Ranking Relevances from Neural Networks for the Selection of Biological Features

Bruno Iochins Grisci[1,2], Mathias J. Krause [2] and Márcio Dorn[1]

[1]Institute of Informatics, UFRGS, Brazil.

[2]Institute of Applied and Numerical Mathematics, KIT, Germany.

bigrisci@inf.ufrgs.br

7th Heidelberg Laureate Forum, September 22 to 27, 2019.

Abstract

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

In the past years, several methods were proposed to tackle the problem of interpreting the learned behavior of artificial neural networks, for much time considered to be "black boxes." 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.

In this work, we propose the use of interpretability methods for feature selection:

- **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].

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), and a solution is the subset of genes accountable for that condition [2].
- **Forensic biology**: the features are specific mutations in the DNA, and one wants to know how they are related to phenotypes (e.g., hair color). 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.

In these problems, an ideal algorithm would also point out to possible outliers (as biological data is subject to contamination and experimental errors), and show the relationship between the features.

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.

- Takes all features at once and considers the relationship between them.
- Does not necessarily discard redundant features.
- The classification performance is a measure of the quality of the function, but not the metric for selection.
- It would require the training of only one neural network, being more efficient than wrappers [2], that need the creation of multiple classifiers.

Relevance Propagation

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.

LRP works with two passes through a trained neural network (Fig. 1):

1. **Feedforward**: goes from the input layer to the output layer and computes its output.
2. **Backward:** sends the output value back through the network structure as a relevance message that is distributed among the neurons in the previous layers [4].

The computation uses several rules that take into account the input domain and layer type, as in in Eq. 1 and 2. For both of them, k and j are the kth and jth layers, a is the output of a neuron, w^+ and w^- are positive and negative weights, α and β are constants that must obey $\alpha - \beta = 1$ and $\beta \geq 0$, and R is the relevance signal.

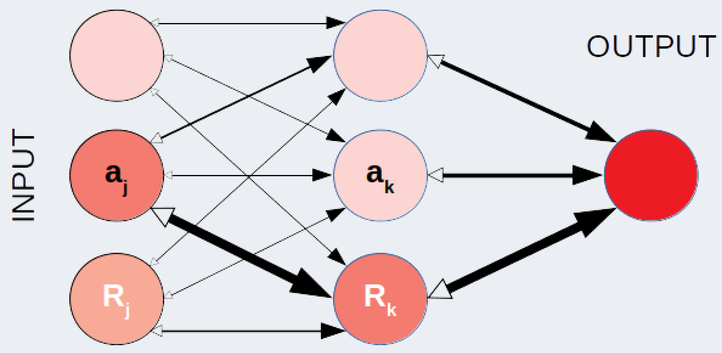

Figure 1: A schematic diagram of the two passes required for computing relevances.

$$R_j = \sum_k \left(\alpha \frac{a_j w_{jk}^+}{\sum_j a_j w_{jk}^+} - \beta \frac{a_j w_{jk}^-}{\sum_j a_j w_{jk}^-} \right) R_k \quad (1)$$

$$R_j = \sum_k \frac{w_{jk}^2}{\sum_j w_{jk}^2} R_k \quad (2)$$

Relevance Aggregation

We can compute the relevance of each feature for each sample with LRP.
But how to select the features global or classwise?

1. Rank the features by their relevance.
2. Give each sample a vote in the form of its rank.
3. Aggregate the ranks.

	S1	S2	S3	Aggregation	Rank
A	1	2	1	$\sqrt[3]{1 \times 2 \times 1} = 1.26$	1
B	2	2	2	$\sqrt[3]{2 \times 2 \times 2} = 2.00$	2
C	4	4	3	$\sqrt[3]{4 \times 4 \times 3} = 3.63$	4
D	3	1	4	$\sqrt[3]{3 \times 1 \times 4} = 2.29$	3

Challenges

- LRP is applied in image and text classification using convolutional neural networks. For fully connected layers, it loses selectivity.
- Summarization of feature relevances through different rank aggregation algorithms.

Results

- **Datasets**:
  3XOR: 200 samples of 20 binary features in the classes zero or one, defined by the XOR operation of only three specific features.
  Leukemia: 72 samples composed of 7129 values of gene expression into acute lymphocytic leukemia (ALL) or acute myelogenous leukemia (AML).
- **Classifier**: (i) Feedforward dense neural networks with SGD. (ii) Tanh and ReLU activation for the hidden layers. (iii) Dropout of 50%.
- **Relevance**: (i) LRP-$\alpha_1\beta_0$ rule for hidden layers. (ii) LRP-w^2 rule for the input layer. (iii) Aggregation with Borda count using geometric mean.

The preliminary results can be seen in Fig. 2. For the 3XOR problem, the three relevant features out of 20 were identified and rank in the top positions. For the leukemia 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 feature X58529 is 60.38, but if we consider only class ALL, it is 12.18, while for class AML it becomes 1,224.09. This may suggest a great difference in the relation of this feature to each of the conditions being studied. Another possibility is the inspection of outliers by observing clashing patterns in the relevances of features.

	RANK	REL	zero	OUT one	2,43 *zero*	2,41 *zero*	2,45 *zero*	2,36 *zero*	2,32 *zero*	1,94 *one*	1,79 *one*	1,69 *one*	1,67 *one*	1,63 *one*	1,54 *one*
R_10	1,38	33,98	1,33	1,43	1	1	1	1	1	0	0	0	0	0	1
R_3	2,08	29,98	2,12	2,05	1	1	1	1	1	0	1	0	0	1	0
R_17	2,09	30,36	2,12	2,05	1	1	1	1	1	1	0	1	1	0	0
x_2	5,29	9,37	5,13	5,46	0	1	0	1	1	0	0	0	0	0	0
x_11	5,97	7,38	6,91	5,16	0	1	1	0	1	0	0	0	1	1	0
x_18	7,66	6,38	8,22	7,13	0	0	0	1	1	1	1	1	1	1	1
x_7	8,34	6,01	7,73	9,00	0	0	0	0	0	1	0	1	0	0	1
x_15	8,74	5,73	7,26	10,53	0	0	0	0	0	1	1	1	0	0	1
x_16	9,00	5,63	8,36	9,70	0	1	1	0	0	1	1	1	1	1	0
x_13	10,37	5,29	10,14	10,60	1	1	0	0	0	0	1	0	1	1	1
x_4	11,23	4,50	11,22	11,23	1	0	1	1	0	1	0	1	1	1	1
x_5	11,36	4,67	12,92	9,99	0	0	0	0	0	0	0	0	0	0	0
x_12	12,03	4,57	10,70	13,53	0	0	0	1	1	1	0	0	1	0	1
x_14	12,68	4,09	12,48	12,89	0	0	0	0	1	0	0	1	1	1	1
x_9	13,69	3,80	13,69	13,68	0	1	0	0	0	1	1	1	1	1	1
x_6	14,25	3,56	13,73	14,79	0	0	1	0	1	0	0	0	1	1	1
x_1	15,36	3,28	15,77	14,96	1	1	1	1	1	0	0	1	1	1	0
x_8	15,38	3,65	14,35	16,49	0	0	1	1	1	0	1	1	1	1	1
x_19	15,88	2,79	18,95	13,30	1	1	0	0	0	0	0	0	0	0	1
x_0	17,29	2,57	18,77	15,93	1	1	0	1	0	1	1	1	0	0	0

(a) 3XOR

	RANK	REL	ALL	OUT AML	21,02 *ALL*	25,75 *ALL*	17,69 *ALL*	24,47 *ALL*	17,77 *ALL*	21,09 *AML*	23,81 *AML*	32,71 *AML*	17,97 *AML*	29,73 *AML*
M19507	1,50	3,21	1,20	2,27	706	199	18	-60	-146	747	22222	14230	27285	5895
M23197	25,51	1,82	13,65	82,61	287	222	73	128	117	1533	629	646	336	1883
U59632_s	27,62	1,87	14,55	92,13	365	360	150	-114	142	126	190	358	269	234
M84371_rna1_s	49,66	1,77	151,29	6,12	779	2338	1681	2305	1995	291	517	126	655	996
U05572_s	53,40	1,90	95,60	17,86	-142	-114	461	-2	80	1178	275	556	466	942
HG3731.HT4001	56,74	1,78	67,35	41,12	126	85	36	-572	-657	-42	-162	-383	-226	-40
X58529	60,38	1,56	12,18	1.224,09	1458	6074	7243	3394	3995	18	235	1190	328	651
M32304_s	78,67	1,72	92,50	58,04	582	1114	369	240	455	1558	971	1060	1024	1926
U50822_rna1_s	81,64	1,51	21,90	968,59	-119	-19	0	-34	-239	-62	79	3	-120	-152
M27891	85,94	1,67	70,21	125,67	107	-177	-125	502	78	17846	10737	14193	2460	19680
M24902	87,73	1,63	202,62	18,18	189	20	-73	14	21	29	187	262	88	263
X66401_cds1	89,28	1,61	76,87	118,32	1758	3172	1192	2947	2072	1048	653	1149	618	1133
X95735	90,44	1,39	28,03	817,82	399	805	252	-152	-61	2007	4403	2871	2122	5949
X70297	92,04	1,66	149,73	36,87	107	218	6	-42	-37	-26	279	346	304	443
X59711	94,82	1,56	65,73	188,84	52	95	43	137	142	-27	48	-27	15	35

(b) Leukemia

Figure 2: Summary of results for the 3XOR problem and the gene expression of a Leukemia dataset. For each dataset the top ranked features are shown, besides five samples from each class (in green and yellow). The intensity of the red cells is proportional to their relevance, and the numerical values are the raw data.

Conclusion

In this work, we propose the use of artificial neural networks, Layer-wise Relevance Propagation, and rank aggregation for the selection of relevant biological features. 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

[1] Bruno Iochins Grisci, Bruno César Feltes, and Marcio Dorn.
Neuroevolution as a tool for microarray gene expression pattern identification in cancer research.
Journal of biomedical informatics, 89:122–133, 2019.

[2] Jun Chin Ang, Andri Mirzal, Habibollah Haron, and Haza Nuzly Abdull Hamed.
Supervised, unsupervised, and semi-supervised feature selection: a review on gene selection.
IEEE/ACM transactions on computational biology and bioinformatics, 13(5):971–989, 2016.

[3] Eduardo Avila, Aline Brugnera Felkl, Pietra Graebin, Cláudia Paiva Nunes, and Clarice Sampaio Alho.
Forensic characterization of brazilian regional populations through massive parallel sequencing of 124 snps included in hid ion ampliseq identity panel.
Forensic Science International: Genetics, 40:74–84, 2019.

[4] Sebastian Bach, Alexander Binder, Grégoire Montavon, Frederick Klauschen, Klaus-Robert Müller, and Wojciech Samek.
On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation.
PloS one, 10(7):e0130140, 2015.

Acknowledgement: