# On the Model Selection of Bernoulli Restricted Boltzmann Machines Through Harmony Search

Joao Papa, Gustavo Rosa
Kelton Costa, Aparecido Marana
Department of Computing
São Paulo State University
Bauru, Brazil
{papa,nilceu}@fc.unesp.br
gth.rosa@uol.com.br
kelton.costa@gmail.com

Walter Scheirer
David Cox
Center for Brain Science
Harvard University
Cambridge, USA
{wscheirer,davidcox}@fas.harvard.edu

## ABSTRACT

Restricted Boltzmann Machines (RBMs) are amongst the most widely pursued techniques in deep learning-based environments. However, the problem of selecting a suitable set of parameters still remains an open question, since it is not straightforward to choose them without prior knowledge. In this paper, we introduce the Harmony Search (HS) optimization algorithm to find out a suitable set of parameters that minimize the reconstruction error of Bernoulli RBMs, which address binary-valued visible and hidden units. The results have shown the suitability of using HS for such task when compared to other optimization techniques.

## Categories and Subject Descriptors

I.2 [**Artificial Intelligence**]: Learning

## Keywords

Bernoulli Restricted Boltzmann Machines; Harmony Search

## 1. INTRODUCTION

Deep learning-based techniques have been paramount in the last years, mainly because of their very high recognition rates in several applications. Restricted Boltzmann Machines (RBMs) are among the most widely used techniques, since they obtained interesting results in different research areas [3]. However, one of the main problems related to RBMs is their parameter selection, which has been hand-tuned in several works out there.

In this work, we focused on employing meta-heuristic optimization techniques for RBM model selection, i.e., parameter fine-tuning. We highlight here techniques based on the Harmony Search (HS) paradigm [2], since they are usually much faster than swarm-oriented techniques (e.g., Particle Swarm Optimization - PSO [4]), which often require all possible solutions to be updated at each iteration. Next sections describe in more details the proposed approach, as well as the experimental section.


*GECCO '15 July 11-15, 2015, Madrid, Spain*

ACM ISBN 978-1-4503-3488-4/15/07.
DOI: http://dx.doi.org/10.1145/2739482.2764628

## 2. METHODOLOGY

We propose to model the problem of selecting suitable parameters for RBM by means of vanilla Harmony Search and some of its variants. The RBM learning step has four parameters: the learning rate $\eta$, weight decay $\lambda$, penalty parameter $\alpha$, and the number of hidden units $n$. Therefore, we have a four-dimensional search space with three real-valued variables, as well as the integer-valued number of hidden units. In regard to the experiments, we employed two datasets: MNIST[1] and CalTech 101 Silhouettes Data Set[23].

We compared the proposed HS-based RBM model selection against with the well-known PSO, a random initialization of parameters (RS), as well as against with Hyperopt library using random search (Hyper-RS) and Tree of Parzen Estimators (Hyper-TPE) [1], and a Bayesian optimization library called Spearmint [7]. Additionally, we have employed two HS variants: (i) Improved Harmony Search (IHS) [5] and (ii) Global-best Harmony Search (GHS) [6].

In order to provide a statistical analysis by means of Wilcoxon signed-rank test [8], we conducted a cross-validation with 10 runnings. In regard to the parameter configuration for each optimization technique, we have used $c_1 = 1.4$, $c_2 = 0.6$ and $w = 0.7$ for PSO, $HMCR = 0.8$, $PAR = 0.7$ and $\varrho = 0.1$ for HS, $PAR_{MIN} = 0.1$, $PAR_{MAX} = 1.0$, $\varrho_{MIN} = 0.1$ and $\varrho_{MAX} = 0.5$ for IHS. GHS has used the same $HMCR$, $PAR_{MIN}$ and $PAR_{MAX}$ values as employed in HS. Additionally, we employed 5 agents over 50 iterations for convergence considering all techniques.

Finally, we set each RBM parameter according to the following ranges: $n \in [5, 100]$, $\eta \in [0.0, 1.0]$, $\lambda \in [0.0, 1.0]$ and $\alpha \in [0.0, 0.001]$. We have employed $T = 100$ as the number of epochs for BRBM learning weights procedure.

## 3. EXPERIMENTAL RESULTS

[1]http://yann.lecun.com/exdb/mnist
[2]https://people.cs.umass.edu/~marlin/data.shtml
[3]Notice we employed a reduced version of the training set concerning MNIST dataset.

This section aims at presenting the experimental results concerning RBM model selection for binary image reconstruction. We compare nine methods for BRBM model selection, including random search, PSO, HS and some variants, Hyperopt and Spearmint. The next sections describe in details the experimental results for each dataset.

## 3.1 MNIST dataset

As aforementioned, we modeled the problem of selecting suitable parameters for BRBMs as an optimization task, being the objective function to minimize the MSE over the training set. Table 1 presents the mean MSE and its standard deviation over the training and test sets concerning the image reconstruction task over MNIST dataset Additionally, we also show the number of calls to the RBM learning procedure to give us an idea about the computational burden of each technique.

**Table 1: Experimental results concerning MNIST dataset.**

| Technique | MSE (training set) | MSE (test set) | #calls |
|---|---|---|---|
| RS | 0.079±0.048 | 0.079±0.048 | 1 |
| HS | 0.040±0.006 | 0.042±0.005 | 55 |
| **IHS** | **0.037±0.006** | **0.039±0.005** | **55** |
| GHS | 0.040±0.009 | 0.042±0.008 | 55 |
| **PSO** | **0.035±0.007** | **0.037±0.005** | **250** |
| Hyper-RS | 0.063±0.002 | 0.064±0.002 | 250 |
| Hyper-TPE | 0.050±0.002 | 0.050±0.002 | 250 |
| Spearmint | 0.047±0.001 | 0.048±0.001 | 250 |

Considering the experimental results, we may drive some conclusions here: HS-based techniques seem to be very suitable for RBM model selection, since they achieved one of the best results among all nine compared techniques, but with a lower computational burden, since they have been about 4.54 times faster than PSO, Hyperopt and Spearmint techniques considering the number of calls to the evaluation function, i.e., the RBM learning algorithm. Other conclusion concerns about IHS: it seems to be more important to update both HMCR and PAR parameters dynamically than to consider the best harmony's values when creating the new harmony memory, as employed by GHS. This can be evidenced when we compare IHS and GHS results, since the latter has obtained the worst results among all HS-based techniques, and it does not employ HMCR and PAR parameters. Therefore, it seems both HMCR and PAR play an important whole when dealing with RBM model selection.

## 3.2 Caltech 101 Silhouettes Dataset

Table 2 presents the MSE of the reconstruction procedure over the training and test sets, as well as the number of calls to the RBM learning algorithm. Based on the Wilcoxon signed-rank test, we can observe two important information: (i) firstly, all metaheuristic techniques have been statistically more accurate than random search, and (ii) PSO, HS and IHS have been considered similar to each other with respect to the MSE over the test set, although PSO has been about 4.54 times more expensive in terms of computational burden than HS-based approaches. It seems the process of employing dynamic PAR and bandwidth values by IHS appears to be a game-changing, since it has obtained the minimum reconstruction error among the HS-based approaches.

**Table 2: Experimental results concerning Caltech 101 Silhouettes dataset.**

| Technique | MSE (training set) | MSE (test set) | #calls |
|---|---|---|---|
| RS | 0.205±0.055 | 0.205±0.055 | 1 |
| **HS** | **0.130±0.025** | **0.130±0.028** | **55** |
| **IHS** | **0.122±0.016** | **0.122±0.017** | **55** |
| GHS | 0.145±0.022 | 0.146±0.024 | 55 |
| **PSO** | **0.113±0.032** | **0.111±0.032** | **250** |
| Hyper-RS | 0.163±0.010 | 0.164±0.012 | 250 |
| Hyper-TPE | 0.159±0.003 | 0.159±0.001 | 250 |
| Spearmint | 0.153±0.003 | 0.279±0.120 | 250 |

# 4. CONCLUSIONS

In this paper, we have shown the suitability in using HS-based approaches for RBM model selection in the context of image reconstruction. The experimental results over two datasets evidenced both the efficiency and effectiveness of HS-based approaches when compared to others in the literature.

# 5. REFERENCES


[1] J. S. Bergstra, D. Yamins, and D. D. Cox. Hyperopt: A python library for optimizing the hyperparameters of machine learning algorithms. In *Python for Scientific Computing Conference*, pages 1–7, 2013.

[2] Z. W. Geem. *Music-Inspired Harmony Search Algorithm: Theory and Applications*. Springer Publishing Company, Incorporated, 1st edition, 2009.

[3] G. Hinton. Training products of experts by minimizing contrastive divergence. *Neural Computation*, 14(8):1771–1800, 2002.

[4] J. Kennedy and R. Eberhart. *Swarm Intelligence*. M. Kaufman, 2001.

[5] M. Mahdavi, M. Fesanghary, and E. Damangir. An improved harmony search algorithm for solving optimization problems. *Applied Mathematics and Computation*, 188(2):1567 – 1579, 2007.

[6] M. G. Omran and M. Mahdavi. Global-best harmony search. *Applied Mathematics and Computation*, 198(2):643 – 656, 2008.

[7] J. Snoek, H. Larochelle, and R. P. Adams. Practical bayesian optimization of machine learning algorithms. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 2951–2959. Curran Associates, Inc., 2012.

[8] F. Wilcoxon. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83, 1945.


## Acknowledgments


The authors are grateful to FAPESP grants #2013/20387-7, #2014/16250-9, #2014/09125-3 and #2014/24491-6, and CNPq grants #303182/2011-3, #470571/2013-6 and #306166/2014-3.