Contribution ID: **1120** Type: **Rapporteured**

In-depth Research on the Interpretable Disruption Predictor in HL-2A

Saturday, May 15, 2021 8:30 AM (17 minutes)

A disruption predictor based on deep learning method is developed in HL-2A. Its structure is specially designed to deal with data from fusion devices, and shows to have a better performance on disruption prediction problem than ordinary models commonly used in computer science[1.]. Based on this deep learning algorithm, in-depth research are carried out in three aspects. Firstly, different structures of the 1.5-D layer are tried in order to find the best way to merge information from different input channels. Secondly, different preprocessing methods are tried to find a better way to deal with the discrete data in the input, e.g., power of NBI heating. These data seems to be harmful to the performance of disruption predictor without a properly preprocessing. Finally, an interpretation method is applied on the disruption predictor, which can be used to give the causes of disruptions along with the disruption alarms.
The model in this research is mainly based on a 1.5-Dimensional Convolutional Neural Network (1.5-D CNN), which is good at dealing with signals from multi-channels with great divergence. The Disruption Predictor uses shots 20000-29999 in HL-2A to train the machine learning model, and uses shots 30000-31999 to optimize hyper parameters. When tested on shots 32000-36000 in HL-2A, it reaches a True Positive Rate (TPR) of 92.2% and a True Negative Rate (TNR) of 97.5% with 30ms before the disruption. A trade-off between TPR and TNR can be realized according to the cost of false alarms and missed alarms during application by changing the alarm threshold, as showed in Fig.1. There is also a trade-off between prediction advance times and accuracy, which is showed in Fig.2.
A comparison is made between the accuracy of four different models to predict the disruptions in HL-2A. And 1.5-D CNN+LSTM shows to perform much better than other traditional CNN-based models. LSTM also shows to perform an important role to achieve better accuracy.
To solve the remaining problems in the algorithm mentioned above, and reach a better performance, in-depth research are carried out in three aspects.
The first problem is how to deal with multi-channels data with great divergence from fusion device. Although the 1.5-D CNN has solved the problem to a certain extent, since information from all channels are merged simply in one layer, this structure may not be able to find some complex multi-channels information. Since that, several different versions of 1.5-D layer are compared to find a better way to merge the information from different channels.
The second problem is about the preprocessing of data. Most of the input channels of disruption predictor are continuous variables, i.e., the distributions of their value locate in a continuous range. But there are some exceptions. For example, the power of NBI heating will always be selected from several individual values, then the sharp rising edge and falling edge will bring troubles to the training of machine learning models. Several different preprocessing method are tried to solve this problem, and the result shows that the model will perform better during the transition stage with properly preprocessing.
Finally, a model interpretation method is applied on the prediction model to decide which of the input channels are more closely related to a certain disruption. And the result shows a good coherence with the disruption causes. For example, the model shows to rely on signals from mirnov probes during MHD instability induced disruptions, and to rely on bolometer signals during radiation induced disruptions. This model interpretation method can be used to automatically give the disruption causes, which will be quite helpful for the active avoidance of disruption. To check the feasibility of this method, several main types of disruption are defined in HL-2A and a simple Bayesian classifier is developed to classify these types of disruptions.

References
[1.] Z. Yang, et al, Nucl. Fusion 60 (2020) 016017

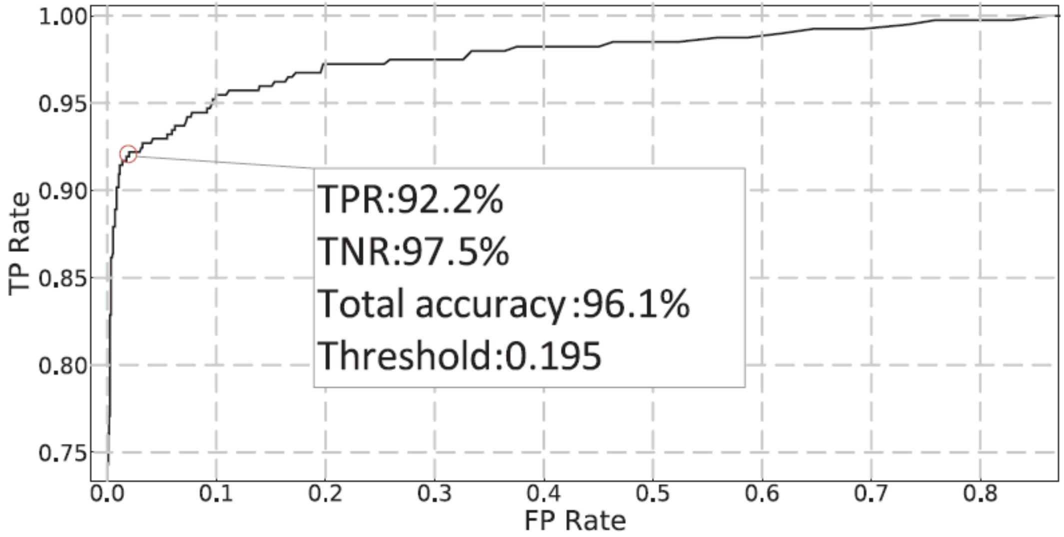

Figure 1: ROC curve of Disruption Predictor on testing set

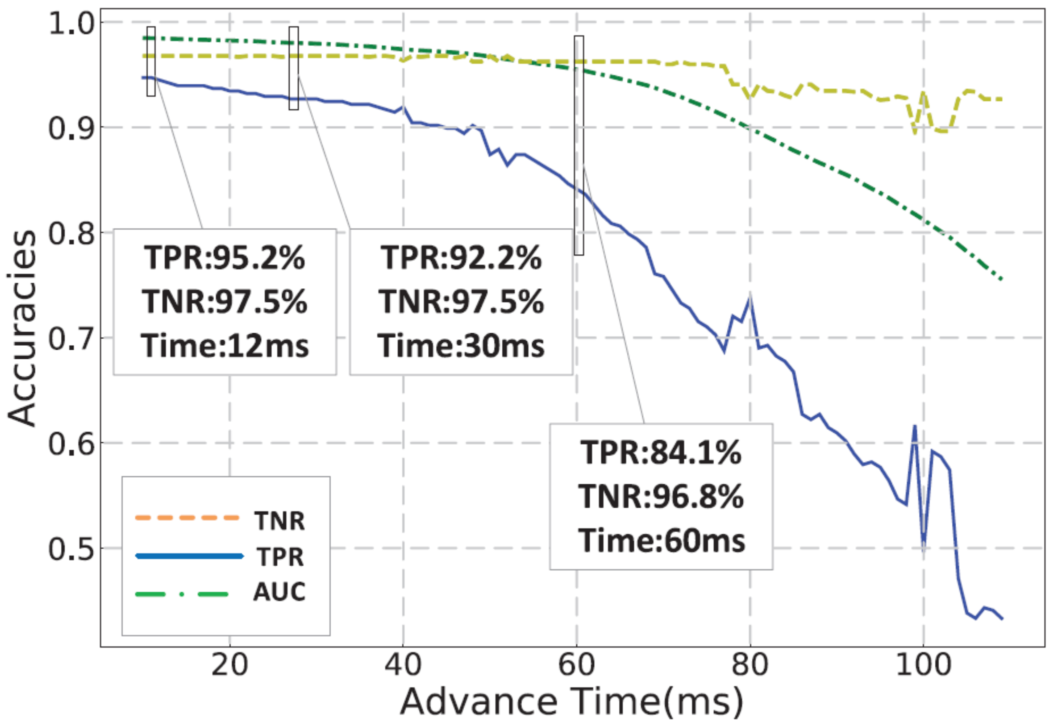

Figure 2: Accuracies reached with different advance times

Affiliation

Southwestern Institute of Physics & Tsinghua University

Country or International Organization

China

Primary authors: Mr YANG, Zongyu; Prof. XIA, Fan (Southwestern Institute of Physics); Dr SONG, Xianming (Southwestern Institute of Physics, Chengdu 610041, People's Republic of China); GAO, Zhe (Tsinghua University); DONG, Yunbo (Southwestern Institute of Physics); HUANG, yao; Dr WANG, Shuo (SWIP)

Presenter: Mr YANG, Zongyu

Session Classification: TH/7 Disruptions, advances in RF modeling, and stellarators

Track Classification: Magnetic Fusion Theory and Modelling