

A topological insight into restricted Boltzmann machines (extended abstract)¹

Decebal Constantin Mocanu ^a Elena Mocanu ^a Phuong H. Nguyen ^a
Madeleine Gibescu ^a Antonio Liotta ^a

^a *Eindhoven University of Technology, Dep. of Electrical Engineering, Netherlands*

1 Introduction

Restricted Boltzmann Machines (RBMs) and models derived from them have been successfully used as basic building blocks in deep neural networks for automatic features extraction, unsupervised weights initialization, but also as standalone models for density estimation, activity recognition and so on. Thus, their generative and discriminative capabilities, but also their computational time are instrumental to a wide range of applications. The main contribution of this paper [4] is to study the above problems by looking at RBMs and Gaussian RBMs (GRBMs) [2] from a topological perspective, bringing insights from network science, an extension of graph theory which analyzes real world complex networks [6].

2 The proposed method

Firstly, we study the topological characteristics of RBMs and GRBMs, showing that these exhibit a small-world topology. We then hypothesize that by constraining the topology to be also scale-free it is possible to reduce the size of ordinary RBMs and GRBMs models, as it has been shown in [1] that scale-free networks are sparse. Thus, we introduce a three stages method to create RBMs and GRBMs with small-world, scale-free topologies while still considering local neighborhoods and data distribution. In the first stage, a scale-free bipartite graph is generated; in the second one, the graph is adjusted to be also small-world; and in the third stage, the graph topology is fitted to the data distribution. We dub the resulting models as complex Boltzmann Machine (XBM) (see Fig. 1a) and Gaussian complex Boltzmann Machine (GXBM), respectively.

An interesting finding is that constraining such XBM and GXBM topologies at their inception leads to intrinsically sparse networks, a considerable advantage to typical state-of-the-art methods in which sparsity is enforced as an aftermath, that is during testing (exploitation) phase (e.g. [3]). In turn, XBM

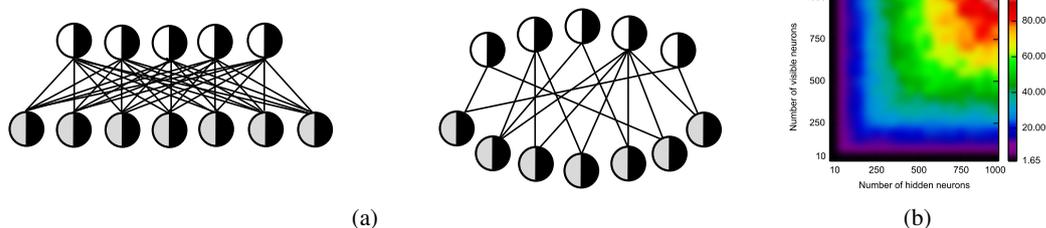


Figure 1: (a) Schematic architecture of: RBM (left) and XBM (right). (b) Studying the relation between the number of weights in RBM and XBM (the heatmap values are given by n_w^{RBM}/n_w^{XBM}), where n_w is the number of weights of the specific model.

¹The full paper has been published in *Machine Learning*, Volume 104, 2016, Pages 243–270, ISSN 1573-0565, <http://dx.doi.org/10.1007/s10994-016-5570-z>.

Table 1: Estimation of the average log-probabilities on the training and testing data obtained from the MNIST digits dataset using AIS [5] on fully connected RBM, XBM and two state-of-the-art sparse models (i.e. $\text{RBM}_{\text{FixProb}}$, $\text{RBM}_{\text{TrPrTr}}$). The RBM result is taken from [5].

No. of CD steps during learning	No. of weights	Model	No. of hidden units	Average shortest path	Average cluster coefficient	No. of pruning iterations	Average train log-probabilities	Average test log-probabilities
from 1 to 25 (variable)	392000	RBM	500	1.52	1	0	-83.10	-86.34
	387955	XBM	27000	2.05	0.156	0	-86.12	-85.21
	391170	$\text{RBM}_{\text{FixProb}}$	27000	2.87	0.053	0	-107.23	-106.78
	3262957	$\text{RBM}_{\text{TrPrTr}}$	27000	2.18	0.076	50	-349.87	-376.92
	10790	XBM	500	2.44	0.082	0	-121.26	-120.43
	10846	$\text{RBM}_{\text{FixProb}}$	500	3.12	0.039	0	-136.27	-135.89
	36674	$\text{RBM}_{\text{TrPrTr}}$	500	2.35	0.071	50	-134.25	-135.76

and GXBM have a considerably smaller number of weights, which further on contributes to considerably faster computational times (proportional to the number of weights in the model), both in the training and testing phases. What is more, we found that the proposed topology imposes an inductive bias on XBMs and GXBMs, which leads to better statistical performance than RBMs and GRBMs. Our comparative study is based on both simulated and real-world data, including the Geographical origin of music dataset, the MNIST digits dataset, CalTech 101 Silhouettes dataset, and the 8 datasets from UCI evaluation suite. We show that, given the same number of hidden neurons, XBM and GXBM have similar or relatively close capabilities to RBM and GRBM, but are considerably faster thanks to their reduced amount of weights. For instance, in a network of 100 visible and 100 hidden neurons, the reduction in weights was by one order of magnitude. A network with 1000 visible and 1000 hidden neurons led to a reduction in weights by two orders of magnitude, as depicted in Fig. 1b. Additionally, we show that given the same amount of weights, XBMs and GXBMs achieve better generative capabilities than fully connected RBMs or GRBMs, due to their higher number of hidden neurons. For the sake of illustration, Table 1 presents a snapshot of XBM performance on the MNIST dataset.

3 Conclusion

In this paper [4], we look at RBMs and GRBMs from a topological perspective, bringing insights from network science. Firstly, we point out that RBMs and GRBMs are small-world bipartite networks. Secondly, by introducing scale-free constraints, we devised two novel sparse models, namely XBMs and GXBMs. These sparse models exhibit much faster computational time than their fully connected counterparts thanks to a smaller number of parameters which have to be computed, at almost no cost in performance. We believe that this will lead to the ability to tackle problems having much higher dimensional data - something that is today unfeasible without performing dimensionality reduction - and we intend to tackle this research direction in the near future.

References

- [1] Charo I. Del Genio, Thilo Gross, and Kevin E. Bassler. All scale-free networks are sparse. *Phys. Rev. Lett.*, 107:178701, Oct 2011.
- [2] G. E. Hinton and R. R. Salakhutdinov. Reducing the Dimensionality of Data with Neural Networks. *Science*, 313(5786):504–507, July 2006.
- [3] Honglak Lee, Chaitanya Ekanadham, and Andrew Y. Ng. Sparse deep belief net model for visual area v2. In J.C. Platt, D. Koller, Y. Singer, and S.T. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 873–880. Curran Associates, Inc., 2008.
- [4] Decebal Constantin Mocanu, Elena Mocanu, Phuong H. Nguyen, Madeleine Gibescu, and Antonio Liotta. A topological insight into restricted boltzmann machines. *Machine Learning*, 104(2):243–270, 2016.
- [5] Ruslan Salakhutdinov and Iain Murray. On the quantitative analysis of deep belief networks. In *In Proceedings of the International Conference on Machine Learning*, pages 872–879, 2008.
- [6] Steven H. Strogatz. Exploring complex networks. *Nature*, 410(6825):268–276, March 2001.