World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:17, No:08, 2023

DeepNIC a Method to Transform Each Tabular Variable into an Independant Image Analyzable by Basic CNNs

Authors : Nguyen J. M., Lucas G., Ruan S., Digonnet H., Antonioli D.

Abstract : Introduction: Deep Learning (DL) is a very powerful tool for analyzing image data. But for tabular data, it cannot compete with machine learning methods like XGBoost. The research question becomes: can tabular data be transformed into images that can be analyzed by simple CNNs (Convolutional Neuron Networks)? Will DL be the absolute tool for data classification? All current solutions consist in repositioning the variables in a 2x2 matrix using their correlation proximity. In doing so, it obtains an image whose pixels are the variables. We implement a technology, DeepNIC, that offers the possibility of obtaining an image for each variable, which can be analyzed by simple CNNs. Material and method: The 'ROP' (Regression OPtimized) model is a binary and atypical decision tree whose nodes are managed by a new artificial neuron, the Neurop. By positioning an artificial neuron in each node of the decision trees, it is possible to make an adjustment on a theoretically infinite number of variables at each node. From this new decision tree whose nodes are artificial neurons, we created the concept of a 'Random Forest of Perfect Trees' (RFPT), which disobeys Breiman's concepts by assembling very large numbers of small trees with no classification errors. From the results of the RFPT, we developed a family of 10 statistical information criteria, Nguyen Information Criterion (NICs), which evaluates in 3 dimensions the predictive quality of a variable: Performance, Complexity and Multiplicity of solution. A NIC is a probability that can be transformed into a grey level. The value of a NIC depends essentially on 2 super parameters used in Neurops. By varying these 2 super parameters, we obtain a 2x2 matrix of probabilities for each NIC. We can combine these 10 NICs with the functions AND, OR, and XOR. The total number of combinations is greater than 100,000. In total, we obtain for each variable an image of at least 1166x1167 pixels. The intensity of the pixels is proportional to the probability of the associated NIC. The color depends on the associated NIC. This image actually contains considerable information about the ability of the variable to make the prediction of Y, depending on the presence or absence of other variables. A basic CNNs model was trained for supervised classification. Results: The first results are impressive. Using the GSE22513 public data (Omic data set of markers of Taxane Sensitivity in Breast Cancer), DEEPNic outperformed other statistical methods, including XGBoost. We still need to generalize the comparison on several databases. Conclusion: The ability to transform any tabular variable into an image offers the possibility of merging image and tabular information in the same format. This opens up great perspectives in the analysis of metadata.

Keywords : tabular data, CNNs, NICs, DeepNICs, random forest of perfect trees, classification
Conference Title : ICDIA 2023 : International Conference on Data Intelligence and Analytics
Conference Location : Paris, France
Conference Dates : August 24-25, 2023