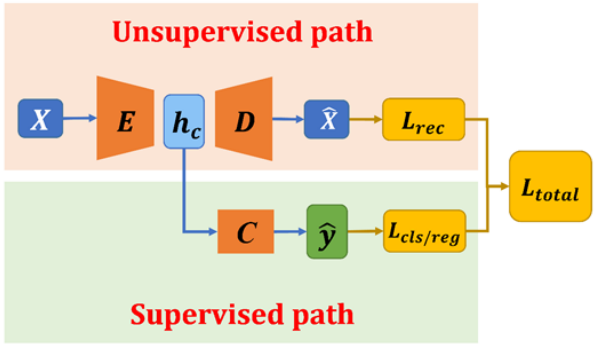
Extraction of Classification/Regression-qualified and Explainable Features for Deep Classifier/Regressor Modeling

Yun Chu & Jakey Blue*

Institute of Industrial Engineering, National Taiwan University
*jakeyblue@ntu.edu.tw

Abstract:

With the continuous innovation in science and technology, data analysis has brought about a wave of paradigm shifts in data volume, computing technology, and modeling methods. In the face of diverse sources and complex structure data exploration technologies, feature engineering is particularly essential. The current mainstream of dimension reduction technologies is mostly unsupervised learning to extract the features that can retain the maximum information within the data. However, the features are not entirely suitable for constructing classification/regression models.



As can be seen in the figure above, this research is motivated to design a framework that integrates the convolutional autoencoder with a deep classification/regression model. Through flattening the hidden feature layer of the convolutional autoencoder to become the input to the classifier and regressor, the two losses are integrated for simultaneous training. The goal is to extract the classification and regression-qualified features as well as to retain the capability of reconstructing the original data. Two data sets: Fashion MNIST and the data from a Chemical Mechanical Polishing (CMP) process are used to validate the proposed method. The feature importance is visualized to identify the key variables that affect the model so as to keep the predictability and explainability in one model. The balance between the two abilities will surely enhance the applicability of deep neural network technology.

Keywords:

Convolutional Neural Networks (CNN), Convolutional Autoencoder (CAE), Feature Extraction, Classifier/Regressor Model, Machine/Deep Learning, Explainable AI (XAI).