Scalable Methods for Deep Belief Learning in Very Large Datasets

James Philbin Oliver Williams Michael Isard
Microsoft Research, Silicon Valley

Deep belief networks(DBN) [1] composed from layers of restricted boltzmann machines have proven a very successful technique in machine learning, demonstrating state of the art performance in learning tasks as diverse as handwriting recognition, dimensionality reduction and the modelling of temporal sequences [1, 2, 3]. Despite their conceptual simplicity and ability to learn complex non-linearities, the significant time and space complexity of these methods has thus far restricted their use to low dimensional problem domains and small datasets.

In this work we introduce a sparse formulation of the problem which is able to exploit domain-specific locality in the inter-layer connections. This significantly reduces memory requirements and is exploited to speed up both the unsupervised contrastive divergence training and the back-propagation used for supervised fine-tuning. Additionally, we show how these modifications can be used to partition the problem for solution over a large number of machines, and allow DBNs to be used in high-dimensional problem domains with very large quantities of training data.

Specifically, we demonstrate our methods on a challenging dataset of medium-resolution video where the task is to build a temporal next-frame predictor.

References

[1] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5789):504–507, 2006.

[2] R. R. Salakhutdinov and G. E. Hinton. Semantic hashing. In *Proceedings of the SIGIR Workshop on Information Retrieval and Applications of Graphical Models*, 2007.

[3] I. Sutskever and G. E. Hinton. Learning multilevel distributed representations for high-dimensional sequences. In *AISTATS*, 2007.

Topics: learning algorithms, visual processing and pattern recognition
Preference: poster