Abstract P39

We propose a new method for feature selection by combining neural network classifiers, machine learning interpretability, and rank aggregation algorithms. It takes into account the interaction between different features, helps with the identification of outliers, and allows class and even sample-specific analysis of which features are relevant for a particular problem. The practical goal is to aid research in Bioinformatics, for datasets that contain a large number of irrelevant features.

Introduction

Several methods were proposed to tackle the problem of interpreting the learned behavior of artificial neural networks. These new algorithms enabled researchers to understand better what and how the neural network has learned and which features in the input space were deemed relevant to perform correct classification. It found many applications in analyzing the function to understand its inner-workings. Several methods were proposed to tackle the problem of interpreting the learned behavior of artificial neural networks. These new algorithms enabled researchers to understand better what and how the neural network has learned and which features in the input space were deemed relevant to perform correct classification. It found many applications in analyzing the function to understand its inner-workings.

Problems of Interest

We are interested in problems of feature selection within Bioinformatics:

Gene selection: the expression levels of thousands of genes are available for a few samples that contain specific conditions (e.g., diseases) [2].

Forensic biology: how mutations in the DNA are related to phenotypes. One of the goals is the forensic characterization of Brazilian regional populations [3].

Cancer immunotherapy: identifying the relevance of different immune system cells in response to cancer.

General View

We need a function that maps all the features to the studied condition and to be capable of investigating this function to understand its inner-workings.

For each sample, rank the features by their relevance.

Relevance Aggregation

Layer-wise Relevance Propagation (LRP) [4] is an algorithm for interpretation capable of identifying the specific features responsible for the network’s output for each input sample. It found many applications in analyzing the inner-workings and quality of image and text classifiers.

How to select the features global or classwise?

Dimensionality reduction: irrelevant or redundant features in the data are discarded.

Improve the accuracy of classifiers, reduce memory consumption and processing time.

Further interpretation: the original meaning of the features is preserved [1].

Mathematical Formulation

Let's consider the two passes required for computing relevances.

First pass:

For the synthetic data, the relevant features were identified and ranked in the top positions. For the eye color dataset, this distinction is not so clear, but there is an order that emerged, and the ranks of the features vary according to the classes. The average global rank of SNP02 is 21.43, if we consider only the class Blue, it is 2.16, while for class Intermediate it becomes 60.49. This may suggest a great difference in the relation of this SNP to each of the phenotypes being studied. Another possibility is the inspection of outliers by observing clashing patterns in the relevances of features.

Results

Datasets:

Synthetic: 1000 samples, 5 relevant, 5 redundant, and 990 irrelevant features, 3 classes.

Eye color: 72 samples, 67 SNP, 3 colors.

Classifier: (i) Feedforward dense neural network with SGD. (ii) ReLU activation. (iii) Dropout of 50%.

Relevance: (i) LRP-αβjk rule for hidden layers. (ii) LRP-αβjk for the input layer. (iii) Aggregation with Borda count.

For the synthetic data, the relevant features were identified and ranked in the top positions. For the eye color dataset, this distinction is not so clear, but there is an order that emerged, and the ranks of the features vary according to the classes. The average global rank of SNP02 is 21.43, if we consider only the class Blue, it is 2.16, while for class Intermediate it becomes 60.49. This may suggest a great difference in the relation of this SNP to each of the phenotypes being studied. Another possibility is the inspection of outliers by observing clashing patterns in the relevances of features.

Challenges

Training good classifiers with small datasets.

Generalization.

For connected layers, LRP loses selectivity.

Summarization of feature relevances through different rank aggregation algorithms.

Conclusion

Although still in its early stages, the results obtained from the two described experiments are encouraging. The next steps are testing different interpretation and rank algorithms, and the development of new network structures or propagation rules that do not lose selectivity.

References


Acknowledgement:

Selection of Biological Features by Ranking the Relevances of Neural Network Inputs

Bruno Iochins Grici1,2, Mathias J. Krause 2 and Márcio Dorn3

1Institute of Informatics, UFRGS, Brazil.
2Institute of Applied and Numerical Mathematics, KIT, Germany.
3Institute of Computational and Bioinformatics, BioISI, University of Minho.

21st EMBL PhD Symposium: Facing the Future, November 28 to 30, 2019.

Figure 1: A schematic diagram of the two passes required for computing relevances.

Figure 2: LRP results for the classification of digit 3 as 3 and as 8. [4]

Figure 3: For each dataset the top ranked features are shown. The intensity of the red (positive) and blue (negative) cells is proportional to their relevance, and the numerical values are the raw data.

Figure 4: The original meaning of the features is preserved [1].