25/11/2024-02/12/2024

WEEK 3

ONE STAGE CONVOLUTIONAL NEURAL NETWORK BASED METHOD

This week, we started researching the use of CNN for outlier detection. A CNN (Convolutional Neural Network) is a type of artificial neural network used to capture general patterns in data. It performs convolution and pooling operations on the inputs to extract significant features from the data and passes these features to a classifier or regressor model. CNNs are trained on normal data. When a new data point is introduced, the CNN analyzes it and checks for deviations from the learned normal data distribution, identifying these deviations as outliers.

metin, diyagram, çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Açıklama otomatik olarak oluşturuldu

After transitioning to working with CNN, we decided to switch from the previously used breast cancer dataset to a dataset focused on heart rhythm data to obtain diverse results. This dataset contains three classes: normal, test and abnormal. The model is trained exclusively on normal data.

After training, the model attempts to make predictions on both normal and abnormal data, as well as test data. It performs better on normal data, as evidenced by a lower mean absolute error (MAE). Predictions on test data also show relatively lower errors compared to abnormal data. However, when making predictions on abnormal data, the MAE is significantly higher. This outcome indicates that the model can accurately learn and reconstruct normal data, as the difference between the actual and predicted values for normal data is minimal.

On test data, the model performs reasonably well but still shows more variation than on normal data. In contrast, the larger error observed with abnormal data suggests that the model is not familiar with these patterns. This characteristic makes the model highly useful in monitoring patients' heart rhythms, as it can effectively detect and flag anomalous data, which may indicate potential health concerns.

metin, ekran görüntüsü, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

After performing the operations on each sample, we obtain three values: the mean absolute error for normal data, the mean absolute error for test data, and the mean absolute error for abnormal data. The relationship between these values is as follows: normal <test <abnormal.

metin, yazı tipi, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu

Since we obtained positive and meaningful results with this dataset, we decided to apply the same procedures to the breast cancer dataset as well.

metin, yazı tipi, ekran görüntüsü içeren bir resim

Açıklama otomatik olarak oluşturuldu

When working with the Breast Cancer dataset, we encountered an issue where the data is not labeled as "normal" and "anomalous" as we expected. Therefore, we are unable to use the CNN method as intended.



For this reason, we understood why combining OCSVM and CNN would yield better results. The OCSVM method will first identify the outliers. Then, we will add the outliers detected by OCSVM to the "abnormal" class in the CNN method. Afterward, we will continue our research using these data.

Referencess:

<https://www.mdpi.com/1424-8220/22/22/8764?type=check_update&version=1>

<https://medium.com/@tuncerergin/convolutional-neural-network-convnet-yada-cnn-nedir-nasil-calisir-97a0f5d34cad>

<https://www.geeksforgeeks.org/introduction-convolution-neural-network/>

<https://medium.com/@deonatan27/cnn-knn-and-svm-analysis-f7679da5358b>