Multilayer Perception Hyperparameter Fine-Tuning for Ionospheric VLF Amplitude Data Exclusion

Filip Arnaut and Aleksandra Kolarski

Institute of Physics Belgrade, University of Belgrade, Pregrevica 118, 11080 Belgrade, Serbia
E-mail: filip.arnaut@ipb.ac.rs

Abstract

The analysis of ionospheric amplitude data is affected by various factors, including the influence of solar flare events, instrument malfunction, and other sources of error. These factors collectively contribute to a decrease in the overall quality of the data. The removal of such data is performed manually by researchers, a process that is characterized by its labor-intensive nature and time-consuming requirements. This research paper presents a procedure for fine-tuning of the learning rate (LR), number of epochs and the momentum in Multilayer Perception (MLP) classification models. The proposed method can be utilized as a benchmark for optimizing other hyperparameters in the future.

Introduction

Very low frequency (VLF) ionospheric amplitude data is adversely affected by solar flare events and instrumental errors, including malfunctions. In order to render the VLF amplitude data applicable for further analyses by researchers, it is imperative to eliminate those effects. The process of manually removing these effects is known to be time-consuming, tedious, and labor-intensive. Consequently, there is an advantage in automating this process using machine learning (ML) classification techniques.

In order to accomplish this task, a range of ML techniques can be utilized, including Artificial Neural Networks (ANN) such as Multilayer Perception (MLP) models, as well as traditional ML models like Random Forests (RF) and Support Vector Machines (SVM). The distinction between MLP models and classic models, such as RF models, becomes evident when considering the quantity of hyperparameters that require configuration. In the case of RF models, the researcher is only required to determine the number of trees. Conversely, MLP models necessitate the fine-tuning of multiple hyperparameters, including the learning rate (LR), momentum, and the number of epochs among others. This study aims to demonstrate the process of fine-tuning the LR, number of epochs and the momentum. This initial step will serve as a starting point in identifying the most suitable MLP model for the intended research objective, which involves the automated identification of erroneous data points in ionospheric VLF amplitude data.

Methods and data

The data employed for this study consists of VLF amplitude measurements obtained during the months of September and October of 2011. These measurements capture solar flare events falling within the range of C2.5 to X2.1 in terms of their class. The training dataset for September 2011 consists of 59,344 datapoints after being balanced as to remove any bias (Prusa et al. 2015; Kulkartni et al. 2020; Devi et al. 2020), while the testing dataset for October 2011 comprises 180,071 data points. The training and testing datasets consist of 40 features, encompassing the original VLF amplitude data, X-ray irradiance data, encoded values of the transmitter and receiver, and the local receiver time expressed in decimal points. Statistical features encompass additional features, including rolling window statistics that employ diverse window lengths (5, 20, and 180 minutes) and varying parameters such as mean, median, and standard deviation. The features utilized in the analysis also consisted of the first and second differential data of both VLF amplitude and X-ray measurements.

The process of searching and optimizing hyperparameters was conducted in a two-step manner. The initial step consisted of determining the optimal combination of the LR and the number of epochs, while maintaining a constant momentum hyperparameter of 0.2. The second phase of the procedure involved determining the optimal momentum value, while keeping the LR and number of epochs constant. In the initial phase, the LR and the number of epochs were subject to variation within the ranges of 0.1 to 0.85 and 100 to 500, respectively. The LR varied by increments of 0.1, while the number of epochs varied by increments of 200. The initial iteration of the modeling process consisted of a total of 18 models, whereas the second stage of modeling involved a total of 9 models. The Weka software package (Frank et al. 2016) included several noteworthy parameters, one of which was the number of nodes. In the case of this MLP, this value was determined by dividing the sum of the number of features and classes by 2, resulting in a total of 22 nodes.

The evaluation of all models involved the utilization of standard ML classification methods, including the F-measure, the area under the receiver operating characteristic curve (ROC), the number of Correctly Classified Instances (CCI), the percentage of Incorrectly Classified Instances (ICI), Cohen's kappa statistics, true positive (TP) and false positive (FP) rates, as well as the Matthew's correlation coefficient (MCC). Due to the inherent imbalance in the ML problem at hand, the analysis primarily emphasized the evaluation of the MCC and kappa values. These metrics were chosen as they offer a less biased assessment of the model compared to other measures like the F-measure for imbalanced datasets (Chicco, Jurman, 2020). Furthermore, the F-measure, TP and FP rates were examined subsequent to the evaluation of MCC and kappa statistics. This analysis included an examination of per-class statistics. The ROC value was additionally analyzed in order to ascertain the models' ability to distinguish between different classes.

Results and Discussion

The preliminary analysis involved conducting modeling tests with different LR values and varying the number of epochs. The resulting MCC values ranged from 0.34 to 0.42. Notably, the model trained with an LR value of 0.85 and 500 epochs achieved the highest MCC value among all the models tested. Furthermore, it is worth noting that the aforementioned model exhibited the highest Kappa coefficient, which was measured at 0.36. The weighted TP rates for all models ranged from 0.659 to 0.745. In contrast, the TP rate for the anomalous data class ranged from 0.94 to 0.73, with the best model achieving a TP rate of 0.843. The average weighted F-measure for all 18 models was found to be 0.753. However, the model that was considered the best overall exhibited a higher weighted F-measure value of 0.774, surpassing the mean value. In general, the model with a LR of 0.85 and 500 epochs exhibited satisfactory evaluation metric statistics. It demonstrated a higher F-measure compared to the average F-measure of all the models, as well as higher values for the MCC and Kappa coefficient.

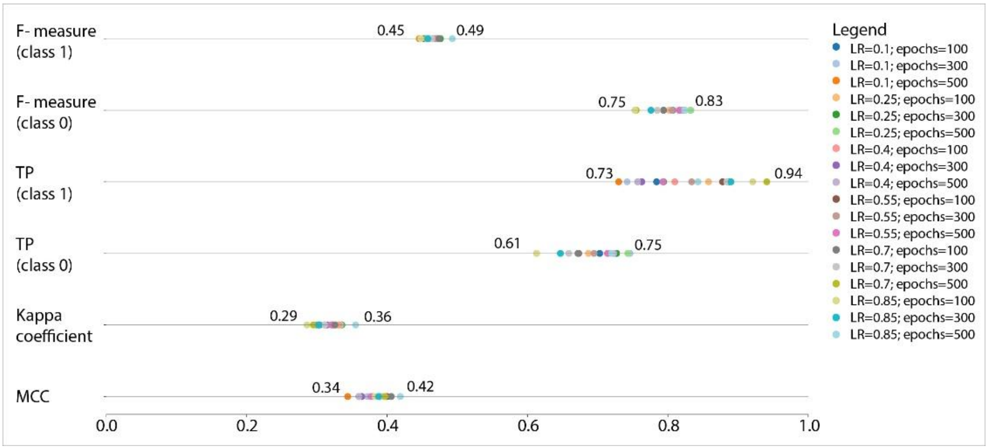

Fig. 1. Evaluation metrics for the first phase of the hyperparameter tuning.

The second iteration of modeling was conducted using a fixed LR and epoch values of 0.85 and 500, respectively. However, the momentum was not constant and varied from 0.1 to 0.9 in increments of 0.1. The observed CCI values exhibited a greater upper limit of 84%. However, the model responsible for generating these outcomes demonstrated inaccurate MCC and Kappa values, rendering it unsuitable for further analysis. As a consequence, the model was excluded from consideration. In contrast, among the models considered in the second round, the model with a momentum value of 0.2, which was also employed in the initial modeling phase, emerged as the best model overall.

The figure depicting the performance of the best overall model (LR= 0.85, trained for 500 epochs and with a momentum of 0.2) was created to illustrate both successful and unsuccessful classifications. Figures 2a and 2b in the upper panels depict the pre-processing outcomes conducted by the researchers, whereas the bottom panels illustrate the display generated by the MLP classification. In the case of the satisfactory classification example, it is evident that the model successfully classified erroneous data points accurately. In the case of the poorer classification example, the model misclassified a significant number of data points that were determined to be non-anomalous and also inaccurately interpreted the duration of the anomaly.

It is important to note that the time intervals depicted in Figure 2 may not be continuous due to the removal of erroneous data points during the pre-processing stage for MLP classification. Additionally, the figure is presented in instances representing minutes of the day rather than in units of time.

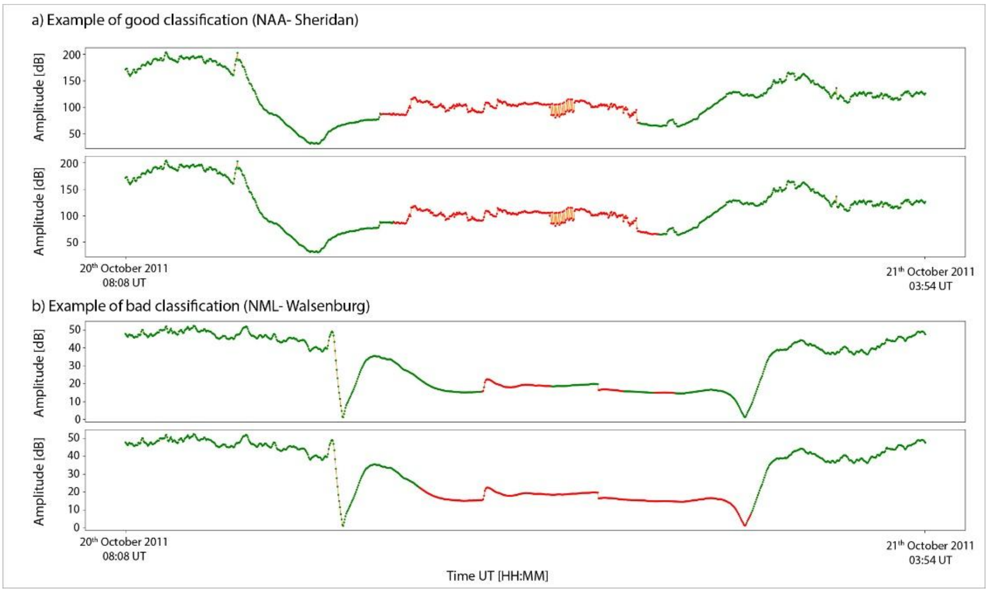

Fig. 2. Examples of good and bad classifications made by the model.

When compared to other ML methods, such as the RF model, the MLP model requires more extensive fine-tuning and computational resources. Consequently, the RF model can be considered a more favorable initial alternative for assessing the suitability and solvability of a given task using ML methods, given the available and used features and data.

Conclusions

Multilayer Perceptron models exhibit the potential for further refinement and future investigation in optimizing additional parameters. The findings suggest that the model with a LR of 0.85, 500 epochs, and a momentum value of 0.2 performed the best among all MLP models evaluated. The model chosen as best and described in the study exhibited the highest MCC and kappa values, as well as F-measure values that surpassed the average value for all models developed.

Once further research has been conducted and fine- tuning have been made, the model can be employed for the purpose of classifying ionospheric VLF amplitudes with supposed high quality. This model can also be used for future comparisons with other models and for the exclusion of future data in the analysis of ionospheric VLF amplitudes.

Acknowledgments

This work was funded by the Institute of Physics Belgrade, University of Belgrade, through a grant by the Ministry of Science, Technological Development and Innovations of the Republic of Serbia.

References

Chicco, D. and Jurman, G., 2020. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC genomics, 21(1), pp.1-13.

Devi, D., Biswas, S. K. and Purkayastha, B., 2020, July. A review on solution to class imbalance problem: Undersampling approaches. In 2020 international conference on computational performance evaluation (ComPE) (pp. 626-631). IEEE.

Frank, E., Hall, M. A., Witten, I. H., 2016. The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.

Kulkarni, A., Chong, D. and Batarseh, F. A., 2020. Foundations of data imbalance and solutions for a data democracy. In Data democracy (pp. 83-106). Academic Press.

Prusa, J., Khoshgoftaar, T. M., Dittman, D. J. and Napolitano, A., 2015, August. Using random undersampling to alleviate class imbalance on tweet sentiment data. In 2015 IEEE international conference on information reuse and integration (pp. 197-202). IEEE.