The Science and the Engineering of Intelligence

Tomaso Poggio,
The MIT Quest
Center for Brains, Minds & Machines,
McGovern Institute for Brain Research,
Computer Science and Artificial Intelligence Agency,
Brain and Cognitive Sciences, MIT
The CBMM-FLAB partnership


Overview

- Motivations: the greatest problem in science, CBMM, the MIT Quest
- A bit of history: Neuroscience and AI, Science and Engineering
- CBMM and the Quest
- AI ethics and its neural bases
- Theory: explaining how deep networks work and what are their properties and limitations.
CBMM’s focus is the Science and the Engineering of Intelligence

We aim to make progress in understanding intelligence, that is in understanding how the brain makes the mind, how the brain works and how to build intelligent machines.
The Center for Brains, Minds and Machines (CBMM) is a multi-institutional NSF Science and Technology Center dedicated to the study of intelligence - how the brain produces intelligent behavior and how we may be able to replicate intelligence in machines. We believe in the synergy between the science and the engineering of intelligence.
Academic and Corporate Partners

IIT
Cingolani

A*star
Chuan Poh Lim

Hebrew U.
Weiss

MPI
Büthoff

Genoa U.
Verri, Rosasco

Weizmann
Ullman

City U. HK
Smale

Google

IBM

Microsoft

Orcam

Siemens

Fujitsu

DeepMind

Honda

Boston Dynamics

NVIDIA

Schlumberger

Mobileye

Intel
EAC Members & Meetings

Demis Hassabis, DeepMind
Charles Isbell, Jr., Georgia Tech
Christof Koch, Allen Institute
Fei-Fei Li, Stanford
Lore McGovern, MIBR, MIT
Joel Oppenheim, NYU
Pietro Perona, Caltech
Marc Raibert, Boston Dynamics
Judith Richter, Medinol
Kobi Richter, Medinol
Dan Rockmore, Dartmouth
Amnon Shashua, Mobileye
David Siegel, Two Sigma
Susan Whitehead, MIT Corporation
Brains, Minds & Machines Summer Course
An intensive three-week course gives advanced students a “deep” introduction to the problem of intelligence

A community of scholars is being formed
Foraging connections between human and machine intelligence research, its applications, and its bearing on society.

The MIT Intelligence Quest will advance the science and engineering of both human and machine intelligence. Launched on February 1, 2018, this effort seeks to discover the foundations of human intelligence and drive the development of technological tools that can positively influence virtually every aspect of society. The Institute's culture of collaboration …we seek to answer the deepest questions about intelligence.
Intelligence: The MIT Quest

CORE: Cutting-Edge Research on the Science + Engineering of Intelligence

Natural Science of Intelligence
- Nobel prize

The Intersection

Engineering of Intelligence
- Turing Award, Fields Medal
Overview

- Motivations: the greatest problem in science, CBMM, the MIT Quest
- A bit of history: Neuroscience and AI, Science and Engineering
- CBMM and the Quest
- AI ethics and its neural bases
- Theory: explaining how deep networks work and what are their properties and limitations.
Logical for MIT...


MIT Intelligence Initiative

CENTER FOR Brains Minds + Machines

Intelligence: The MIT Quest

2008 2012 - 2013 2018
Just a definition: I use the word science to mean natural science.
CBMM’s focus is the natural **Science** and the **Engineering** of Intelligence

We aim to make progress in understanding intelligence, that is in understanding how the brain makes the mind, how the brain works and how to build intelligent machines. We believe that the science of intelligence will enable better engineering of intelligence.
Neuroscience-Inspired Artificial Intelligence

Demis Hassabis,1,2,* Dharshan Kumaran,1,3 Christopher Summerfield,1,4 and Matthew Botvinick1,2

1DeepMind, 5 New Street Square, London, UK
2Gatsby Computational Neuroscience Unit, 25 Howland Street, London, UK
3Institute of Cognitive Neuroscience, University College London, 17 Queen Square, London, UK
4Department of Experimental Psychology, University of Oxford, Oxford, UK
*Correspondence: dhcontact@google.com
http://dx.doi.org/10.1016/j.neuron.2017.06.011

The fields of neuroscience and artificial intelligence (AI) have a long and intertwined history. In more recent times, however, communication and collaboration between the two fields has become less commonplace. In this article, we argue that better understanding biological brains could play a vital role in building intelligent machines. We survey historical interactions between the AI and neuroscience fields and emphasize current advances in AI that have been inspired by the study of neural computation in humans and other animals. We conclude by highlighting shared themes that may be key for advancing future research in both fields.

The successful transfer of insights gained from neuroscience to the development of AI algorithms is critically dependent on the interaction between researchers working in both these fields, with insights often developing through a continual hand- ing back and forth of ideas between fields. In the future, we
Two Main Recent Success Stories in AI
Demis Hassabis, master of the new machine age

Mural Ahmed

The creator of the AI game-playing program makes all the right moves, writes Mural Ahmed

The victories have a human mastermind in Demis Hassabis, co-founder and chief executive of DeepMind. He describes Mr Lau as the “Roger Federer of Go”, and for some the computer program’s achievement is akin to a robot taking to the laws of Wimbleden and beating the legendary tennis champion.

“I think it is pretty huge but, ultimately, it will be for

Washington Post, 2016-02-01

The US Federal Trade Commission’s recently updated report on privacy was released on February 11th. It is the first such report since 1999 and is intended to provide a comprehensive overview of the current state of US privacy law. The report examines the evolution of privacy laws and regulations in the United States, including the impact of technology on privacy, and the role of the government in protecting privacy.

The report begins with an overview of the history of privacy law in the US, tracing its development from the early 20th century to the present day. It notes that privacy law has evolved significantly over this period, with the emergence of new technologies and the increased digitization of personal information.

The report then examines the various federal and state laws that govern privacy in the US. It highlights the importance of the Fair Credit Reporting Act (FCRA), which regulates the use of credit reports and credit scores, and the Children’s Online Privacy Protection Act (COPPA), which aims to protect the privacy of children online.

The report also looks at the role of the government in protecting privacy, including the efforts of the Federal Trade Commission (FTC) to enforce privacy laws and regulations. It notes that the FTC has taken enforcement action in a number of high-profile cases, including the settlement with Facebook in 2011 over its handling of user data.

Finally, the report considers the future of privacy law in the US, and the challenges and opportunities that lie ahead. It concludes that privacy law is an area of ongoing evolution, and that it will be important for policymakers to keep pace with technological advances and the changing social landscape.
Two Main Recent Success Stories in AI
DL and RL come from neuroscience

RECEPTIVE FIELDS AND FUNCTIONAL ARCHITECTURE IN TWO NONSTRIATE VISUAL AREAS (18 AND 19) OF THE CAT

DAVID H. HUBEL AND TORSTEN N. WIESEL
Neurophysiology Laboratory, Department of Pharmacology, Harvard Medical School, Boston, Massachusetts
(Received for publication August 24, 1964)
Background: State-of-the-art Machines ("Deep Learning") Have Emerged From the Brain's Visual Processing Architecture

Brains / Minds
(ventral visual stream)

What's the engineering of the future?

Machines

State of the Art ResNets

Desimone & Ungerleider 1989; vanEssen+Movshon

NSF Site Visit, May 15-16, 2017
"Hubel-Wiesel" models include

Hubel & Wiesel, 1959; Fukushima, 1980; Wallis & Rolls, 1997; Mel, 1997; LeCun et al., 1998; Riesenhuber & Poggio, 1999; Thorpe, 2002; Ullman et al., 2002; Wersing and Koerner, 2003; Serre et al., 2007; Freeman and Simoncelli, 2011....
Overview

• Motivations: the greatest problem in science, CBMM, the MIT Quest

• A bit of history: Neuroscience and AI, Science and Engineering

• CBMM and the Quest

• AI ethics and its neural bases

• Theory: explaining how deep networks work and what are their properties and limitations.
Visual intelligence, video ergo sum
Visual Intelligence

MODULE ONE

Visual Stream
FOVEA, DEEP FEED-FORWARD NETWORKS, BACK PROJECTIONS

T. Poggio, J. DiCarlo, M. Livingstone, S. Ullman

MODULE TWO

Brain OS
WORKING MEMORY, VISUAL ROUTINES, ATTENTION

G. Kreiman, M. Wilson, B. Desimone

3½D Sketch

MODULE THREE

Cognitive Core
INTUITIVE PHYSICS, GEOMETRY (3D), INTUITIVE PSYCHOLOGY

J. Tenenbaum, N. Kanwisher, Spelke

Towards Symbols
LONG-TERM PLANNING, ABSTRACTION, LANGUAGE

B. Katz, A. Barbu, S. Ullman, J. Tenenbaum

MODULE FOUR

Depository of vision routines...
synthesizing routines as needed
Within The CORE Intersection: CBMM + additional “moonshot” projects

- Visual Intelligence (CBMM)
- Development of Intelligence
- New circuits for deep nets in counter streams in cortical areas
- Planning and imagination
- Emotional Intelligence
- Language
Overview

• Motivations: the greatest problem in science, CBMM, the MIT Quest

• A bit of history: Neuroscience and AI, Science and Engineering

• CBMM and the Quest

• AI ethics and its neural bases

• Theory: explaining how deep networks work and what are their properties and limitations.
AI and ethics

• Too much about
  - AI more dangerous than nuclear bombs
  - the trolley problem

• More pressing issues:
  - What to publish/not publish
  - Jobs lost to machines

• Future:
  - how to build ethical machines
  - can the brain teach us how?
Studies with fMRI revealed that particular areas of the brain are associated with particular cognitive events such as our moral emotions and ethical reasoning.
Overview

• Motivations: the greatest problem in science, CBMM, the MIT Quest

• A bit of history: Neuroscience and AI, Science and Engineering

• CBMM and the Quest

• AI ethics and its neural bases

• Theory: explaining how deep networks work and what are their properties and limitations.
Computation in a neural net

\[ f(x) = f_L(\ldots f_2(f_1(x))) \]
Theories of Deep Learning (STATS 385)

Stanford University, Fall 2017

The spectacular recent successes of deep learning are purely empirical. Nevertheless intellectuals always try to explain important developments theoretically. In this literature course we will review recent work of Bruna and Mallat, Mhaskar and Poggio, Papan and Elad, Bolcskei and co-authors, Baraniuk and co-authors, and others, seeking to build theoretical frameworks deriving deep networks as consequences. After initial background lectures, we will have some of the authors presenting lectures on specific papers. This course meets once weekly.

Instructors:

David Donoho  Hatef Monajemi  Vardan Papyan
Deep nets: a theory is needed
(after alchemy, chemistry)
Computation in a neural net

Rectified linear unit (ReLU)

\[ g(y) = \max(0, y) \]
Deep nets architecture and SGD training

Rectified linear unit (ReLU)

\[ g(y) = \max(0, y) \]

\[ L(w) \]
Gradient descent

\[ \argmin_w \sum_i \ell(z_i, f(x_i; w)) = L(w) \]

One iteration of gradient descent:

\[ w^{t+1} = w^t - \eta_t \frac{\partial L(w^t)}{\partial w} \]

learning rate
DLNNs: three main scientific questions

Approximation theory: when and why are deep networks better - no curse of dimensionality — than shallow networks?

Optimization: what is the landscape of the empirical risk?

Generalization by SGD: how can overparametrized networks generalize?
Theory I:
Why and when are deep networks better than shallow networks?

\[ f(x_1, x_2, ..., x_8) = g_3(g_{21}(g_{11}(x_1, x_2), g_{12}(x_3, x_4))g_{22}(g_{11}(x_5, x_6), g_{12}(x_7, x_8))) \]

\[ g(x) = \sum_{i=1}^{r} c_i |<w_i, x> + b_i|_+ \]

Suppose that a function of \( d \) variables is compositional. Both shallow and deep network can approximate \( f \) equally well. The number of parameters of the shallow network depends exponentially on \( d \) as \( O(\varepsilon^{-d}) \) with the dimension whereas for the deep network dance is dimension independent, i.e. \( O(\varepsilon^{-2}) \)

Mhaskar, Poggio, Liao, 2016
When can the curse of dimensionality be avoided
When is deep better than shallow

\[ f(x_1, x_2, \ldots, x_8) = g_3(g_{21}(g_{11}(x_1, x_2), g_{12}(x_3, x_4)), g_{22}(g_{11}(x_5, x_6), g_{12}(x_7, x_8))) \]
Microstructure of compositionality

target function

approximating function/network
Hierarchically local compositionality

\[ f(x_1, x_2, ..., x_8) = g_3(g_{21}(g_{11}(x_1, x_2), g_{12}(x_3, x_4))g_{22}(g_{11}(x_5, x_6), g_{12}(x_7, x_8))) \]

**Theorem (informal statement)**

Suppose that a function of \( d \) variables is hierarchically, locally, compositional. Both shallow and deep networks can approximate \( f \) equally well. The number of parameters of the shallow network depends exponentially on \( d \) as \( O(\varepsilon^{-d}) \) with the dimension whereas for the deep network dance is \( O(d\varepsilon^{-2}) \).
Binary Tree NN vs Shallow NN 8D
Locality of constituent functions is key **not** weight sharing: CIFAR
Why are compositional functions important?

Which one of these reasons:

Physics (Max Tegmark)?
Neuroscience? <= tp
Evolution?

Locality of Computation

What is special about locality of computation?

Locality in “space”? 
Locality in “time”?
Theorem (informal statement)

Replacing the RELUs with univariate polynomial approximation, Bezout theorem implies that the system of polynomial equations corresponding to zero empirical error has a very large number of degenerate solutions. The global zero-minimizers correspond to flat minima in many dimensions (generically unlike local minima). Thus SGD is biased towards finding global minima of the empirical risk.
Observation (theory and experiment): deep polynomial networks show same puzzles as RELU nets

Poggio et al., 2017
Theory of Deep Learning IIb: Optimization Properties of SGD

by

Chiyuan Zhang\textsuperscript{1} Qianli Liao\textsuperscript{1} Alexander Rakhlin\textsuperscript{2} Brando Miranda\textsuperscript{1} Noah Golowich\textsuperscript{1} Tomaso Poggio\textsuperscript{1}

\textsuperscript{1}Center for Brains, Minds, and Machines, McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA, 02139. \\
\textsuperscript{2}University of Pennsylvania
The set of polynomial equations above with $k = \text{degree of } p(x)$ has a number of distinct zeros (counting points at infinity, using projective space, assigning an appropriate multiplicity to each intersection point, and excluding degenerate cases) equal to $Z = k^n$ the product of the degrees of each of the equations. As in the linear case, when the system of equations is underdetermined – as many equations as data points but more unknowns (the weights) – the theorem says that there are an infinite number of global minima, under the form of $Z$ regions of zero empirical error.
Global and local zeros

\[ f(x_i) - y_i = 0 \quad \text{for } i = 1, \ldots, n \]
\[ \nabla_w \sum_{i=1}^{N} (f(x_i) - y_i)^2 = 0 \]

\( n \) equations in \( W \) unknowns with \( W >> n \)

\( W \) equations in \( W \) unknowns

There are a very large number of zero-error minima which are highly degenerate unlike the local non-zero minima.
Langevin equation

\[ \frac{dw}{dt} = -\gamma_t \nabla V(w(t), z(t)) + \gamma_t' dB(t) \]

with the Boltzmann equation as asymptotic “solution”

\[ p(w) \sim \frac{1}{Z} = e^{-\frac{V(w)}{T}} \]
When is deep better than shallow

\[ f_{t+1} = f_t - \gamma_t \nabla V(f_t, z_t), \quad \nabla V(f_t, z_t) = \frac{1}{|z_t|} \sum_{z \in z_t} \nabla V(f_t, z). \]

We define a noise “equivalent quantity”

\[ \xi_t = \nabla V(f_t, z_t) - \nabla I_{S_n}(f_t), \]

and it is clear that \( \mathbb{E} \xi_t = 0. \)

We write Equation 6 as

\[ f_{t+1} = f_t - \gamma_t (\nabla I_{S_n}(f_t) + \xi_t). \]
GDL selects larger volume minima
GDL ~ SGD (empirically)
GDL selects degenerate minima
Concentration because of high dimensionality

Poggio, Rakhlin, Golovitc, Zhang, Liao, 2017
When is deep better than shallow?

- There are many zero minimizers with overparametrized deep networks because of Bezout theorem.

- SGDL finds with very high probability large volume, flat zero-minimizers; empirically SGD behaves in a similar way.

- Flat minimizers correspond to degenerate zero-minimizers and thus to global minimizers.

---

**SGDL and SGD observation: summary**

Poggio, Rakhlin, Golovitc, Zhang, Liao, 2017
Theory III: How can underconstrained solutions generalize?

Model #params: 9370

![Diagram showing the relationship between the number of training examples and error on CIFAR-10. The graph plots the training error, test error, and the test-training error difference. The x-axis represents the number of training examples, and the y-axis represents the error on CIFAR-10. The model parameters are 9370.](image.png)
Classical Generalization Bounds

With probability $\geq (1 - \delta) \\forall f$

$$|E(\ell) - E_S(\ell)| \leq C_{N,\delta},$$
Very good generalization!

MNIST with different initializations
Three Theory Questions: Summary of Answers

• **Approximation theorems**: for compositional functions deep but not shallow networks avoid the *curse of dimensionality*.

• **Optimization remarks**: SGD finds with high probability global minima which are degenerate.

• **Generalization**: The gradient dynamics of deep networks near global minima converges to minimum norm solution for each layer of weights.
Musings on Near Future Breakthroughs

• new architectures/class of applications from basic DCN block (example GAN + RL/DL + …)

• new semisupervised training framework, avoiding labels: implicit labeling…predicting next “frame”…

• new basic supervised block/circuit ?

• new learning algorithm (Shim) instead of SGD …
General musings

The evolution of computer science

• there were programmers
• there are now labelers
• there may be schools for bots…
The first phase (and successes) of ML: supervised learning, big data: $n \rightarrow \infty$

from programmers…
…to labelers…
…to computers that learn like children…

The next phase of ML: implicitly supervised learning, learning like children do, small data: $n \rightarrow 1$