

rFBP: Replicated Focusing Belief Propagation algorithm

Nico Curti^{*1}, Daniele Dall'Olio^{†3}, Daniel Remondini³, Gastone Castellani², and Enrico Giampieri¹

1 eDIMESLab, Department of Experimental, Diagnostic and Specialty Medicine of Bologna University
2 Department of Experimental, Diagnostic and Specialty Medicine of Bologna University
3 Department of Physics and Astronomy of Bologna University

DOI: [10.21105/joss.02663](https://doi.org/10.21105/joss.02663)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Editor: [Ariel Rokem](#) ↗

Reviewers:

- [@justusschock](#)
- [@DanielLenz](#)

Submitted: 19 August 2020

Published: 20 October 2020

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](#)).

Summary

The rFBP project implements a `scikit-learn` compatible machine-learning binary classifier leveraging fully connected neural networks with a learning algorithm (*Replicated Focusing Belief Propagation*, rFBP) that is quickly converging and robust (less prone to brittle overfitting) for ill-posed datasets (very few samples compared to the number of features). The current implementation works only with binary features such as one-hot encoding for categorical data.

This library has already been widely used to successfully predict *source attribution* starting from GWAS (*Genome Wide Association Studies*) data. That study was trying to predict the animal origin for an infectious bacterial disease inside the H2020 European project COMPARE (Grant agreement ID: 643476). A full description of the pipeline used in this study is available in the abstract and slides provided into the [publications](#) folder of the project.

Algorithm application on real data:

- *Classification of Genome Wide Association data by Belief Propagation Neural network*, [CCS Italy 2019](#), Conference paper
- *Classification of Genome Wide Association data by Belief Propagation Neural network*, [CCS Italy 2019](#), Conference slides

Statement of need

The learning problem under ill-posed conditions can be tackled through statistical mechanic models joined with the so-called Large Deviation Theory (Baldassi & Braunstein, 2015; Baldassi et al., 2016b; Monasson & Zecchina, 1995a, 1995b; Parisi, 2007). In general, the learning problem can be split into two sub-parts: the classification problem and the generalization one. The first aims to completely store a pattern sample, i.e., a prior known ensemble of input-output associations (*perfect learning*, Baldassi et al.) (Baldassi et al., 2016a; Krauth & Mézard, 1989). The second one corresponds to compute a discriminant function based on a set of features of the input which guarantees a unique association of a pattern.

From a statistical point-of-view many Neural Network models have been proposed and spin-glass models have emerged as the most promising ones. Starting from a balanced distribution of the system, generally based on Boltzmann distribution, and under proper conditions, we can prove that the classification problem becomes a NP-complete computational problem (Blum & Rivest, 1992). A wide range of heuristic solutions to that type of problems were proposed

*co-first author

†co-first author

(Baldassi, Braunstein, Brunel, & Zecchina, 2007; Braunstein & Zecchina, 2006; Huang & Kabashima, 2014).

In this project we show one of these algorithms developed by Baldassi et al. (Baldassi et al., 2016a) and called *Replicated Focusing Belief Propagation* (rFBP). The rFBP algorithm is a learning algorithm developed to justify the learning process of a binary neural network framework. The model is based on a spin-glass distribution of neurons put on a fully connected neural network architecture. In this way each neuron is identified by a spin and so only binary weights (-1 and 1) can be assumed by each entry. The learning rule which controls the weight updates is given by the Belief Propagation method.

A first implementation of the algorithm was proposed in the original paper (Baldassi et al., 2016a) jointly with an open-source Github repository. The original version of the algorithm was written in [Julia language](#). Julia is certainly an efficient programming language but it is not part of most machine learning developers' tool of choice. To broaden the scope and use of the method, a C++ implementation was developed with a joint Cython wrap for Python users. The C++ language guarantees better computational performances against the Julia implementation and the Python version enhances its usability. This implementation is optimized for parallel computing and is endowed with a custom C++ library called [Scorer](#), which is able to compute a large number of statistical measurements based on a hierarchical graph scheme. With this optimized implementation and its [scikit-learn](#) compatibility we try to encourage researchers to approach these alternative algorithms and to use them more frequently on real context.

As the Julia implementation also the C++ one provides the entire rFBP framework in a single library callable via a command line interface. The library widely uses template syntaxes to perform dynamic specialization of the methods between two magnetization versions of the algorithm. The main object categories needed by the algorithm are wrapped in handy C++ objects easy to use also from the Python interface.

Acknowledgments

The authors acknowledge COMPARE n. 643476 EU Horizon 2020 (EU) Project.

References

- Baldassi, C., & Braunstein, A. (2015). A max-sum algorithm for training discrete neural networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2015(8), P08008. doi:[10.1088/1742-5468/2015/08/p08008](https://doi.org/10.1088/1742-5468/2015/08/p08008)
- Baldassi, C., Braunstein, A., Brunel, N., & Zecchina, R. (2007). Efficient supervised learning in networks with binary synapses. *Proceedings of the National Academy of Sciences*, 104(26), 11079–11084. doi:[10.1073/pnas.0700324104](https://doi.org/10.1073/pnas.0700324104)
- Baldassi, C., Gerace, F., Lucibello, C., Saglietti, L., & Zecchina, R. (2016a). Learning may need only a few bits of synaptic precision. *Physical Review E*, 93(5). doi:[10.1103/physreve.93.052313](https://doi.org/10.1103/physreve.93.052313)
- Baldassi, C., Ingrosso, A., Lucibello, C., Saglietti, L., & Zecchina, R. (2016b). Local entropy as a measure for sampling solutions in constraint satisfaction problems. *Journal of Statistical Mechanics: Theory and Experiment*, 2016(2), 023301. doi:[10.1088/1742-5468/2016/02/023301](https://doi.org/10.1088/1742-5468/2016/02/023301)
- Blum, A. L., & Rivest, R. L. (1992). Training a 3-node neural network is NP-complete. *Neural Networks*. doi:[10.1016/S0893-6080\(05\)80010-3](https://doi.org/10.1016/S0893-6080(05)80010-3)

- Braunstein, A., & Zecchina, R. (2006). Learning by message passing in networks of discrete synapses. *Physical Review Letters*, 96(3). doi:[10.1103/physrevlett.96.030201](https://doi.org/10.1103/physrevlett.96.030201)
- Huang, H., & Kabashima, Y. (2014). Origin of the computational hardness for learning with binary synapses. *Physical Review E*, 90(5). doi:[10.1103/physreve.90.052813](https://doi.org/10.1103/physreve.90.052813)
- Krauth, W., & Mézard, M. (1989). Storage capacity of memory networks with binary couplings. *Journal De Physique*, 50, 3057–3066. doi:[10.1051/jphys:0198900500200305700](https://doi.org/10.1051/jphys:0198900500200305700)
- Monasson, R., & Zecchina, R. (1995a). Weight space structure and internal representations: A direct approach to learning and generalization in multilayer neural networks. *Physical Review Letters*, 75(12), 2432–2435. doi:[10.1103/physrevlett.75.2432](https://doi.org/10.1103/physrevlett.75.2432)
- Monasson, R., & Zecchina, R. (1995b). Learning and generalization theories of large committee-machines. *Modern Physics Letters B*, 09(30), 1887–1897. doi:[10.1142/S0217984995001868](https://doi.org/10.1142/S0217984995001868)
- Parisi, G. (2007). Mean field theory of spin glasses: Statics and dynamics. Retrieved from <http://arxiv.org/abs/0706.0094>