**MDNN: A Multimodal Deep Neural Network for Predicting Drug-Drug Interaction Events**

The interaction of multiple drugs might lead to many serious events, which results in high medical costs and may sometimes cause injuries. Accurate Drug-Drug Interaction (DDI) predictions can assist the clinicians make effective decisions and many AI based techniques have proposed to predict the DDI associated events. Most of the existing methods do not pay more attention towards the positive correlations between the DDI events, targets, enzymes and other multimodal data. Also, cross-modality complementarity of multimodal data has not been considered. This paper proposes the Multimodal Deep Neural Network to address this problem. The DDI event prediction is formulated as a multi-class classification problem.

In the Multimodal Deep Neural Network, a two-pathway framework is being designed to obtain the drug multimodal representation, which includes drug knowledge graph (DKG) based pathway and heterogeneous feature (HF) based pathway. They provide complementary information to each other. The DKG-based pathway employs a Graph Neural Network (GNN) to extract topological structure information and semantic relationships between drugs from the drug knowledge graph. The semantic feature score between the drug and tail entity is computed. Messages propagated from the neighborhood is aggregated to refine the drug embeddings. In the final step, the drug embedding, and the neighborhood embedding is aggregated. Heterogeneous features are used to calculate the drug similarity using the DDI events. The Jaccard similarity measurement is used to calculate the similarity of the pairwise drug using feature vectors. The HF-based pathway extracts all the predictive information to enhance the performance of the learned models from different modalities. To explore the complementary among drug multimodal representation, a multimodal fusion neural layer is designed. The embedding of the multimodal fusion with multiple fully connected layers is used for the DDI event prediction. Activation function used is softmax. To accelerate the convergence, batch normalization layers are added, and dropout layers and L2 regularization and early stopping strategies are added to enhance the generalization ability and avoid over-fitting. This is done for the optimization of the model.

The main contribution of this paper is that the MDNN mines the inter-modality similarities from many sources and learns representations from multimodal data. The model is quite stable and with a drug knowledge graph, MDNN makes use of topological structural data and semantic relations. The embedding propagation layer is used to exploit the first-order connectivity information for drug representations. When multi-task analysis is done on the model, experimental results is better than the results of the previously proposed methods. This proves that structural information and heterogenous features can be utilized to improve the accuracy of the prediction of the DDI events. However, the model cannot incorporate all the structural information completely if the size of the sampled neighbor is less. In cases where the size of the sampled neighbor is more, the model is prone to be affected by noise. When large number of GNN layers is used, it decreases the model performance because large amount of noise is being brought into the model. When the embedding dimension is too large, the model will experience over-fitting.

**DOC3 - Deep One Class Classification using Contradictions**

Research using deep learning for one class problems has increased dramatically as a result of the recent success of deep learning-based techniques for various machine learning problems. But most of these works use inductive learning. Because of this, the underlying models perform poorly and are data-hungry when there is a limited availability of data. For problems with limited training data availability, learning from contradictions paradigm has been effective, but it is mostly limited to multi-class or binary class problems. In this paper, ‘Learning from contradictions’ for one class problems has been introduced. The Deep One Class Classification is proposed using the contradictions of DOC algorithm. The generalization performance of one class formulations is analyzed under inductive and universum settings using Rademacher complexity-based bounds.

A different learning paradigm is adopted by the universum learning. The main goal of ‘learning from contradictions’ is minimizing the generalization error on the test data having both normal and abnormal samples. The DOC3 algorithm searches for a solution where the margin errors are reduced, correlation and the generalization error are minimized. The two hyperparameters introduced in the DOC3 algorithm is CU (maximizing the contradiction on Universum samples) and ∆ (insensitive loss). The model selection here is simplified by setting the ∆ (insensitive loss) to 0 and optimally tuning the CU. During the training process, the universum samples are also provided in addition to the samples from normal class. It mainly focuses on minimizing the generalization error while maximizing the contradictions on the universum samples. The generalization error can be improved using DOC to learn under the universum settings. Empirical Rademacher Complexity is used as the capacity measure of the hypothesis class. It is seen that the unseen anomalous samples are contradicted by the universum samples. The DOC3 searches for the solution where the Σ(∞) is low between the training and universum samples. This ensures that the ERC is lower and thus the generalization is improved compared to the DOC. Also, the universum setting controls the hypothesis class complexity. This is done by constraining the spaces in which both the normal and anomalous samples lie. For the once class hinge loss, the maximum contradiction on the universum samples can be achieved when the universum samples lie on the decision boundary. This motivates the one class loss using contradictions by introducing an insensitive loss to relax the constraint. But the main limitation in this is the handling of the absolute term in LU. In this paper, the LU is simplified by rewriting it as a sum of 2 hinge functions. This is done by creating two artificial samples for every universum sample. This results in the universum loss being equal to the sum of two hinge functions with the margins.

The DOC3 provides an improved generalization over the DOC. This is done by deriving the Empirical Rademacher Complexity (ERC). Also, the DOC3 provides a solution where there is a significant reduction in the correlation between the training and universum samples. However, the effectiveness of the DOC3 algorithm depends mainly on the type of the universum being used. Also, there is a need to improve the understanding of the “good” universe sample. The success of the DOC3 model mainly depends on carefully tuning the hyperparameters. Hence, there is a need for better mechanisms to developed for model selection to yield optimal models.