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

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

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

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.

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) [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.

- 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.

The computation uses 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.

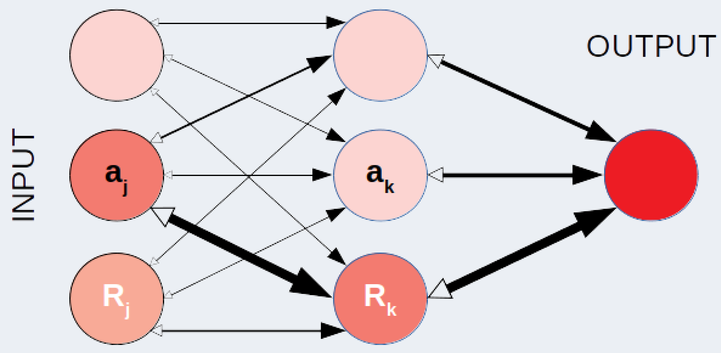

Figure 1: A schematic diagram of the two passes required for computing relevances.

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

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

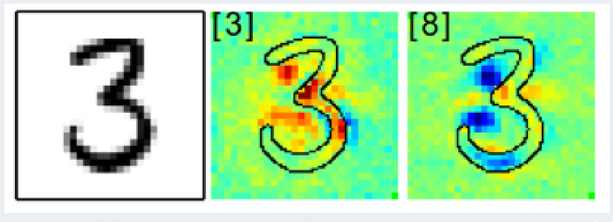

Figure 2: LRP results for the classification of digit 3 as 3 and as 8. [4]

Relevance Aggregation

How to select the features global or classwise?

1. For each sample, 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

Results

- **Datasets**:
  Synthetic: 1000 samples, 5 relevant, 5 redundant, and 990 irrelevant features, 3 classes.
  Eye color: 72 samples, 67 SNPs, 3 colors.
- **Classifier**: (i) Feedforward dense neural network with SGD. (ii) ReLU activation. (iii) Dropout of 50%.
- **Relevance**: (i) LRP-$\alpha_2\beta_1$ rule for hidden layers. (ii) LRP-w^2 rule 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_A is 21.43, but 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.

			Pred	0.98	0.98	0.99	0.99	1.0	1.0	1.0	1.0	0.99	0.99	1.0	1.0
	RANK	0	1	2											
RED009	55,72	11,12	7,90	147,52	-0,19	2,48	-3,65	-4,14	7,71	2,47	2,56	0,08	-1,21	-2,87	2,20	5,86
RED007	90,19	263,97	5,55	1,06	5,12	3,78	-1,95	-9,43	-1,74	-1,33	-4,40	9,39	4,32	1,84	6,55	10,35
REL001	115,94	341,58	2,11	4,12	-2,39	-3,36	4,04	2,18	-1,65	-2,95	-3,75	3,53	0,73	1,38	2,35	2,69
REL003	121,92	360,26	2,94	2,57	-2,38	-1,63	2,67	4,58	3,85	2,92	2,43	-3,24	-3,62	-1,83	-1,08	-3,88
RED008	123,08	365,77	1,03	2,42	-0,79	0,83	-2,07	3,12	6,71	4,08	7,07	-8,25	-4,55	-4,40	-3,86	-6,42
REL004	130,90	10,56	6,67	373,83	-1,92	-2,83	2,46	4,33	-8,02	-2,66	-2,94	-0,81	1,67	3,72	-3,54	-5,04
REL005	130,97	383,53	4,53	4,85	5,89	3,05	-1,35	-5,56	-3,66	-1,31	-1,82	4,49	2,98	1,06	2,80	2,49
REL002	162,39	475,75	5,44	5,98	4,64	1,91	-2,89	-1,76	-0,80	-1,94	2,94	-2,70	0,56	-2,12	-1,15	-3,22
RED010	217,58	18,02	14,79	617,21	-6,24	-4,58	3,64	5,84	-1,87	-1,34	-2,51	-0,73	-0,53	2,12	-1,89	-1,19
IRR468	231,08	417,09	82,74	192,93	1,26	-0,81	-1,17	-0,40	2,25	0,54	0,38	1,60	-1,89	-0,34	-0,35	0,25
IRR426	247,33	360,24	348,03	35,15	-0,54	0,35	-0,35	-1,22	-0,78	-0,52	-1,10	0,08	0,75	0,55	-0,99	2,99
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
IRR910	774,53	616,67	807,60	898,90	-0,42	1,13	3,34	-0,07	-0,57	-0,47	0,61	1,14	-0,55	-0,17	1,23	0,89

(a) Synthetic

				Prediction	0.71	0.90	0.93	1.00	1.00	0.87	0.87	0.90	0.96	0.96	0.91	0.91	0.91	1.00	1.00
	RANK	BLUE	INTER	DARK															
SNP01_C	17.00	1.09	62.51	3.38	0.5	0.5	0	0	0	0.5	0.5	0.5	0	1	0.5	0.5	0.5	1	1
SNP02_A	21.43	2.16	60.49	12.33	0.5	0.5	0	0	0	0.5	0.5	0.5	0	1	0.5	0.5	0.5	1	1
SNP01_T	21.92	2.86	79.44	4.20	0.5	0.5	1	1	1	0.5	0.5	0.5	1	0	0.5	0.5	0.5	0	0
SNP02_G	38.46	3.96	63.67	42.73	0.5	0.5	1	1	1	0.5	0.5	0.5	1	0	0.5	0.5	0.5	0	0
SNP03_G	44.03	31.31	107.20	20.80	0	1	1	0.5	0	1	1	0.5	0.5	1	0.5	0.5	0	0	0.5
SNP04_T	54.51	15.30	127.91	38.76	0.5	1	0.5	0	0	0.5	0.5	0.5	0.5	1	0.5	0.5	0	0	0
SNP05_G	57.88	27.57	136.40	35.67	0.5	0	0.5	0	0	0.5	0.5	0.5	0	0	0.5	0.5	0.5	0	0.5
SNP06_G	58.80	48.70	69.87	58.36	1	0.5	0.5	0	0	0.5	0.5	0.5	0	0	1	0.5	0	0.5	0
SNP07_G	60.30	37.23	146.19	31.38	0.5	0	0	0.5	1	0.5	0	1	0.5	0.5	0.5	0	0	0	0.5
SNP03_T	61.24	12.91	112.26	60.00	1	0	0	0.5	1	0	0	0.5	0.5	0	0.5	0.5	1	1	0.5
SNP08_T	61.87	59.80	143.67	25.16	1	1	1	0.5	1	1	1	1	1	1	1	1	1	1	0.5
SNP04_C	62.47	23.27	156.47	37.22	0.5	0	0.5	1	1	0.5	0.5	0.5	0.5	0	0.5	0.5	1	1	0
SNP09_T	63.54	42.50	143.74	36.30	0.5	1	0.5	0	0	0.5	0.5	0.5	0.5	1	0.5	0.5	0	0	1
SNP10_C	67.38	103.27	83.44	43.45	0.5	1	0.5	0	0	0.5	0.5	0.5	0.5	1	0.5	0.5	0	0	1
SNP08_C	69.37	55.43	183.80	23.09	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0.5
SNP11_C	71.40	55.46	152.54	41.38	0.5	0.5	0.5	0.5	0	1	0	0.5	0	0	0	1	0	0	1
SNP12_G	72.51	183.59	62.63	25.91	0	0	0	0	0	0	0	0	0	0	0.5	0	0	0	1
SNP09_C	76.17	63.14	128.30	58.17	0.5	0	0.5	1	1	0.5	0.5	0.5	0.5	0	0.5	0.5	1	1	0
SNP13_A	76.55	183.17	117.99	8.37	1	1	1	1	1	0.5	1	0.5	1	0.5	1	1	1	1	0.5
SNP14_C	78.86	60.67	174.71	43.09	1	0.5	1	1	0.5	1	1	0.5	0.5	1	1	1	1	0.5	1
SNP15_G	81.60	178.56	108.71	24.46	0.5	0	0.5	1	1	0.5	1	0.5	0.5	0.5	0.5	1	1	1	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
SNPxx_G	265.13	301.30	184.06	285.81	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SNPyy_T	266.98	233.00	201.26	312.89	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SNPzz_G	274.75	324.03	146.17	311.26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

(b) Eye color

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.

Challenges

- Training good classifiers with small datasets.
- Generalization.
- For fully 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

[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: