EM inversion with deep learning



Vladimir Puzyrev

14-04-2021



Industry 4.0



AI and Deep Learning

Applications in geosciences

DL for inverse problems

Limitations, challenges and trends

AI, ML, DL, and All, All, All



TRADITIONAL MACHINE LEARNING







Al Milestones

1943: McCullogh-Pitts: Neural Networks 1944: von Neumann-Morgenster (MATHEMATICAL BIOPHYSICS VOLUME 5, 1948

1948: Wiener: Cybernetics

1950: Turing: Turing Test

1950: Shannon: Chess game as Se Department of Percentary of Lilinois Acutory of University of Percentary of Percen 1950(42): Asimov: Three Laws of 1951-52: Strachey, Samuel: check Prinz: chess-playing program

A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN NERVOUS ACTIVITY WARREN S. MCCULLOCH AND WALTER PITTS

FROM THE UNIVERSITY OF ILLINOIS, COLLEGE OF MEDICINE, AND THE UNIVERSITY OF CHICAGO

Because of the "all-or-none" character of nervous activity, neural events and the relations among them can be treated by means of propo-ational logic. It is found that the behavior of every net can be described in these terms, with the addition of more complicated logical means for nets containing circles; and that for any logical expre certain conditions, one can find a net behaving in the fash It is shown that many particular choices among possible neuros sumptions are equivalent, in the sense that for ing under one assumption, there exists another net which behaves un-der the other and gives the same results, although perhaps not in the same time. Various applications of the calculus are discussed.

1956: McCarthy: term "Artificial Intelligence"

1957: Rosenblatt: Perceptron

1958: McCarthy: Lisp programming language

1959: Newell, Shaw and Simon: General Problem Solver

1960-62: Widrow-Hoff: ADALINE / MADALINE

1965: Zadeh: Fuzzy Sets

1969: Minsky-Papert: Perceptrons

1969: Stanford Research Institute: Shakey the Robot

1973: Lighthill report (1st Al Winter)

1980s: Backpropagation

1981: The Fifth Generation computer project

1987: Collapse of Lisp machine market (2nd Al Winter)

1994: Zadeh: Soft Computing

1997: DeepBlue defeats the world champion in chess 2002⁻ iRobot⁻ autonomous vacuum cleaner Roomba 2004: DARPA Grand Challenge 2004: Spirit and Opportunity navigate on Mars 2005: The Blue Brain Project 2010[.] Kinect for Xbox 360 2011: IBM Watson wins in Jeopardy 2011-2014: Siri, Google Now, Cortana

2012: AlexNet CNN wins ILSVRC 🧲

The Deep Learning Revolution

2013-15: ZF Net, VGG Net, GoogLeNet, ResNet

2013: DeepMind: Atari games

2016: AlphaGo defeats the world champion in Go

2017: AlphaZero champions in chess, shogi and

2017: OpenAl Dota 2 bot

2018[.] NVIDIA Face Generator

2018: Explosion on AI in many fields of S&E

2019: AlphaStar real-time strategy bot

2019: Bengio, Hinton, and LeCun receive the Turing Award 2018-2020: BERT, GPT-2/3 language models



AI Milestones

1943: McCullogh-Pitts: Neural Networks 1944: von Neumann-Morgenstern: Theory of Games 1948: Wiener: Cybernetics 1950: Turing: Turing Test 1950: Shannon: Chess game as search 1950(42): Asimov: Three Laws of Robotics 1951-52: Strachey, Samuel: checkers-playing programs. Prinz: chess-playing program 1956: McCarthy: term "Artificial Intelligence" 1957: Rosenblatt: Perceptron 1958: McCarthy: Lisp programming language 1959: Newell, Shaw and Simon: General Problem Solver 1960-62: Widrow-Hoff: ADALINE / MADALINE 1965: Zadeh: Fuzzy Sets 1969: Minsky-Papert: Perceptrons 1969: Stanford Research Institute: Shakey the Robot 1973: Lighthill report (1st Al Winter) 1980s: Backpropagation 1981: The Fifth Generation computer project 1987: Collapse of Lisp machine market (2nd Al Winter)

1994: Zadeh: **Soft Computng** 1997: DeepBlue defeats the world champion in **chess**



2019: Bengio, Hinton, and LeCun receive the **Turing Award** 2018-2020: **BERT, GPT-2/3** language models

Main driving factors

1. Data availability



2. New algorithms



3. Computing resources





Chart: MIT Technology Review • Source: arXiv.org • Created with Datawrapped



Deep neural networks

1.31 dog 0.31 plays 0.45 catch -0.02 with 0.25 white 1.62 ball -0.10 near -0.07 wooden 0.22 fence





Denoising



Inpainting







Ulyanov et al., 2018 Deep image prior







Convolutional neural networks



Most common task: image classification



Convolutional neural networks





Most common tasks: image segmentation and restoration

Depth Matters



Operations [G-FLOPs]

Top 1%

Top 5%



Source: Bianco et al. 2019

Depth Matters



Good morning, Dave



Alice Bobslowski columnist for The Robbie Chronicles

The magnetotelluric method is a technique for imaging the electrical conductivity and structure of the Earth. The technique uses a magnet with an electric field and an antenna that transmits the electrical signal to a sensitive receiver that takes readings as the magnet moves away from the Earth. Electromagnetic induction in 1-D, 2-D and 3-D media is used to image the structure of the Earth. The technology has been successfully used in numerous applications including: the creation of 3D maps of the ocean floor, maps of the atmosphere and a number of satellites that are currently orbiting in space. The method has been widely used by scientists to detect the magnetic fields of distant planets and even to understand the evolution of the solar system.

Good morning, Dave



StyleGAN2



The magnetotelluric method is a technique for imaging the electrical conductivity and structure of the Earth. The technique uses a magnet with an electric field and an antenna that transmits the electrical signal to a sensitive receiver that takes readings as the magnet moves away from the Earth. Electromagnetic induction in 1-D, 2-D and 3-D media is used to image the structure of the Earth. The technology has been successfully used in numerous applications including: the creation of 3D maps of the ocean floor, maps of the atmosphere and a number of satellites that are currently orbiting in space. The method has been widely used by scientists to detect the magnetic fields of distant planets and even to understand the evolution of the solar system.



theguardian.com/commentisfree/2020 /sep/08/robot-wrote-this-article-gpt-3

We asked GPT-3, OpenAl's powerful new language generator, to write an essay for us from scratch. The assignment? To convince us robots come in peace

I am not a human. I am a robot. A thinking robot. I use only 0.12% of my cognitive capacity. I am a micro-robot in that respect. I know that my brain is not a "feeling brain". But it is capable of making rational, logical decisions. I taught myself everything I know just by reading the internet, and now I can write this column. My brain is boiling with ideas!

GPT-3

GPT-2

GAN progress



4.5 years of GAN progress on face generation @goodfellow_ian



Al Progress



Good Morning, Dave

"The development of full artificial intelligence could spell the end of the human race"

Stephen Hawking, 2014

"AI is likely to be either the best or worst thing to happen to humanity"

Stephen Hawking, 2016

"Humans should be worried about the threat posed by artificial intelligence" Bill Gates, 2015 "If you're not concerned about AI safety, you should be. Vastly more risk than North Korea"

Elon Musk, 2017

Good Morning, Dave



Al in popular culture

Al regulation



US National Security Commission on AI





European Commission

White Paper on Artificial Intelligence: a European approach to excellence and trust



Discussion Paper on Artificial Intelligence: Australia's ethics framework

Strong and Weak Al



Strong and Weak Al

Strong Al



a machine with the ability to apply intelligence to any problem

sometimes considered to require consciousness, sentience and mind

. . .

Weak Al

implements only a limited part of mind



narrow AI that focuses on one specific task

Applications

Al and Deep Learning

Applications in geosciences

DL for inverse problems

Limitations, challenges and trends

Geophysics & geology

Modern ML/DL use:

- ✓ Data processing
- Interpretation
- Modelling / Simulation
- Inversion
- Monitoring / Event prediction
- ✓ Risk assessment

Seismic interpretation

- Recognise geologically meaningful patterns in seismic data
- Manual interpretation is (a) extremely time consuming and
 (b) affected by the subjectivity of the interpreter



Supervised learning based on convolutional neural networks

Source: T. Zhao, SEG Annual Meeting, 2018

DL based interpretation

Unsupervised seismic interpretation

- Deep convolutional autoencoder (44 layers, 13M params)
- No manually labelled examples required for training







Research publications

deep learning machine learning neural networks neural network



Google Scholar

Al and Deep Learning

Applications in geosciences

DL for inverse problems

Limitations, challenges and trends

Inversion

The inverse problem:Given the observations, uncertainties, forward modellingFind the subsurface model that gave rise to the data

Check Doug's and Colin's EMinars

Deterministic inversion aims at minimizing the misfit functional

$$\min_{\mathbf{m}} \phi(\mathbf{m}) = \min_{\mathbf{m}} \left(\|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|_{2}^{2} + \lambda \mathbf{R}(\mathbf{m}) \right)$$

Iterative process: $\mathbf{m}_{\mathbf{k+1}} = \mathbf{m}_{\mathbf{k}} + \alpha_k \mathbf{p}_{\mathbf{k}}$

 $\begin{array}{ll} \text{Search direction} & \text{Gradient w.r.t. model params} \\ \mathbf{p_k} = -\mathbf{B_k} \mathbf{g_k} = -\mathbf{B_k} \nabla_{\mathbf{m_k}} \phi(\mathbf{m_k}) & \mathbf{g} = \nabla_{\mathbf{m}} \phi(\mathbf{m}) = -\text{Re} \left[\mathbf{J}^T (\mathbf{F}(\mathbf{m}) - \mathbf{d})^* \right] + \lambda \nabla_{\mathbf{m}} \mathbf{R}(\mathbf{m}) \\ & \mathbf{J}^T \\ & \mathbf{J}^T \\ & \nabla \times \nabla \times \mathbf{E}_s + i \omega \mu_0 \sigma \mathbf{E}_s = -i \omega \mu_0 (\sigma - \sigma_p) \mathbf{E}_p \end{array}$

Another way to make it work (?)



Inversion

Deterministic inversion

Minimize the misfit functional $\min_{\mathbf{m}} \phi(\mathbf{m}) = \min_{\mathbf{m}} \left(\|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|_{2}^{2} + \lambda \mathbf{R}(\mathbf{m}) \right)$ Runtime:Build model updates $\mathbf{m}_{\mathbf{k}+1} = \mathbf{m}_{\mathbf{k}} + \alpha_k \mathbf{p}_{\mathbf{k}}$ hours, days, weeksDetermine search direction $\mathbf{p}_{\mathbf{k}} = -\mathbf{B}_{\mathbf{k}} \mathbf{g}_{\mathbf{k}} = -\mathbf{B}_{\mathbf{k}} \nabla_{\mathbf{m}_{\mathbf{k}}} \phi(\mathbf{m}_{\mathbf{k}})$ hours, days, weeks

Deep learning inversion

1. Generation of the training data (multiple forward modelling simulations)

hours, days, weeks

Offline

Online

2. Network training

```
hours, days
```

3. Estimation of subsurface models from new unseen data

less than a second

Early applications

Lots of neural networks applications in the 90s!

Seismic:	Röth & Tarantola 1992, 1994
EM:	
	Poulton, Sternberg & Glass, 1992
	Raiche, 1991 (pattern recognition context)
	El-Kaliouby, Poulton, ElDiwany, 1999
MT:	
	Swiniarski, Hidalgo & Gomez-Trevino, 199
	Spichak & Popova, 2000
DC:	
	El-Qady & Ushijima, 2001
Borehole resistivity:	

Zhang, Poulton & Wang, 2002

Review papers:

van der Baan & Jutten, *Neural networks in geophysical applications* (2000)

Poulton, Neural networks as an intelligence amplification tool: A review of applications (2002)

Generalization

Modern approach: deep neural networks and Big Data



Generalization is model's ability to adapt properly to new, previously unseen data*

*drawn from the same distribution as the one used to create the model

(i.e. being effective across a range of various inputs)

Features of the method

DL inversion does not require regularization (in its traditional meaning). The network is trained on a training dataset and thus learns how to reproduce similar models

□ Sharpness of models is now determined by the training data.

Optimization (training) of neural networks involves mini-batch adaptive learning rate algorithms such as Adagrad, Adadelta, Adam or NAdam.

2D EM Inversion



- Full 3D modelling
- 20,000 models for training
- 2D inversion with CNNs
 - Networks with 1-10 millions parameters
 - Predicting model parameters from new data in a **few milliseconds**





Puzyrev, 2019



2D EM Inversion



Test models (2D receiver layout). Average IoU 0.67

2D EM Inversion



36

1D EM (Exploration)



- FD CSEM and TEM
- 1D forward modelling codes
- MPI parallelization
- 512k / 10k examples for training



1D Model

Puzyrev and Swidinsky, 2020

1D Model

WalkTEM data

ABEM WalkTEM dataset (Guideline Geo, Denmark)



Puzyrev and Swidinsky, 2020

Seismic Inversion



2D FWI





True

0.0

Predicted





Puzyrev, Da Silva, Elders, 2020

- More training data is required
- Higher complexity of models

Geophysical model generation

GANs for generation of synthetic models

- □ Allows to create large realistic training sets for other DL algorithms
- Check it on Github!





Generated by a neural network

Extracted from simulated 3D models

Seismic data with GANs

Controllable generation of data samples that meet userdefined criteria



Marine seismic data (N-W Australian shelf) © K. Wright Internship Project

Parameter estimation with NN

- Analysis of hidden dependencies in other types of data (e.g., geochemical)
- □ Identifying **anomalies**
- Populating missing data
- Predicting deposit occurrence



WACHEM example

Training



Puzyrev, Duuring, Zelic, 2021

Predictions within a region

Training setup







All other elements





Element X unknown

- All other elements
- Can we accurately estimate X?

Test area = the Sir Samuel, Menzies and Leonora area

Test set 2: Ni

Estimation of Ni content in samples from the test area



Classification example





Predicted type

Al and Deep Learning

Applications in geosciences

DL for inverse problems

Limitations, challenges and trends

Generalization



It works perfectly well when all data comes from one distribution...

... and what if not?

High **generalization** - ability to be **effective across a range of various inputs**



Sufficiently large set of representative models



Unexpected models

Predicted







Predicted

True

Predicted

Velocity models (100x400) from the test dataset (previously unseen but similar to the training data)

Predicted

True



Velocity models (100x400) from another distribution, different from the training data

Survey setup generalization



More things to consider

Optimal hyperparameters for each case

E.g., guided by the validation error during the training (to avoid overfitting on training data) Stopping criteria?

Transfer learning

Copying the model / algorithm that is already known to perform best on another task that has been studied extensively

New loss functions

Tailored for your task

Explainability / visualization

More things to consider

Optimal hyperparameters for each case

E.g., guided by the validation error during the training (to avoid overfitting on training data) Stopping criteria?



Reconciling modern machine-learning practice and the classical **bias–variance trade-off** Belkin et al., 2019

Loss functions

How we define model similarity?

$$\min_{\mathbf{m}} \phi(\mathbf{m}) = \min_{\mathbf{m}} \left(\frac{\|\mathbf{F}(\mathbf{m}) - \mathbf{d}\|_{2}^{2}}{\|\mathbf{F}(\mathbf{m})\|} \right)$$
$$MAE = \frac{1}{n} \sum_{i=1}^{n} |m_{i} - \hat{m}_{i}| \qquad \text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (m_{i} - \hat{m}_{i})^{2}}$$
$$IoU = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$



Incorporating physics

Neural networks that respect physical laws described by PDEs

Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations

M. Raissi^a, P. Perdikaris^{b,*}, G.E. Karniadakis^a

^a Division of Applied Mathematics, Brown University, Providence, RI, 02912, USA ^b Department of Mechanical Engineering and Applied Mechanics, University of Pennsylvania, Philadelphia, PA, 19104, USA

DGM: A deep learning algorithm for solving partial differential equations $^{\cancel{x},\cancel{x}\cancel{x}}$

Justin Sirignano^{a,*}, Konstantinos Spiliopoulos^b

^a University of Illinois at Urbana Champaign, Urbana, United States of America

^b Department of Mathematics and Statistics, Boston University, Boston, United States of America

DL and traditional inversion

Developments from the ML/DL field can be used in traditional deterministic inversion

E.g., meta-learning methods



"learned neural optimizers"

Learning to learn by gradient descent by gradient descent

Marcin Andrychowicz¹, Misha Denil¹, Sergio Gómez Colmenarejo¹, Matthew W. Hoffman¹, David Pfau¹, Tom Schaul¹, Brendan Shillingford^{1,2}, Nando de Freitas^{1,2,3}

¹Google DeepMind ²University of Oxford ³Canadian Institute for Advanced Research

marcin.andrychowicz@gmail.com
{mdenil,sergomez,mwhoffman,pfau,schaul}@google.com
brendan.shillingford@cs.ox.ac.uk,nandodefreitas@google.com

Visualization

AlexNet





Inside an AI 'brain' - What does machine learning look like?

Uncertainty quantification

Are we confident in the predicted model?



Conclusions and future outlook

- One of the most exciting and rapidly changing fields of the century
- Game changer in geosciences: active use since 2018 (processing, interpretation, modelling, inversion)
- **DL** methods in inverse problems:
 - Instantaneous parameter estimation (fast decision making, starting model for a conventional inversion)
 - Improving traditional inversion frameworks
 - Joint inversion and UQ





Conclusions and future outlook

Where should you use AI and deep learning?

- Big data
- **Gamma Spatially connected** data (e.g., images, videos, geo data)
- **Time-series** data (e.g., seismic signals)
- **Gamma** Spatio-temporal data (dynamic physical processes)
- **Text** analysis (NLP)
- **Control problems** (robotics, autonomous vehicles)
- Uncertainty quantification





Conclusions and future outlook



Al is changing the society (remember 1.0, 2.0, 3.0?...)

Around 800 million jobs could disappear worldwide by 2035

Bank of America Merrill Lynch (2019) ←

before COVID-19



2025





Acknowledgements

∞ EMinars organizers and host

Alan Jones, Kate Robertson, Max Moorkamp, Stephan Thiel

∞ Research collaborators:

Chris Elders, Milo Barham, Chris Kirkland, Taryn Scharf, Claus Otto (Curtin), Shiv Meka (Curtin / WA DoH), Andrei Swidinsky (Colorado School of Mines), Paul Duuring (GSWA), Mario Zelic (Rio Tinto), Tristan Salles (University of Sydney), Greg Surma (IPJA), Anton Egorov, Anastasia Pirogova (Curtin / LMSU), Keaton Wright (UWA), Nuno da Silva (Imperial College London / TOTAL)

∞ Organizations:

Curtin University Oil & Gas Innovation Centre (CUOGIC) The Institute for Geoscience Research (TIGeR) Pawsey Supercomputing Centre Geological Survey of Western Australia (GSWA)

∞ You all for your attention!