ExpDNN: Explainable Deep Neural Network

Chi-Hua Chen (✉ chihua0826@gmail.com)

Fuzhou University <https://orcid.org/0000-0001-7668-7425>

Research Article

Keywords: Explainable neural network, Explainable layer, Neural network

Posted Date: April 2nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-299913/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

In recent years, deep neural networks have been applied to obtain high performance of prediction, classification, and pattern recognition. However, the weights in these deep neural networks are difficult to be explained. Although a linear regression method can provide explainable results, the method is not suitable in the case of input interaction. Therefore, an explainable deep neural network (ExpDNN) with explainable layers is proposed to obtain explainable results in the case of input interaction. Three cases were given to evaluate the proposed ExpDNN, and the results showed that the absolute value of weight in an explainable layer can be used to explain the weight of corresponding input for feature extraction.

Full Text

This preprint is available for [download as a PDF](#).

Figures

Figure 1