Deep Learning : Towards Building Truly Intelligent Machines

Saurabh V. Pendse

North Carolina State University

May 9, 2013

Introduction

Deep learning refers to a class of algorithms that attempt to learn multiple levels of representation of increasing complexity or abstraction.

- Model the structure of biological learning systems, like the mammalian brain.
- Switch features of the brain into a learning model which can :
  1. deal with high dimensional data
  2. support fast learning algorithms
  3. give the system a lot of data 'so it can discover by itself what some of the concepts in the world are'.

Core Problems in AI

"Finding a good representation of the massive amount of knowledge about the world is hard enough, it is compounded by the need to efficiently extract contextually relevant knowledge depending on the situation."

-Jeff Hawkins, Redwood Neuroscience Center

Two of the core problems in AI :

- Knowledge Representation
- Learning

Motivation I

1. **Depth and Performance**
   - Relationship between architecture depth and performance.
   - Lower depth increases architecture complexity.
2. **Architecture of the Brain**
   - The most advanced learning system in existence.
   - The brain has a deep architecture.
   - The neocortex is remarkably regular & hierarchical.
3. **Nature of Cognitive Processes**
   - Humans organize their ideas and concepts hierarchically.
   - Simpler concepts are composed into abstract ones.

History

Perceptrons (Rosenblatt, 1960 : Artificial brain)

- One hand-crafted feature layer. Recognition done by learning weight vectors combining all the features.
- Single layer structure limited the class of learnable functions.

Neural Networks (Hinton, 1985)

- Original fixed feature layer replaced by several hidden layers.
- Could learn more complicated functions : Backpropagation.

Support Vector Machines (Vapnik, 1995)

- Replace the hand-crafted feature layer in Perceptrons into a feature layer generated using the kernel trick.
- Works well when the data has a simple structure.
- Choice of kernel function dictates performance.

Breakthrough

Saurabh V. Pendse

Geoffery Hinton describes SVMs as *"a temporary digression"*.

Before 2006, attempts at training deep architectures failed : yielded worse results than shallow ones (with 1 or two hidden layers).

Overcoming the drawbacks of backpropagation :

- Keep the efficiency and simplicity of backpropagation for adjusting the weights, but use it for modeling the structure of the sensory input.
- Learn $p(input)$ and not $p(label|input)$.

If you want to do computer vision, first learn computer graphics.

Restricted Boltzmann Machines (RBMs) I

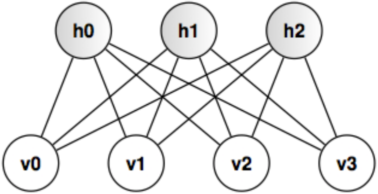

Figure : RBM - A generative stochastic neural network (Hinton, 2006)

- Consists of one visible layer and one hidden layer.
- Neurons are binary stochastic bernoulli variables.
- No lateral connectivity between neurons.
- Hidden units are conditionally independent given the visible states.

Advantage: Quickly get an unbiased sample from the posterior distribution when given a data-vector.

Restricted Boltzmann Machines (RBMs) II

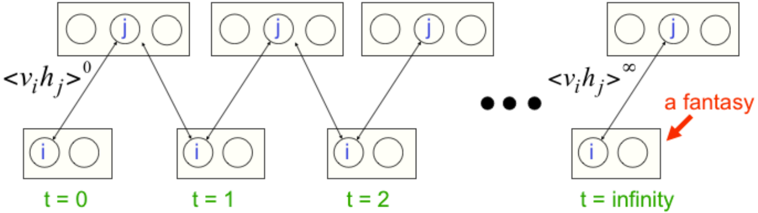

Figure : Maximum likelihood learning algorithm for a RBM

- Start with a training vector on the visible units.
- Alternate between updating all the hidden units in parallel and updating all the visible units in parallel.

$$\frac{\partial \log p(v)}{\partial w_{ij}} = < v_i h_j >^0 - < v_i h_j >^\infty$$

The Pre-training Principle (DBNs)

Layer-by-layer pre-training algorithm (Hinton, 2006)

- Each layer in the deep network is pre-trained in a completely unsupervised manner.
- The layers are stacked one on top of other. The output representation at each layer is the input for the next layer.
- Supervised training is used on the entire network to fine-tune all the layers.

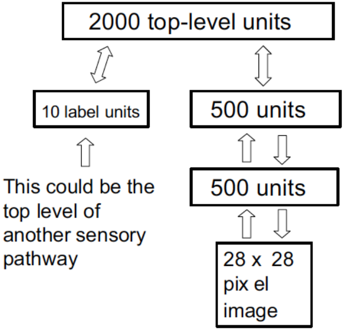

Figure : A Deep Belief Network formed by stacking RBM units.

Model a joint distribution of the images and their labels.

Other Deep Architectures I

Autoencoders

- Similar in functional form to RBMs. However, their interpretation and procedures for training them are considerably different.
- A denoising criterion is often used for regularization (*manifold assumption*).
- Performs equal or better than RBMs on classification tasks.

Convolutional Neural Networks

- Variants of Multi-Layer Perceptrons which draw inspiration from biology.
- Exploit spatial correlation in the data by enforcing local connectivity between neurons.
- Spatial invariance achieved higher up in the hierarchy. (Typically used for vision problems.)

Other Deep Architectures II

Hierarchical Temporal Memory

- Modeled very closely on the mammalian neo-cortex.
- Comprises of a tree-shaped hierarchy composed of nodes.
- It is a memory system that learns in space and time.
- Requires time-varying data, and relies on storing a large set of patterns and sequences.

Sparse Coding Methods & Sparse Distributed Representations (SDRs)

- SDR - a mathematical model for the human long-term memory.
- Sparse coding is a neural code which is based on SDRs.
- Goal is to represent input vectors as a parse approximate weighted linear combination of basis vectors.
- Learn classifier on the training set defined on the basis vectors.

Applications

Deep learning algorithms have been applied to a variety of domains, such as :

- **Vision** - outperformed traditional SVMs on the standard MNIST benchmark (Hinton 2007, Bengio2008, Vincent2010).
- **Information Retrieval** - Semantic hashing (Salakhutdinov2009) performs document similarity search in a time that is independent of the document collection.
- **Automatic Speech Recognition** - DBNs outperform state-of-the-art ASR systems based on HMMs (Mohamed2009).
- **Natural Language Processing** - CNNs have been used to define a unified architecture for NLP that learns features that are relevant to the tasks at hand given very limited prior knowledge (Collobert2008).

Deep learning algorithms significantly outperform traditional

Challenges

Existing deep learning models are still far from being as capable as the human brain.

- Existing models cannot cope with new tasks for which they have not been specifically trained.
- Scalability issues.
- Multiple categories (of the order of thousands).
- Contextual Relevance.
- Integrating multiple input data streams.

Trends & Future Work

- **Cognitive Computing -** a result of advances in deep learning.
  - Making machines think like humans.
  - Embedding deep learning algorithms in silicon circuitry.
  - Learning through experiences, finding correlations, creating hypotheses.

- **Analytics Systems -** Grok, from Numenta Inc. Provide Organizations with analytical expertise to handle explosive growth of data.
  - Automated Data Modeling. (Deploy and scale rapidly)
  - Adaptive learning. (Adjust to changing data streams)
  - Automated action. (Achieve peak optimization)
  - Action Intelligence. (Profit in real-time)

Conclusion

- Deep learning is a very promising field of AI that is very much in its infancy.
- Significant research progress on deep learning algorithms has been made in the past decade.
- Better theoretical understanding of brains.
- Moving closer towards the ultimate goal of building truly intelligent machines.
- Silicon-based learning systems.
  - Sensors to sample environment
  - Trained on real-world patterns
- Self-driving car, Google Brain, Siri, Vision systems, Intelligent security systems, Non humanoid robots, Next generation Microprocessors.

Thank you.