Artificial Neural Networks: Recent Advances, Current Trends and Open Problems

Eduardo Bezerra
ebezerra@cefet-rj.br
09/mar/2018
Roadmap

- Recent Advances
- Current Trends
- Open Problems
Recent Advances
Backpropagation

Learning representations by back-propagating errors

David E. Rumelhart*, Geoffrey E. Hinton† & Ronald J. Williams*

* Institute for Cognitive Science, C-015, University of California, San Diego, La Jolla, California 92093, USA
† Department of Computer Science, Carnegie-Mellon University, Pittsburgh, Philadelphia 15213, USA
“We show that standard multilayer feedforward networks with as few as a single hidden layer and arbitrary bounded and nonconstant activation function are universal approximators”
“if you can understand the paper, you are better than many people in ML. It took 10 years until people understand what they were talking about”.  – Jeff Hinton

Citações: 2012: 77; 2016: 2133; 2017: 3932
LeNet

Gradient-Based Learning Applied to Document Recognition
Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner

1998
Neural Nets Renaissance

2006
Reducing the Dimensionality of Data with Neural Networks
G. E. Hinton and R. R. Salakhutdinov

2009
Learning multiple layers of representation
Geoffrey E. Hinton
Department of Computer Science, University of Toronto, 10 King’s College Road, Toronto, M5S 3G4, Canada

ImageNet: A Large-Scale Hierarchical Image Database
Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li and Li Fei-Fei
Dept. of Computer Science, Princeton University, USA
{jiadeng, wdong, rsocher, jial, li, feifei}@cs.princeton.edu

2009
Learning Deep Architectures for AI
Yoshua Bengio

2009
Large-scale Deep Unsupervised Learning using Graphics Processors
Rajat Raina
Anand Madhavan
Andrew Y. Ng
Computer Science Department, Stanford University, Stanford CA 94305 USA
Deep Learning Explosion

2010s

Also known as: Revenge of the Sith Neural Nets!
Basic Idea: Hierarchy of Features

Composition of functions across layers is a cornerstone of deep neural nets.

Credits: Andrew Ng apud The Learning Machines (Nature, 2014)
Deep Neural Networks for Acoustic Modeling in Speech Recognition

Geoffrey Hinton, Li Deng, Dong Yu, George Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara Sainath, and Brian Kingsbury
Object Recognition (AlexNet)

2012

dropout
rectified linear activation units (ReLU)

AlexNet
8 layers

ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
University of Toronto
kri@cs.utoronto.ca

Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca
Object Recognition (AlexNet)

Credits: Mathew Zeiler (Clarifai)

ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

2012
Deep Nets

- VGGNet: 19 layers
- GoogleNet: 22 layers
- ResNet: 152 layers
Deep Nets

Credits: https://medium.com/@Lidinwise/the-revolution-of-depth-facf174924f5
Semantic Segmentation

Fully Convolutional Networks for Semantic Segmentation

Jonathan Long*  Evan Shelhamer*  Trevor Darrell
UC Berkeley
{jonlong,shelhamer,trevor}@cs.berkeley.edu
Generative Adversarial Nets

2014
Deep Reinforcement Learning

2015

AlphaGo
Success Factors

- More computer power (GPUs)
- More Data Available
- New training techniques & methods
- Tons of $
“What was wrong in the 80’s is that we didn’t have enough data and we didn’t have enough computer power”

Geoffrey Hinton
The Connectionist Invasion
Current Trends
Multimodal Learning

A Bimodal Learning Approach to Assist Multi-sensory Effects Synchronization
R. Abreu, J. Santos, E. Bezerra, submitted to IJCNN2018
Multimodal Learning

A close up of a hot dog on a bun.

A bathroom with a toilet and a bath tub.

A vase filled with flowers sitting on a table.

I promise not to use Google Translator to do my homeworks.

Eu prometo não usar o Google Translator para fazer minhas lições de casa.
Learning to Learn (Meta learning)

Welcome to the Hyperparameter Jungle!

- Optimization technique
- Amount of epochs
- Amount of layers, hidden units
- Type of each layer, type of activation function
- Cost function, regularization technique,
- Learning rate, momentum
- Early stopping, weight decay, minibatch size
- ....

Y. Bengio. Practical Recommendations for Gradient-Based Training of Deep Architectures, 2012
Learning to Learn (Meta learning)

- Current:
  ML expertise + data + computation

- Future:
  data + 100X computation

Credits: https://becominghuman.ai/jeff-deans-talk-on-large-scale-deep-learning-171fb8c8ac57
Use of Prior Knowledge

- Error curve is flatening → add prior knowledge!
  - Visual common-sense for scene understanding using perception, semantic parsing and reasoning, S Aditya, Y Yang, C Baral, C Fermuller, Y Aloimonos - 2015.
Open Problems
Is Winter Coming Again?!

Credits: Monty Barlow
Unsupervised Learning

Current models are hungry for labeled data. Today’s DL is supervised learning.

“The Revolution Will Not be Supervised.” – Yann Lecun
Deep RL Takes Too Long to Train

RL systems require a gazillion trials!
Neural Nets need a Vapnik!

The theory about generalization properties of ANNs is not completely understood.
Neural Nets need a Vapnik!
Natural Language Understanding

- Headlines:
  - Enraged Cow Injures Farmer With Ax
  - Hospitals Are Sued by 7 Foot Doctors
  - Ban on Nude Dancing on Governor’s Desk
  - Iraqi Head Seeks Arms
  - Local HS Dropouts Cut in Half
  - Juvenile Court to Try Shooting Defendant
  - Stolen Painting Found by Tree
  - Kids Make Nutritious Snacks

- Why are these funny?

Source: CS188
"If a mother has a son, then the son is younger than the mother and remains younger for his entire life."

"If President Trump is in Washington, then his left foot is also in Washington,"

“There’ll be a lot of people who argue against it, who say you can’t capture a thought like that. But there’s no reason why not. I think you can capture a thought by a vector.” – Geoff Hinton
PPCIC – CEFET/RJ
Programa de Pós-Graduação em Ciência da Computação
http://eic.cefet-rj.br/ppcic

THANKS!

Eduardo Bezerra (ebezerra@cefet-rj.br)
Backup slides
Transfer Learning

- Higher slope
- Higher asymptote
- Higher start

---

With transfer
Without transfer
ML Tribes

- There are several ML tribes!
  - Symbolists (rule-based systems)
  - Evolutionists (evolutionary computation, GAs)
  - Analogists (SVMs, k-NN, …)
  - Bayesians (Bayes update rule)
  - Connectionists (ANNs)
[...] hard problems are easy and the easy problems are hard. The mental abilities of a four-year-old – recognizing a face, lifting a pencil, walking across a room, answering a question – in fact solve some of the hardest engineering problems [...], it will be the stock analysts and petrochemical engineers and parole board members who are in danger of being replaced by machines. The gardeners, receptionists, and cooks are secure in their jobs for decades to come.
Artificial Neuron

A model inspired in the real one (biological neuron).

\[ z = \mathbf{w}^T \mathbf{x} + b \quad a = \varphi(z) \quad a = \hat{y} \]
Artificial Neuron - input

\[ x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \]

\[ z = w^T x + b \]

\[ a = \varphi(z) \]

\[ a = \hat{y} \]
Artificial Neuron – parameters

\[ w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \]

\[ z = w^T x + b \quad a = \varphi(z) \]

\[ a = \hat{y} \]
Artificial Neuron – pre-activation

\[ a = \varphi(z) \]

\[ z = w^T x + b \]
Artificial Neuron – activation function

The composition of linear transformations is also a linear transformation.

Nonlinearities are necessary so that the network can learn complex representations of the data.

"The composition of linear transformations is also a linear transformation"
Artificial Neural Net

Feedforward Neural Network
AI Winters
Artificial Neuron

McCulloch-Pitts model (boolean functions, cycles)

1943
NY Times: “the embryo of an electronic computer that [the Navy] expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence.
Perceptrons are Deprecated!

“[…] by the mid-1960s there had been a great many experiments with perceptrons, but no one had been able to explain why they were able to recognize certain kinds of patterns and not others.”

1969

siblings(X, Y) :- parent(Z, X) and parent(Z, Y), 😈
The First AI Winter

1973
The Lighthill report

1974
DARPA cuts AI funding
"it is comparatively easy to make computers exhibit [...] intelligence tests or playing checkers, and difficult or impossible to give them the skills of a one-year-old when it comes to perception and mobility."
The Second AI Winter

1990s