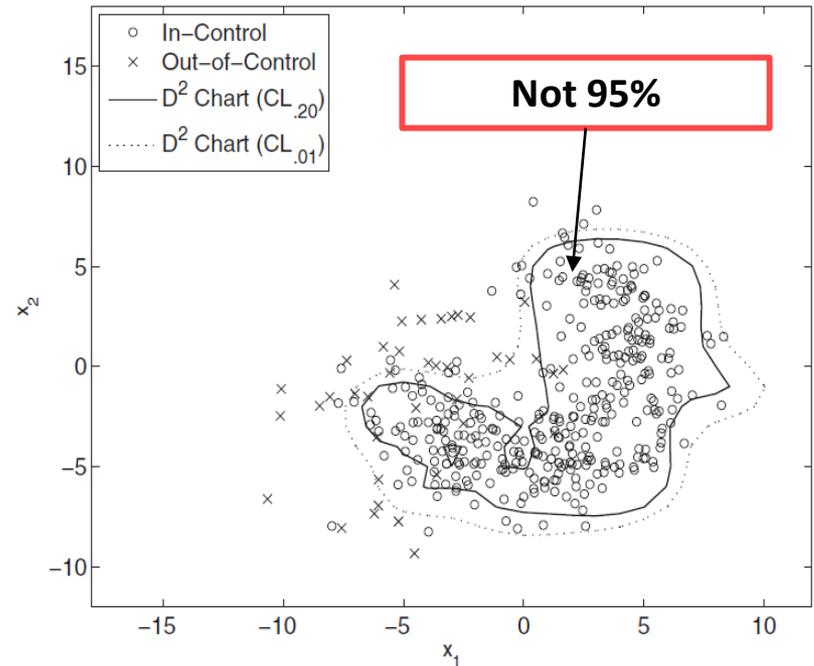
Subject to

$$\begin{aligned} \|\varphi(X_i) - \varphi(a)\|^2 &\leq R^2 + \xi_i \\ \xi_i &\geq 0, i = 1, 2, \dots, n \end{aligned}$$

Limitations

- Continuous variable
- Convex space assumption of decision variable

Limitation of SVDD-OCC



Reference : Chang, W. C., Lee, C. P., & Lin, C. J. (2013). A revisit to support vector data description (SVDD). *Technical report*, National Taiwan University.

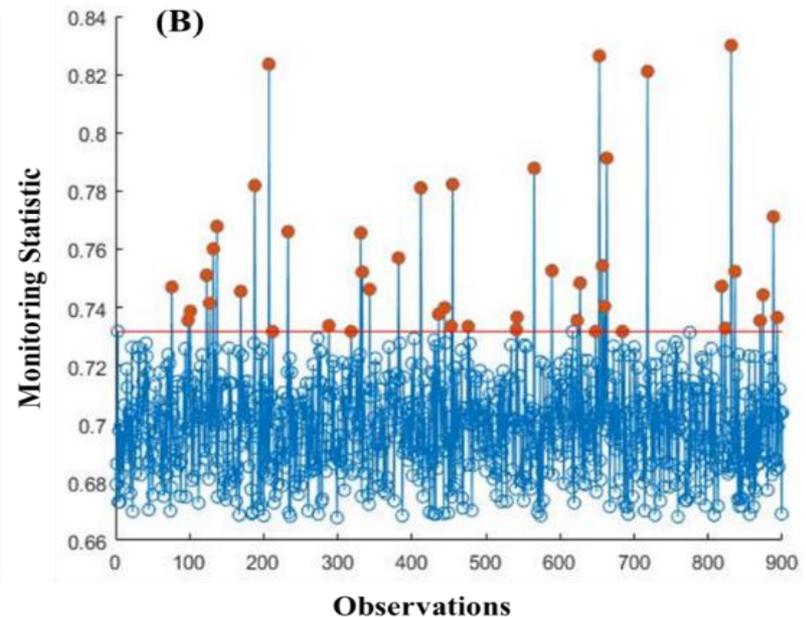
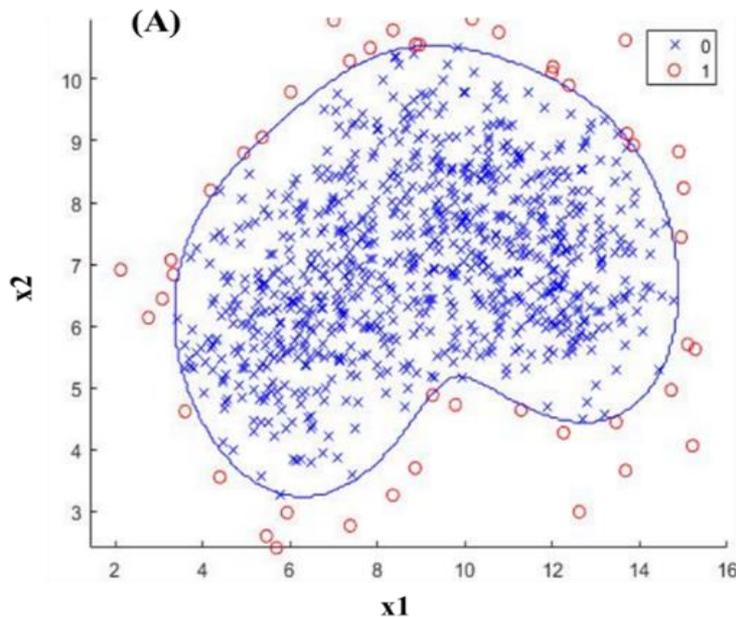
Optimal False Alarm-Controlled Data Description

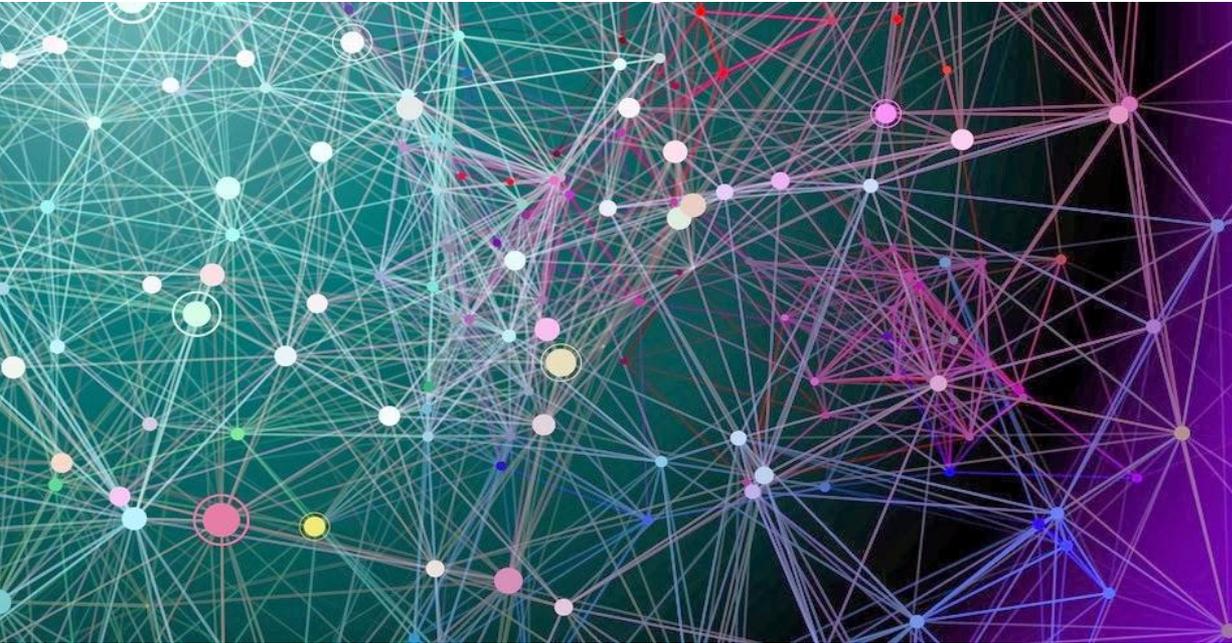
Minimize R^2

Subject to

$$\|\boldsymbol{\varphi}(X_i) - \boldsymbol{\varphi}(C)\|_2^2 \leq R^2 + MZ_i \quad i \in \{1, 2, 3, \dots, N\}$$

$$\sum_{i=1}^N Z_i = \alpha \cdot N$$





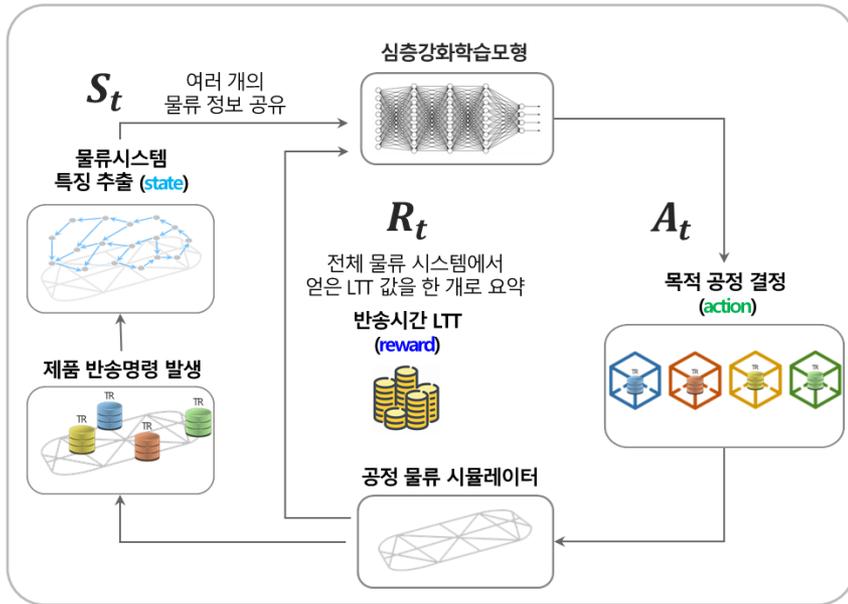
Industrial Projects

Projects



Samsung Electronics

Material Logistics Automation with Reinforcement Learning

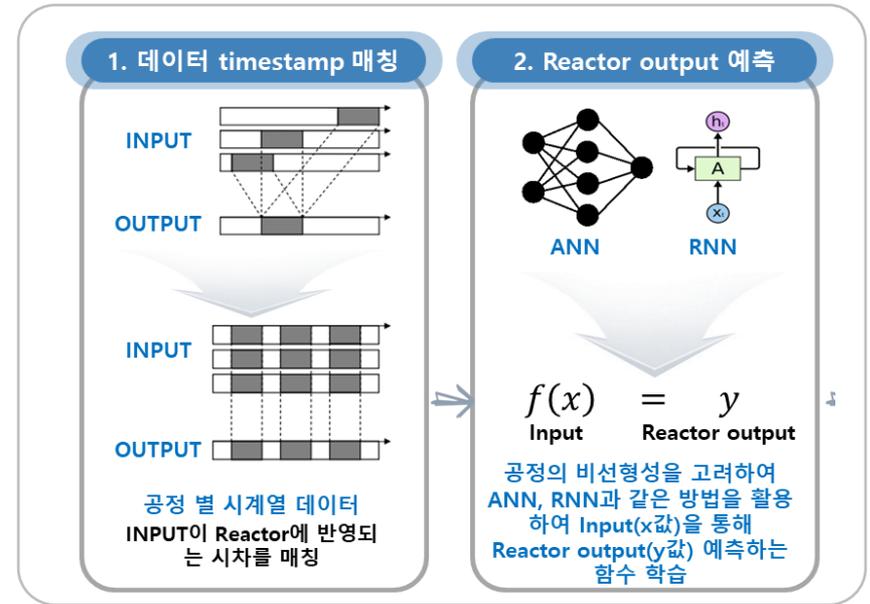


- Reducing the lead time in material logistics of semiconductor manufacturing production
- Applying reinforcement learning methods to the system for dynamic dispatching system



SK Nexlene

Autonomous Chemical Production Control with Deep Learning

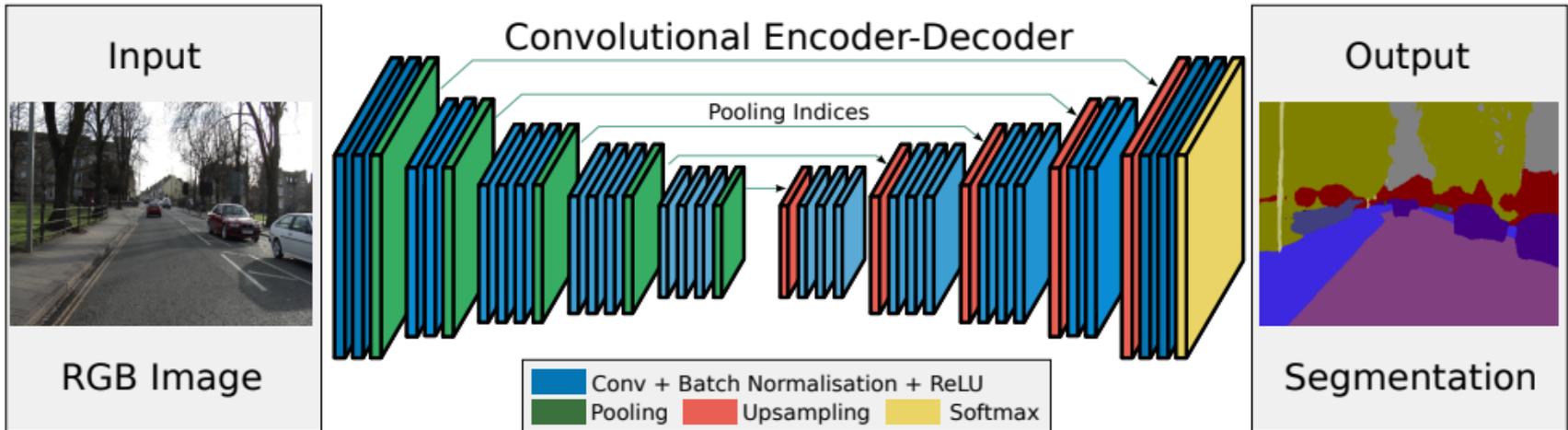


- Improving the productivity of chemical compound with deep neural network and optimization
- Predictive modeling with deep neural network and input optimization via mathematical optimization



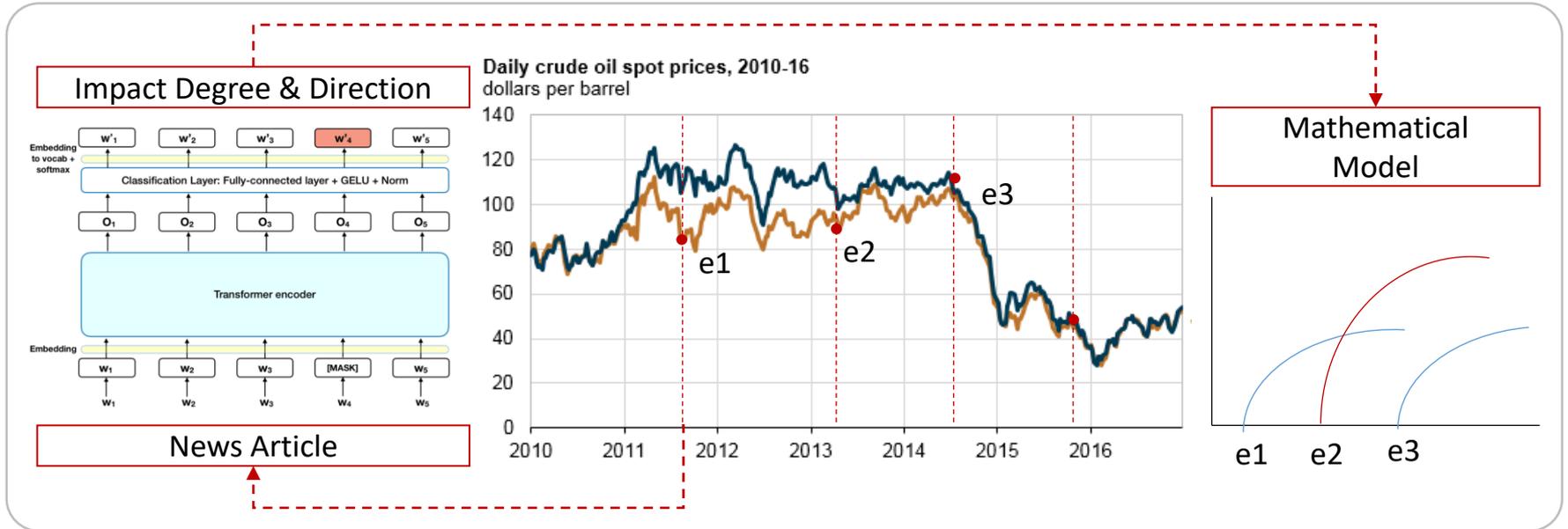
Road Obstacles Detecting with Deep Neural Network

- ❖ Segnet: Fully Convolutional Neural Network, Pooling Index Sharing for Computational Performance
- ❖ Deeplab v3+: Depthwise separable convolution, Astrous convolution, Spatial pyramid pooling
- ❖ Detection: We labeled each pixel of images and applied the segmentation methods-> MIoU 0.7
- ❖ Future works: enhancing the computational performance for real time application





International Oil Price Analysis based on News Events and Financial Factors



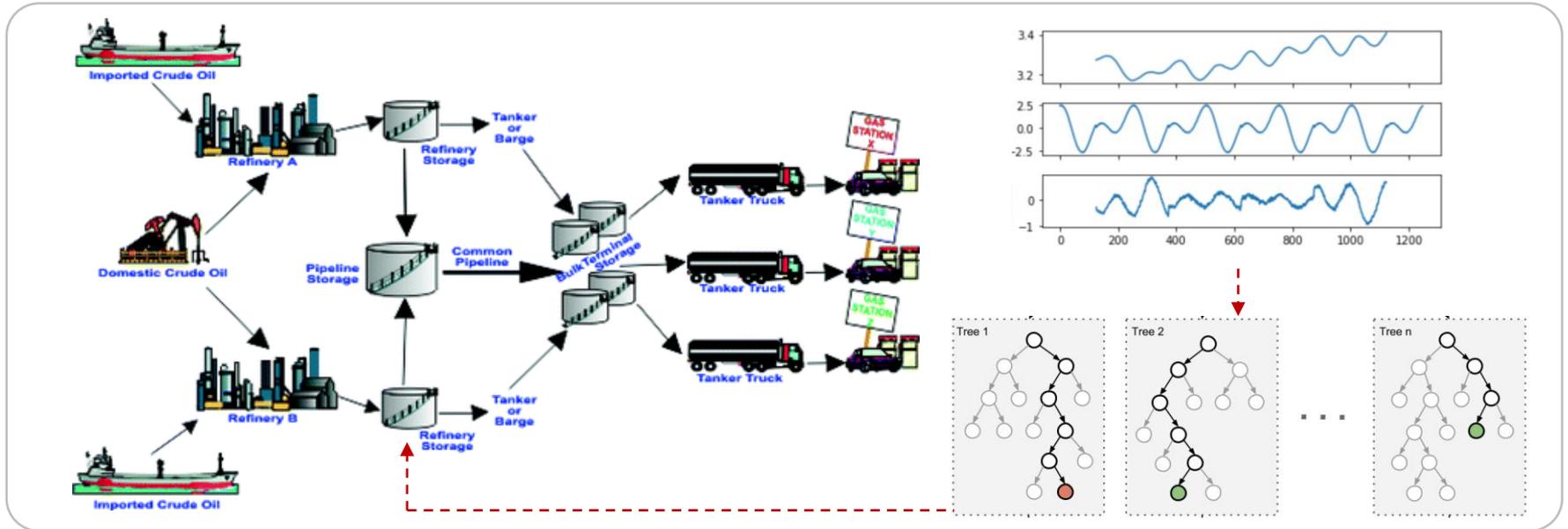
- We use the deep learning-based natural language transformer to classify impact degree and direction from Reuters news articles.
- We try to calculate the impact and sustaining time of each event with mathematical model. The optimization model assumes the decaying impact of the events.
- We try to reflect the financial factors with the mathematical model.

Projects



SK innovation

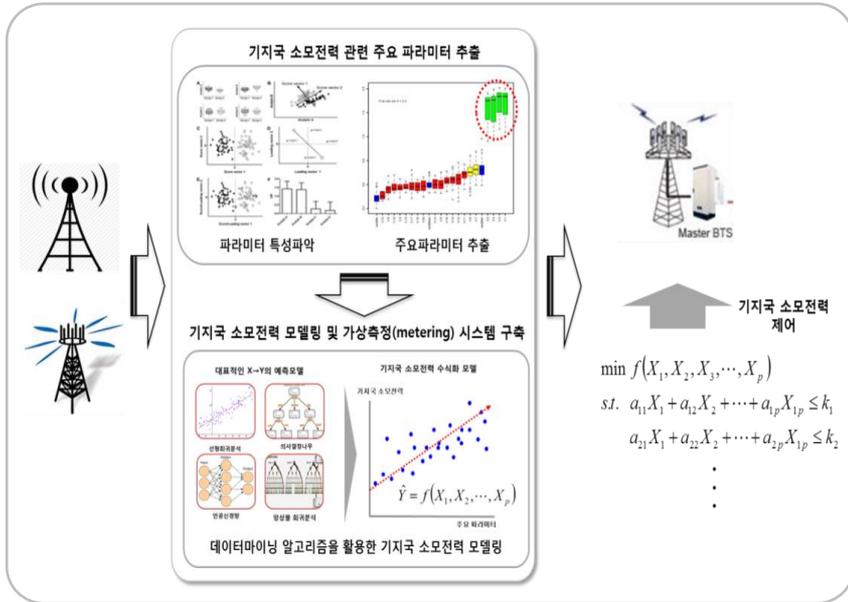
Demand Forecasting and Stock Optimization for Domestic Oil Products



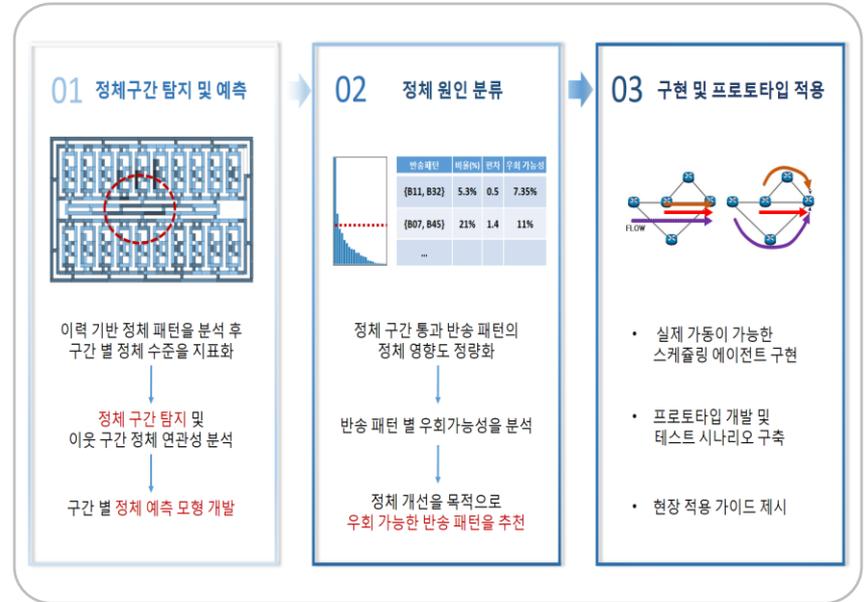
- Forecasting demand of domestic oil products with machine learning algorithm.
- We transform the time series data into multivariate data with modifications and apply the gradient boosting method for robust prediction performance.
- To enhance the performance, we utilize some price information like international oil price or domestic price for each oil product.

Projects

Samsung Electronics Demand Forecasting for Base Stations



Samsung Electronics Optimizing Semiconductor Material Logistics



- Predicting the energy consumption of base stations
- Random Forest, Boosting Tree, Lasso, Elastic Net
- Adaptive training for time-varying problem
- The prediction results satisfied 90% accuracy criteria

- Congestion prediction and optimize material route
- Neural Network, Random Forest, Boosting Tree, K-NN
- Optimizing the material logistic path considering the Predicted congestion



Future Plans

Future Plans

Short-term Plans

Writing

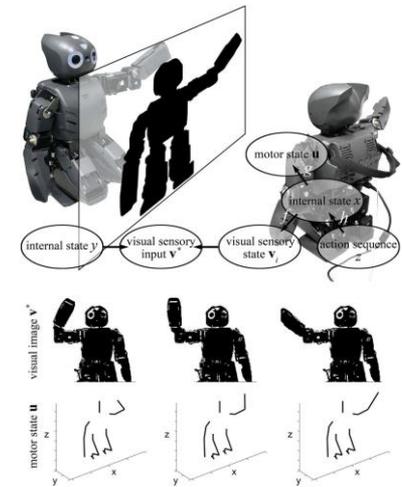
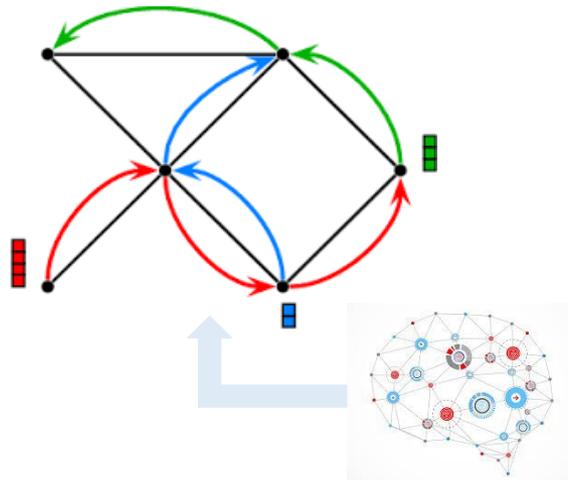
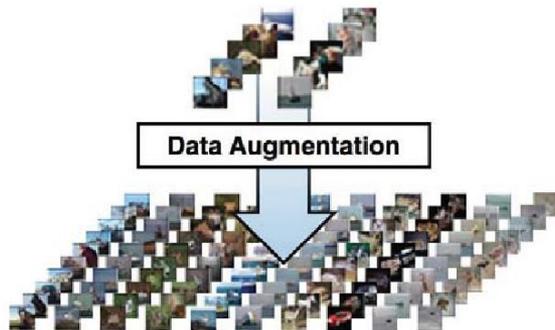
- Multiple Instance Neural Network for Cancer Diagnosis with Sparse Attention
- Chemical Manufacturing Automation with Attention-based Recurrent Neural Network and Optimization
- Autonomous Fork Lift Control with Deep Neural Network
- Liver Cancer Diagnosis using Neural Network with Lifestyle Features

Ongoing

- Robust Graph Convolutional Neural Network with Graph Data Augmentation
- Data Augmentation for Glaucoma Diagnosis with Deep Neural Network
- Online Loss Prediction and Hard Example Mining for Robust Neural Network
- Dynamically Regularized Label Propagation for Robust Semi-supervised Learning

Long-term Plans

- Robust Deep Learning with Data Augmentation for Real-world Applications
- Discrete Optimization with Deep Learning for Efficient Decision Making
- Industrial Production Control with Data-driven Modeling



thank
you

