ISSN 2186-7437

NII Shonan Meeting Report

No. 2014-5

Deep Learning: Theory, Algorithms, and Applications

Pierre Baldi
Kenji Fukumizu
Thomaso Poggio

May 19-22, 2014

National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-Ku, Tokyo, Japan

Deep Learning: Theory, Algorithms, and Applications

Organizers:
Pierre Baldi (University of California in Irvine)
Kenji Fukumizu (The Institute of Statistical Mathematics)
Tomaso Poggio (Massachusetts Institute of Technology)

May 19-22, 2014

The ability to learn is essential to the survival and robustness of biological systems. There is also growing evidence that learning is essential to build robust artificial intelligent systems and solve complex problems in most application domains. Indeed, one of the success stories in computer science over the past three decades has been the emergence of machine learning and data mining algorithms as tools for solving large-scale problems in a variety of domains such as text analysis, computer vision, robotics, and bioinformatics. However, we are still far from having a complete understanding of machine learning and its role in AI, and plenty of challenges, both theoretical and practical, remain to be addressed.

Complex problems cannot be solved in one single step and often require multiple processing stages in both natural and artificial systems. For instance, visual recognition in humans is not an instantaneous process and requires activation of a hierarchy of processing stages and pathways. The same is true for all the best performing computer vision systems available today. Thus deep learning architectures, comprising multiple, adaptable, processing layers are important for the understanding and design of both natural and artificial systems and, today, are at the forefront of machine learning research. In the past year alone, deep architectures and deep learning have achieved state-of-the-art performance in many application areas ranging from computer vision, to speech recognition, to bioinformatics.

It is this recent wave of progress that provides the relevant context for the proposed meeting which will focus on all aspects of deep architectures and deep learning, with a particular emphasis on understanding fundamental principles because there is still very little theoretical understanding of deep learning, in spite of the recent progress. Thus a major thrust of the meeting will be to foster theoretical analysis of deep learning. In addition to theory, topics to be covered will include also algorithms and applications. The primary intellectual focus of the meeting will be on deep learning in artificial systems. However, deep learning draws some of its inspiration from, and has close connections to, neuroscience. Thus presentations and discussions bridging learning in natural and artificial learning systems will also be encouraged.

Schedule

Day 1 (May 19)

Welcome, Introduction, Overview, History:
(Pierre Baldi, Tomaso Poggio, and Kenji Fukumizu)

1 THEORY (Monday Morning) Chair: Kenji Fukumizu

Talks:

- Basic principles of Self-Organization and Supervised L earning (Shun-ichi Amari)
- Autoencoders for Structured Data (Alessandro Sperduti)
- Deep Targets and Dropout (Pierre Baldi)

Spotlights:

- Non Linear Dynamics of Learning in Deep Linear Networks (Surya Ganguli)
- Stochastic Optimization (Takayuki Osogami)
- Black Box and Representation Aspects of Neural Networks (Klaus-Robert Muller)

Panel: Chair and Speakers

2 NEUROBIOLOGY (Monday Afternoon) Chair: Hiroyuki Nakahara

Talks:

- Stochasticity of Biological Synapses (Erik De Schutter)
- A Mathematical Theory of Semantic Cognition (Surya Ganguli)
- Learning, Decision-Making, and Neural Coding (Hiroyuki Nakahara)

Spotlights:

- Circuits and Large Scale Architecture of the Brain (Shimon Edelman)
- Towards Modelling Cortical Response Properties Using Multilayer Models of Natural Images (Aapo Hyvrinen)
- Artificial Neural Networks versus Biological Neural Networks (Pierre Baldi)
- Brain Machine Interfaces (Klaus-Robert Muller)

Panel : Chair and Speakers

Day 2 (May 20)

3 ALGORITHMS (Tuesday Morning) Chair: Klaus-Robert Muller

Talks:

- What is the Information Content of an Algorithm? (Joachim Buhmann)
- M-Theory (Tomaso Poggio)
  Generative vs Discriminative Scaling to Big Data (Klaus-Robert Muller)

Spotlights:

- Going from Text Analysis to Structural Analysis (Michal Rosen Zvi)
- Dropout (Sepp Hochreiter)
- Long-Short Term Memory Units (Sepp Hochreiter)
- Evolutionary/Developmental Neural Networks, Because That's What Worked For Us (Eric Mjoslness)

Panel: Chair and Speakers

4 SYMBOLIC/SEMANTIC VS STATISTICS/CONNECTIONIST (Tuesday Afternoon) Chair: Paolo Frasconi

Talks:

- Languages for Machine Learning: What Role for Neural Networks? (Paolo Frasconi)
- Cognitive Architectures, Expressible as or Hybridized With Neural Networks and/or Graphical Models (Eric Mjolsness)
- A Design for a Brain? (Shimon Edelman)

SPECIAL SESSION ON OPEN PROBLEMS (Paolo Frasconi, Shimon Edelman, Eric Mjolsness)

Day 3 (May 21)

5 APPLICATIONS (Wednesday Morning) Chair: Pierre Baldi

Talks:

- Applications in Physics, Chemoinformatics, and Bioinformatics (Pierre Baldi)
- Applications in Genetics and Quantum Chemistry (Klaus-Robert Muller)
- Big Data in Neuroscience (Joachim Buhmann)

Spotlights:

- Bayesian Optimization in Materials Informatics (Koji Tsuda)
- Applications to Health Care (Michal Rosen Zvi)

- Decoding EEG, MEG, and EMG Data: Three-way Analysis and Deep Learning (Aapo Hyvrinen)
- Large Scale Identification of Brain Cells (Paolo Frasconi)

Panel: Chair and Speakers

6 EXCURSION and OPEN PROBLEMS (Wednesday Afternoon)

Day 4 (May 22)

7 APPLICATIONS (Thursday Morning) Chair: Takio Kurita

Talks:

- What is Learned by Convolutional Networks for Image Recognition Tasks? (Takayuki Okatani)
- Recent developments of Deep Learning in Natural Language Processing (Kevin Duh)
- Limitation of Neural Network Approach in Natural Language Processing (Kunihiko Sadamasa)

Spotlights:

- Expected Applications of Deep Learning (Hideki Asoh)
- Deep Density-Ratio Estimation (Masashi Sugiyama)
- Wasserstein Means: Efficient Detection of Invariances with Optimal Transport (Marco Cuturi)

Panel: Chair and Speakers

Overview of Talks

Basic principles of Self-Organization and Supervised Learning

Shun-ichi Amari, RIKEN Brain Science Institute

Deep learning uses concatenation of layers of neural networks, which are trained by both unsupervised (self-organization) and supervised learning. Lots of techniques (tricks) are used to make deep learning effective. In order to understand the basic principles of learning between two consecutive layers, we analyze simple models of 1) self-organization with no recurrent connections, 2) RBM learning and 3) recurrently connected neural networks. We show the common principle included in them as well as their differences. In particular, we analyze the Gaussian Boltzmann machine and prove that the contrast divergence minimization has the same performance as the original machine.

As for supervised back-prop learning, the singularity existing in the parameter space is shown to be a major obstacle for learning. We analyze its dynamics in a neighborhood of singularity. We show that the natural gradient learning method is free from such difficulties.

Autoencoders for Structured Data

Alessandro Sperduti, Universita degli Studi di Padova

We start by introducing linear autoencoder networks for structured inputs, such as sequences, trees, and graphs. We show that the problem of training an autoencoder has a closed form solution which can be obtained via the definition of a linear dynamical system modelling the structural information present in the dataset of structures. Relationship with principal directions is discussed. We then present an application to pre-training of Recurrent Neural Networks.

Stochastic Optimization

Takayuki Osogami, IBM Research, Tokyo

We extend the multinomial logit model (MLM) to represent some of the empirical phenomena that are frequently observed in the choice made by humans. These phenomena include the similarity effect, the attraction effect, and the compromise effect. We formally quantify the strength of these phenomena that can be represented by our choice model, which illuminates the flexibility of our choice model. We then show that our choice model can be represented as a restricted Boltzmann machine and that its parameters can be trained effectively from data. Our numerical experiments suggest that we can train the parameters of our choice model in such a way that it represents the typical phenomena of choice.

Spike-timing dependent plasticity (STDP), which is first postulated and later confirmed experimentally, is considered as a fundamental mechanism for learning. However, this phenomena has little theoretical underpinning so far. Using a Boltzmann machine extended to a temporal domain, we show the homeostatic

STDP rule can be derived from a single objective function called a contrastive free energy. Several sequence of patterns can be stored in a system and later retrieved with appropriate cue presentations. Our model gave the theoretical underpinning for STDP as the Boltzmann machine did for the Hebb's rule.

A part of this research was supported by JST, CREST.

Black Box and Representation Aspects of Neural Networks

Klaus-Robert Muller, TU Berlin

I first remind the audience of an earlier work by Braun et al 2008, relevant dimension estimation (RDE) (for a recent extension see Montavon et al 2013). RDE allows to quantify the dimension and noise of kernel methods.

We apply RDE to better understand the representation that trained deep networks implement (Montavon et al 2011).

Finally I briefly discuss ongoing work how to deconstruct deep networks in order to arrive at pixel-wise explainations (Bach et al. submitted).

Learning, Decision-Making, and Neural Coding

Hiroyuki Nakahara, RIKEN Brain Science Institute

I suggest that neural computations use different kinds of learning at multiple depths in various forms, i.e., deep learnings. The first example is our recent study. Using reinforcement learning framework as a powerful approach to develop quantitative understanding of social decision making. we recently investigated neural computations for learning to simulate others' decisions (Suzuki et al 2012). Also, I briefly mention our past studies on dopamine-guided learning, i.e., parallel learning for skills (Nakahara et al 2001), and representational learning combined with reinforcement learning (Nakahara and Hikosaka 2012; Nakahara 2014).

Nakahara et al 2001. J. Cog Neurosci 13(5): 626-647; Nakahara and Hikosaka 2012. Neurosci Res 74(3-4): 177-183; Suzuki et al 2012 Neuron 74: 1125-1137; Nakahara 2014 Current Opinion in Neurobiology 25: 123-129.

Towards Modelling Cortical Response Properties Using Multilayer Models of Natural Images

Aapo Hyvärinen, University of Helsinki

An important property of visual systems is to be simultaneously both selective to specific patterns found in the sensory input and invariant to possible variations. Selectivity and invariance (tolerance) are opposing requirements. It has been suggested that they could be joined by iterating a sequence of elementary selectivity and tolerance computations. It is, however, unknown what should be selected or tolerated at each level of the hierarchy. We approach this issue by learning the computations from natural images. We propose and estimate a probabilistic model of natural images that consists of three processing layers. Two natural image data sets are considered: image patches, and complete visual scenes downsampled to the size of small patches. For both data sets, we find that in the first two layers, simple and complex cell-like computations are performed. In the third layer, we mainly find selectivity to longer contours; for patch data, we further find some selectivity to texture, while for the down-sampled complete scenes, some selectivity to curvature is observed. Based on our article in J Physiology (Paris), 2013.

Brain Machine Interfaces

Klaus-Robert Muller, TU Berlin

I briefly introduce the field of Brain Computer Interfacing and discuss statistical estimation problems involved. In particular my focus is on the classification of event related potentials. Here shrinkage estimation has been highly successful. I point out the limits of analytic shrinkage of Ledoit and Wolf 2004, and introduce the novel extension of Bartz et al 2003 and demonstrate its usefulness in BCI.
Further reading:

Blankertz, B., Lemm, S., Treder, M., Haufe, S., Muller, K.-R., Single-trial analysis and classification of ERP components - A tutorial, Neuroimage, 56 (2), 814-825 (2011)

Bartz, D., Muller K.-R., Generalizing Analytic Shrinkage to Arbitrary Covariance Structures, in Advances in Neural Information Processing Systems 26 (NIPS), (eds.) L. Bottou, C.J.C. Burges, M. Welling, Z. Ghahramani and K.Q. Weinberger (2013)

Design for a Brain?

Shimon Edelman, Cornell University

Two approaches to neural computation, deep networks (DN) and reinforcement learning (RL), have recently led to impressive advances both in applied fields such as vision and robotics and in theoretical thinking about the neurobiology of corresponding cognitive faculties. The successes of DN, in particular, have revived the expectations that human-like artificial intelligence is just around the corner. I examine such claims critically in two ways: first, by comparing the tasks at which DN and RL excel to those that arise from real sequential behaviors such as foraging and language; and second, by comparing the architecture of DN and RL models to the real neuroanatomy behind those behaviors. I then use lessons drawn from these comparisons to outline what may be a more promising approach to the understanding and modeling of the computational nature and the brain basis of complex sequential behaviors.

Applications in Genetics and Quantum Chemistry

Klaus-Robert Muller, TU Berlin

The talk presents a recent line of research of usage of Machine Learning for atomistic simulations. We demonstrate that kernel ridge regression as well as deep networks can achieve remarkable results achieving up to chemical accuracy at a speed up of approx 10^6. In addition I report on recent results in materials science, where DOS at Fermi energy is predicted.

Deep Density-Ratio Estimation

Masashi Sugiyama, Tokyo Institute of Technology

Density ratios have been demonstrated to be useful in solving various machine learning tasks such as non-stationarity adaptation, outlier detection, feature extraction, and conditional density estimation. In this talk, I will overview the framework of density ratio estimation and explain how deep learning can be utilzed in this scenario.

Decoding EEG, MEG, and EMG Data: Three-way Analysis and Deep Learning

Aapo Hyvärinen, University of Helsinki

The talks discussed possible application of deep learning in analysis of biomedical signals characterized by rich spectral properties. Technical results were from our paper in NeuroImage (2013), whose abstract is as follows: We propose a new data-driven decoding method called Spectral Linear Discriminant Analysis (Spectral LDA) for the analysis of MEG, EEG, and EMG.The method allows investigation of changes in rhythmic neural activity as a result of different stimuli and tasks. The introduced classification model only assumes that each gbrain stateh can be characterized as a combination of neural sources, each of which shows rhythmic activity at one or several frequency bands.

Furthermore, the model allows the oscillation frequencies to be different for each such state. We present decoding results from 9 subjects in a four-category classification problem defined by an experiment involving randomly alternating epochs of auditory, visual and tactile stimuli interspersed with rest periods. The performance of Spectral LDA was very competitive compared with four alternative classifiers based on different assumptions concerning the organization of rhythmic brain activity. In addition, the spectral and spatial patterns extracted automatically on the basis of trained classifiers showed that Spectral LDA offers a novel and interesting way of analyzing spectrospatial oscillatory neural activity across the brain. All the presented classification methods and visualization tools are freely available as a Matlab toolbox.

What is Learned by Convolutional Networks for Image Recognition Tasks?

Takayuki Okatani, Tohoku University

In this talk I present three experiments that we have conducted hoping to lead to better understanding of why convolutional neural networks (CNNs) work so well for many visual recognition tasks. An intuitive, widely accepted explanation is that CNNs can better deal with global shapes of objects than previous BoF-based approaches. I show two experimental results supporting this explanation. We show that CNNs do not (appear to) perform very well for material recognition, for which the global shape of an object does not play a key role. We also show that recognition performance does deteriorate as global shape information is eliminated by randomly shuffling patches to corrupt images.

Toward better understanding of CNNs, we also show that CNNs do learn good internal representation of images; using it as category-level attributes in a zero-shot learning setting yields accuracy surpassing the state-of-the-art.

Recent developments of Deep Learning in Natural Language Processing

Kevin Duh, Nara Institute of Science and Technology

Two strands of deep learning research are popular in Natural Language Processing. The first focuses on end-to-end prediction tasks; the second explores how semantics might be learned by neural networks. I will survey the main challenges of learning semantics, mainly how to learn semantics of words and how to model composition. While current methods usually assume vector representation of words and simple neural network operators for composition, many open problems remain.

Expected Applications of Deep Learning

Hideki Asoh, National Institute of Advanced Industrial Science and Technology

Deep learning is successfully being applied in many areas such as image recognition, speech recognition, natural language processing, bio-informatics, chamo-informatics, material-informatics, high energy physics etc. However almost all the problems which they treat are limited to classification and prediction. In this talk, I suggest two other expected directions of application of deep learning: 1) reinforcement learning with state representation learning, and 2) modeling acquisition and usage of natural language based on hidden internal representation of sensory-motor information (Bayesian Linguistics).

Wasserstein Means: Efficient Detection of Invariances with Optimal Transport

Marco Cuturi, Kyoto University

We describe in this talk recent developments in the fast computation of Wasserstein distances between empirical measures and show the application of this technique to the numerical resolution of the Wasserstein barycenter problem as described by Agueh and Carlier.