Mathematical Modeling and Algorithm Application
ISSN: 3006-0842 | Vol. 6, No. 1, 2025

Study of Drug Attribute Prediction Model Based on the Neural Network of Divided Attention Mechanism Graph

Haicheng Gao, Jingyi Zheng, Xinyu Song, Bingyu Li

School of Information Engineering, Yancheng Teachers University, Yancheng 224000, China

Abstract: Accurate drug molecular classification is essential for understanding biological activity and guiding clinical applications in drug discovery and development. Traditional methods such as Support Vector Machines (SVM) have demonstrated some success, yet they often struggle with graph-structured data, particularly when confronted with sparse node features and redundant information. Despite significant advances in topology-based drug attribute prediction and machine learning, classification accuracy and efficiency remain critical challenges. Traditional models often fail to capture the intricate relationships within molecular graphs, especially when it comes to feature extraction and information processing. In this study, we focus on enhancing drug molecular classification using graph-based models. This paper begins by preprocessing a graph dataset and manually extracting feature vectors. Dimensionality reduction with t-SNE is then applied to visualize the molecular distributions. A baseline binary classification model using SVM with a Gaussian kernel is implemented, achieving an accuracy of 69.67%. However, this performance is suboptimal. To address this, we introduce an end-to-end Graph Neural Network (GNN) model, incorporating a basic GNN layer, a pooling layer, and a classifier. This model improves accuracy to 89.19%. To further tackle challenges such as node feature sparsity and information redundancy, we augment the model with a frequency division attention mechanism. Additionally, we expand the dataset using Graph Generative Adversarial Networks (GraphGANs), increasing the sample size to 940. The resulting model achieves an accuracy of 92.41%, demonstrating significant performance improvements. Finally, we conduct robustness and sensitivity tests to thoroughly evaluate the models strength's and limitations. Our results reveal the effectiveness of the proposed approach in overcoming key challenges in drug molecular classification and emphasize the potential of GNNs in advancing computational drug discovery.

Keywords: Support Vector Machine, Graph Neural Network Frequency Division Attention Mechanism GraphGANs

1. Introduction

This study presents a novel approach to drug molecular classification using Graph Neural Networks (GNNs). After initially applying traditional SVM with limited success (69.67% accuracy), we developed an end-to-end GNN model, improving accuracy to 89.19%. To address challenges like node feature sparsity and information redundancy, we incorporated a frequency division attention mechanism and augmented the dataset using GraphGANs, achieving a significant performance boost to 92.41%. Our results highlight the potential of GNNs in enhancing drug molecular classification and offer valuable insights for future applications in drug discovery.

2. Study of drug attribute prediction model based on the neural network of divided attention mechanism graph

2.1. Data preprocessing

The data is first loaded from the provided MMM_data dataset in MATLAB format. This dataset contains information for 188 compound molecules, where each molecule is represented as a graph structure. In this structure, denotes the adjacency matrix, represents the node labels may include additional auxiliary information. This information could include topological features of the graph (e.g., node degree, aggregation coefficient) or chemical properties (e.g., molecular fingerprints)[1]. To clarify, each molecule is labeled as either mutagenic aromatic or heteroaromatic. The feature vector $X_i$each molecule is extracted from its auxiliary information.

$$X_i = [x_1, x_2, \cdots, x_n] \tag{1}$$

In order to preliminarily observe the distribution of graph data in space, the t-SNE data embedding scheme is adopted to reduce the dimension of the original data[2]. The t-SNE pairwise differentiated all data points, effectively minimizing the joint distribution of high and low dimensional data. The initial distance relationship between data points is reproduced during the data embedding and thus in the transition to low dimensions.

$X_i$is known to be the feature vector of the molecule, and the ultimate goal is to estimate a two-dimensional $y = [y_{i,1}, y_{i,2}]$, In this way, the pairwise distances are preserved. These two new dimensions are completely different from any initial dimension and are defined as representing data points in low dimensions in order to retain the key similarities or dissimilarities in the high-dimensional space.

$$p_{i,j} = \frac{p_{i|j} + p_{j|i}}{2N_D} \tag{2}$$

where, $p_{i|j}$ represents the probability of the pairwise distance in the original space. The goal is to extract the $\mathrm{x_i}$ and $x_i \cdot d(x_i, x_j)^2$ from each closest pair $p_{i,j}$. The distance relationship between the data points xi and $x_i \cdot d(x_i, x_j)^2$ is expressed by the conditional probability pi|j. This relationship is modeled with the t-student distribution as follows:

$$p_{i,j} = \frac{e^{-d(x_i,x_j)^2/2{\sigma_i}^2}}{\sum_{k \neq i} e^{-d(x_i,x_k)^2/2{\sigma_k}^2}} \tag{3}$$

where, σ_{i} represents a Gaussian distributed bandwidth centered on $x_i \cdot d(x_i, x_j)^2$ Denotes the Euclidean distance

between two data points x_i and x_j in the original feature space. The q_i, q_j represents the probability of a pairwise distance in a low-dimensional space, calculated as follows:

$$q_{i,j} = \frac{(1 + ||y_i - y_j||^2)^{-1}}{\sum k \neq l (1 + ||y_k - y_l||^2)^{-1}} \quad (4)$$

The values of all single $q_{i,j}$ constitute the set of low-dimensional joint probability distributions Q. The values of all single $p_{i,j}$ constitute the set of low-dimensional joint probability distributions P. The Kullback-Leibler divergence (DKL) model was used to calculate the loss of information when using P to estimate Q. DKL was optimized to minimize differences between Q and P, by Figure 1.

$$D_{KL}(P|Q) = \sum i,j \, (p_{i.j} \times log^{\frac{p_{i,j}}{q_{i,j}}}) \quad (5)$$

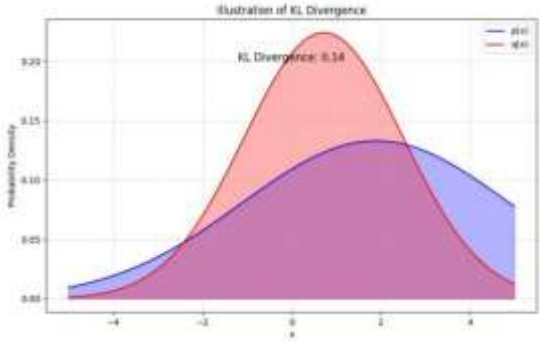

Figure 1 Characterization of molecules

In this paper, gradient descent was used to optimize DKL, obtained from Figure 2.

$$\frac{\partial D_{KL}}{\partial y_i} = 4 \sum j(p_{i,j} - q_{i,j}) (y_i - y_j)(1 + ||y_i - y_j||^2)^{-1} \quad (6)$$

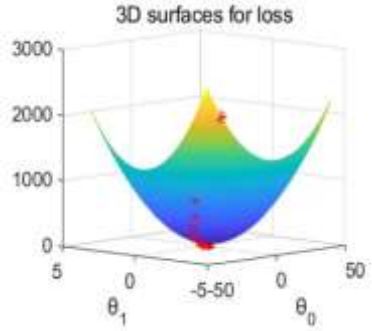

Figure 2 gradient descent

Finally, the distribution of the molecules is shown in Figure 3.

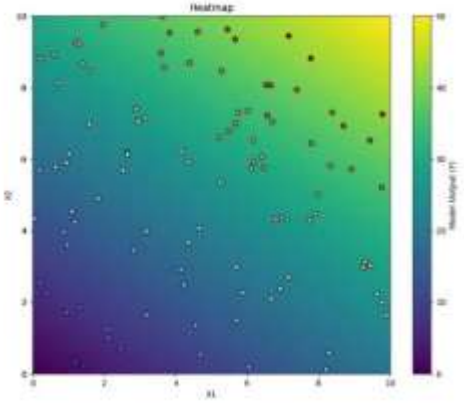

Figure 3 Molecular distribution

By the Figure3 We know that the distribution of molecules in space can be roughly divided into two categories in two dimensions, so this paper adopts a support vector machine-based scheme to solve the graph data by two classification.

2.2. Support vector machine

This paper first defines the optimization problem of SVM. For dichotomy problems, the SVM aims to find a maximum-spaced hyperplane [3], Making the two classes of samples as separate as possible. hypothesis $y_i \in \{-1,1\}$, Show the category label of the i-tb molecule, w is the weight vector, b is the bias term, and the optimization problem of SVM can be expressed as:

$$min_{w,b} \frac{1}{2} ||w||^2 \quad (7)$$

$$y_i(w^T X_i + b) \geq 1 \quad (8)$$

Next, choose the kernel function. Because the graph data is not linearly separable, so a kernel function K can be selected to map the data to a high-dimensional space, so that the data is linearly separable in that space [4]. The commonly used kernel functions are linear kernel K (x_i , x_j) = x_i x_j , polynomial kernel K (x_i , x_j) = (λ x_i , x_j + c) d, and the Gaussian kernel is used here.

$$K_{(X_i,X_j)} = e^{-\lambda ||X_i - X_j||^2} \quad (9)$$

Finally, the SMO algorithm is used to train the model, since the original problem of SVM is to find a maximally interval hyperplane, by Figure 4 Note that its dual form is usually optimized as:

$$max_\alpha \sum_{i=1}^{n} \sum_{i=1}^{n} \alpha_i \, \alpha_j y_i y_j x_i{}^T x_j \quad (10)$$

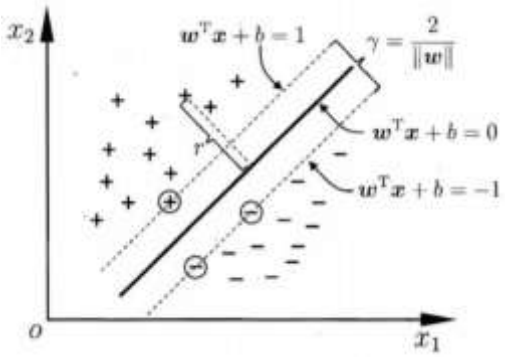

Figure 4 SVM, hyperplane

Therefore, in each iteration, only two variables α_i and α_j are selected, while the other $\alpha_k$are kept unchanged, so that the original problem is reduced to a sub problem of two variables, which can be solved analytically.

This paper uses a heuristic selection strategy to select the two variables that violate the most serious KKT condition, or the two variables that cause the largest decrease in the objective function. The specific strategy can be to select a α i that violates the KKT condition and then find another α_i so that both α_i and α_j can maximize the objective function.

$$\alpha_i y_i + \alpha_j y_j = -\sum k \neq i,j \, \alpha_k y_k = const \quad (11)$$

For a new molecular Xi, its classification decision rule is:

$$f(X_{new}) = sign\left(\sum_{i=1}^{N} \alpha_i y_i K(X_i, X_{new}) + b\right) \quad (12)$$

where, N: the training sample; α_i: a Lagrange multiplier that satisfies the KKT condition.

2.3. Solution based on the SVM model

The preprocessed molecular graph data were input into the model for processing. First, the original dataset was split following an 8:2 ratio, meaning that a total of 150 molecules were divided into the training set, with 38 molecules assigned to the test set. Next, the model's training loss and test accuracy were visualized, as shown in Figure 5. It can be observed that the model's accuracy for this problem is around 69%, indicating relatively poor performance.

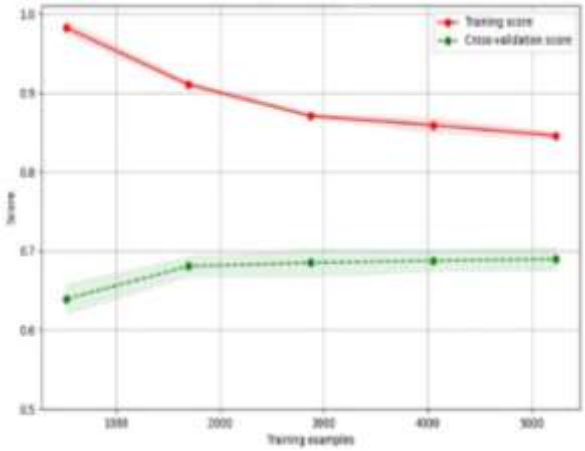

Figure 5 Model accuracy

SVM's poor performance on graph data may stem from suboptimal feature selection for drug molecules, failing to capture key relationships between molecular structure and activity or distinguish mutagenic aromatic and heteroaromatic compounds. Additionally, SVM relies heavily on the kernel function, and for high-dimensional, nonlinear graph data, even powerful kernels like Gaussian may struggle to make the data linearly separable. SVM is also sensitive to parameters like regularization and kernel choice, making it difficult to find the optimal configuration for best results[5].

Based on the above analysis, the deep learning related model with stronger feature extraction ability and relatively few hyperparameters is considered to further address this problem.

2.4. Solution based on the SVM model

In order to adapt the data to the model, this paper first treats each molecule as a graph, with atoms as nodes and chemical bonds as edges. The adjacency matrix provides the structural information of the graph, and each node (atom) may also have the chemical attribute information, and here the feature vector encoding form Xi.

By the Figure 6, It appears that the graph neural network model construction includes the underlying GNN layer [6], Pooling the layer, and the classifier.

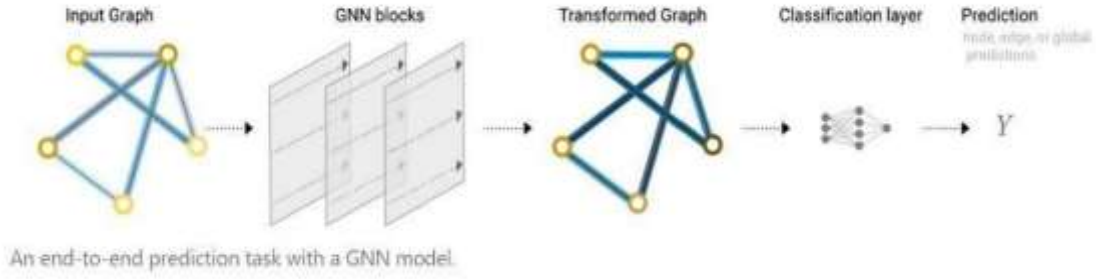

Figure 6 GNN architecture

First, build the underlying GNN layer,

$$h_i^{(l)} = \sigma\left(W^{(l)} h_i^{(l-1)} + \sum_{j \in N_i} A_{ij}^{(l)} h_i^{(l-1)}\right) \tag{13}$$

where h_i (l-l) is the hidden state (feature vector) of node i in layer l; W (l) is the weight matrix of the layer;$\sigma$is the activation function; Ni is the neighbors of nodes i; is the attention weight for the adaptive learning of the adjacency matrix, which is used to measure the influence between nodes.

To extract a fixed-size representation from the entire graph, a pooling layer can be used, with maximum pooling or average pooling at the figure level

$$h_G = POOLING\left(\left\{h_i^{(L)} | v_i \in G\right\}\right) \tag{14}$$

HG: the representation of Figure G; L is the number of layers of the GNN.

Finally, the classifier is built: input the graph-level feature vector h_G into the classifier (such as the fully connected layer) to predict the molecular categories:

$$\hat{y} = soft\,max(Uh_G + b) \tag{15}$$

2.5. AGNN model based on the split-frequency attention mechanism

This paper breaks through the existing graph neural network model in dealing with node features sparsity and information redundancy graph structure of the challenge, and the GNN fusion Attention mechanism, can give different nodes and edges of different attention weight, can enhance the ability to capture key information, at the same time reduce the dependence on redundant information, so as to improve the application of the model in complex network analysis.

The GNN model is based on a frequency-sharing attention mechanism, where neighbor information aggregation for each node is performed through weighted summation. The weights are determined by the frequency-sharing attention mechanism, enabling the model to adjust the importance of neighboring nodes based on their characteristics.

$$e_{ij} = Leaky\ Re\,L\,U\left(a\left[|h_i||h_j|\right]\right) \tag{16}$$

where a: is a learnable weight vector; vector splicing operation; LeakyReLU: is a nonlinear activation function used to increase the expression power of the model.

To ensure that the sum of the attention weights for all the neighbor nodes is 1, Need to normalize these weights. Normalization was performed using the softmax function:

$$\alpha_{ij} = \frac{exp\left(e_{ij}\right)}{\sum k \in N_i\, exp(e_{ik})} \tag{17}$$

where Ni: represents the set of neighbors of node i; αij: represents the attention weight after normalization

By applying attention weights, the weighted features of neighboring nodes are used to derive the new feature representation h for node i.

$$h_i^{'} = \sigma\left(\sum_{j \in N_i} W h_j\right) \tag{18}$$

where σ: is an activation function; W: is a learnable weight matrix for transforming node features.

To enhance the expression ability and stability of the model, the model finally uses the attention mechanism.

$$h^{'} = ||_{k=1}^{K} \sigma\left(\sum_{j \in N_i} \alpha_{ij}^{k} W^{k} h_j\right) \tag{19}$$

So far, the GNN architecture based on the split-frequency attention mechanism is completed.

3. Results

3.1. Solution of the GNN model based on the split-frequency attention mechanism

The dataset was divided into training and test sets at an 8:2 ratio. Specifically, out of the total 940 molecular samples, 752 molecules were allocated to the training set, and the remaining 188 molecules comprised the test set. The data was then fed into the frequency-partitioned attention-based GNN model (constructed in prior steps) for graph data classification. Finally, the classification accuracies of the three models were compared, as summarized in Table 1.

The accuracy of the three models is finally compared, as shown in Table 1.

Table 1 Data dimension reduction model comparison

Model	Accuracy	Sensitivity	Specificity
SVM	69.67	68.92	67.25
GNN	89.19	85.76	89.19
Attention—GNN	92.41	93.87	93.64

By the table 1, The comparative analysis shows that compared with the traditional method SVM, the GNN model has significantly improved the classification accuracy, and also significantly improved the generality of the model. Among them, Sensitivity and Specificity, two indicators used to evaluate the effect of the model for true and false classes, also from68% Up to89%。

Considering the challenges faced by the existing graph neural network model in dealing with the graph structure data with node feature sparsity and information redundancy, the classification effect of the model can be improved after establishing the Attention GNN model. Among them, the accuracy from 89% promote 92%, And the generality is more stable, and the model is more robust.

4. Conclusions

In this study, the problem of drug molecular classification was deeply studied by applying data preprocessing, support vector machine (SVM), graph neural network (GNN), and GNN based on GraphGANs data expansion and GNN of frequency division attention mechanism. The results show that compared with the traditional SVM method, the GNN model achieves a significant improvement in the classification accuracy and improves the accuracy to90% the left and right sides. However, the GNN model still faces challenges in handling node feature sparsity and information redundancy. Expand the data set through GraphGANs technology to 940. After the molecules, the GNN model combined with the frequency attention mechanism further optimized the classification effect, which was finally realized92.41% accuracy, showing more stable generality and better robustness. These findings not only provide new perspectives in the field of drug molecular classification, but also provide useful references and guidance for the application of graph neural networks in related fields.

References

[1] Cao, Y. Research on Drug Property Prediction Based on Topology and Machine Learning [D]. Wuhan Textile University, 2022.

[2] Xu, Z. Application of t-SNE Algorithm in Dimensionality Reduction of High-Dimensional Data [D]. Xiamen University, 2022.

[3] Yu H, Kim S. SVM Tutorial-Classification, Regression and Ranking[J]. Handbook of Natural computing, 2012, 1: 479-506.

[4] Liang S, Hua Z, Li J. Enhanced self-attention network for remote sensing building change detection[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2023, 16: 4900-4915.

[5] Shen, Y. Research and Application of Support Vector Machine Multi-Classifiers [D]. Jiangnan University, 2019.

[6] Han K, Wang Y, Guo J, et al. Vision gnn: An image is worth graph of nodes[J]. Advances in neural information processing systems, 2022, 35: 8291-8303.