# Review-5

## A Learning Algorithm for Boltzmann Machines

Increasing interest in parallel search and connection-based methods has motivated efficient learning algorithms. Learners aim to represent the intricate aspects of data by communicating the compatibility of entities within themselves. One such learner is the Boltzmann Machine which is trained using an energy-based objective. The work presents a novel algorithm for training Boltzmann Machines which is based on energy compatibilities between the inputs and representations. Learning is carried out in an architecture consisting of visible and hidden computational units to predict the probabilities of spatial representations. Probabilistic estimates are used to update parameters of the network. The network is tested on a number of bit encoding tasks.

Training of Boltzmann Machines is greatly hindered by computational speed and the choice of a suitable update rule. While the binary threshold rule serves as a potential candidate, it does not allow the network to be generalized to different inputs and tasks. An alternate method for training Boltzmann Machines is presented which consists of an update rule based on probabilistic inference. The update rule is obtained from relative entropy (KL-divergence) between the distributions of visible units when the network is trained and when the network is running freely. The information theoretic metric yields stable first order updates and reduced local convergence in comparison to the steepest descent method. When trained and tested with novel information theoretic metric on various binary code encoding tasks, the method demonstrates suitable performance. In the case of 40-10-40 encoder, the network achieves 98.6% success rate within 1200 cycles.

The Boltzmann Machine utilizes an energy-based objective which depicts two key benefits in comparison to steepest descent methods. Firstly, the method can approximate first and second order derivatives accurately. Secondly, the metric is an apt update based on the similarity between the two distributions for generalization. While the objective is a suitable candidate for learning, it presents two shortcomings. (1) Since the update consists of probabilistic measures, the magnitude of updates may tend to grow large. One alternative of bounding the weights is by utilizing noise clamping. However, the approach does not present significant results depicting its suitability. (2) In the case of high-dimensional tasks such as the 40-10-40 encoder, the network performs well but mostly depends on the annealing schedule to capture finer aspects of representations. The dependency of the network on annealing indicates room for architectural improvement.

Learning informative representations efficiently is an open problem. The work highlights two novel components which motivate future work in this direction. The architecture employs an information-based objective which solely depends on probabilistic estimates for parameter updates. While steepest descent methods have been known to work well, probabilistic learning reduces the need for complex objectives. Additionally, local representations in connection-based methods motivate the need for parallelized learning in the network. This would aid in learning complex concepts by distributing the spatial data as a means of efficient hardware.

Advances in learning representations have significantly advanced the speed and efficiency of Boltzmann Machines. The work proposes one such method which makes use of an information-theoretic objective based on probabilities estimated using a Boltzmann distribution. The update rule is theoretically compared to steepest descent methods on the basis of noise and approximation errors. Empirically, learning of the Boltzmann Machine utilizing the method demonstrates suitable results on various binar encoding tasks. The algorithm serves as one of the first practical examples of learning representations using generative techniques.