CERN openlab

Summary

Artificial Intelligence for Science, Industry and Society
Mexico City – UNAM, October 21-25, 2019

F.Carminati

Thanks!

- Guy Paic
- The UNAM colleagues
- The UNAM staff
- My co-chairs
- The IOC
- The speakers
- You!

Globally…

A very well-organized meeting

- The venue
- The city
- The transports
- The social event
- The banquet
  (I have been told…)

The scientific programme

- 49 high-level contributions
- Good participation from the public
- Wide spectrum of contributions and fields

Applications in Industry

AI Innovation in the Pharmaceutical Sector - Accelerating Research

AI in Pharma - Why Now?

emerj.com

Business Motives

Finding New Drugs or Treatments: Could be with new data, but often simply involves making the most of older data. (Discovery)

Determine Drugs Likely to Pass Trials: Find common factors among drugs or compounds in order to determine which are most promising and worthy of investment. (Prioritization)

OECD/CERN 2019, Mexico City Emerj

Display a menu

Factors Encouraging Adoption

A Push for Personalized Medicine: This trend requires matching genetic data with drugs - in other words: Finding new patterns across more data.

Data Overload and Pattern Recognition: Finding patterns in terabytes of data is a natural fit for ML.

Ability to Decide and Adopt: Unlike a hospital setting , patients, and executives all need to be harma firms have relatively clear e. who pays, who benefits).

Mexico City Emerj

AI Applications In Oil Extraction, Review and Case of Study: Convolutional Neural Networks Applied to Seismic Image Analisis

Exploration

Seismic Data

Oil Corrosion/Risk Analysis

Materials Management

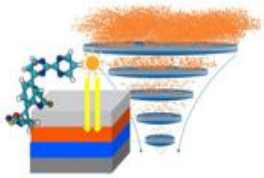
Data Analysis

ININ

Towards inverse design in chemistry: from prediction to deep generative models

How can we find molecules according to functionality?

High throughput virtual screening (HTVS)



Quantum chemistry, Gaussian Process prediction and molecular structure interpretation.

Explicit optimization

Variational Autoencoders, exploring and optimizing in latent space

Implicit optimization

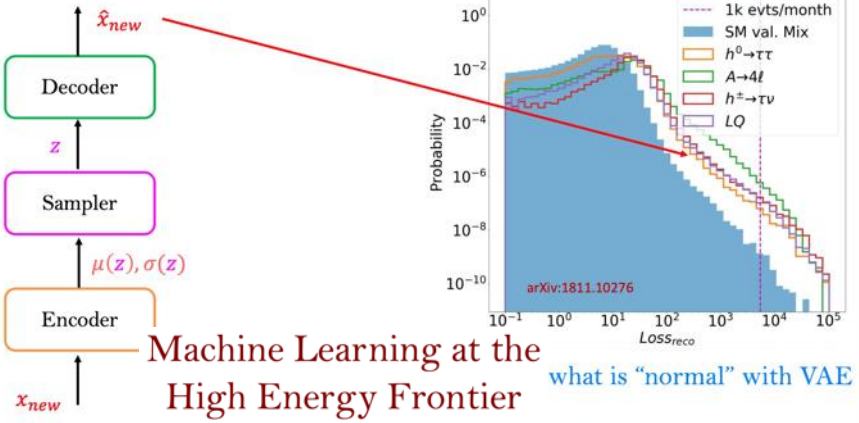
Reinforcement learning and generative adversarial networks

Applications in Fundamental Science

Anomaly Detection with VAEs

M.Kagan

- What if we don't know what a new signal looks like?
  - Goal: Find rare events that differ from "standard"



Machine Learning at the High Energy Frontier

what is "normal" with VAE

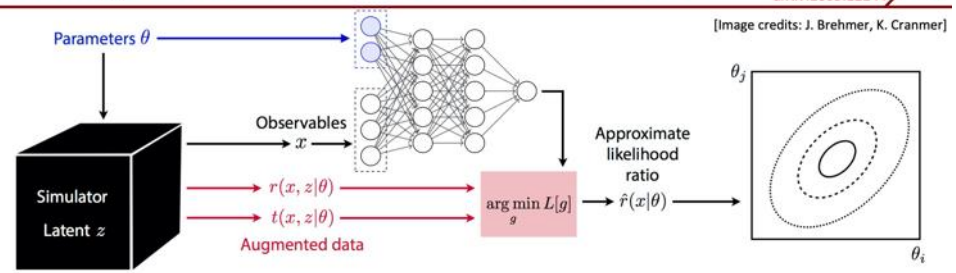
$$\mathcal{L}_{reco}(x_{new};\theta,\phi) = \mathbb{E}_{q_\phi(z|x_{new})}[\log p_\theta(x_{new}|z)]$$

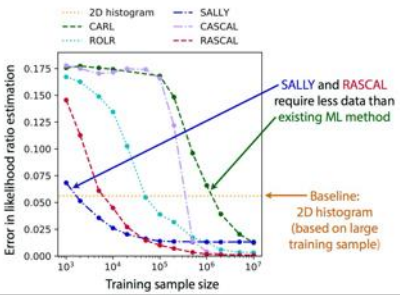
Look for anomalies as poorly reconstructed events

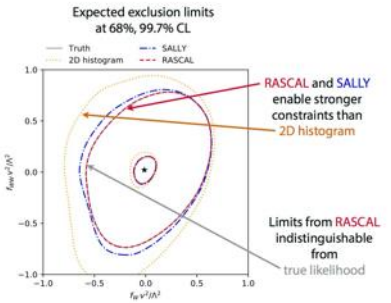
Mining Gold for Likelihood-Free Inference

arXiv:1805.12244

[Image credits: J. Brehmer, K. Cranmer]





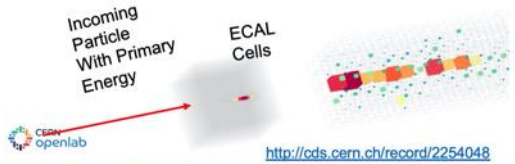


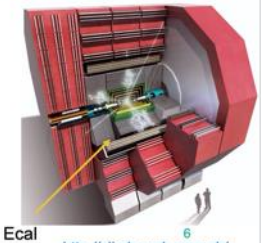
Data set

Compact Linear Collider CLIC

F.Carminati

- **Proposed** linear particle accelerator
- **Calorimeter data set** developed for ML applications
- Events as selected cells around the barycenter of particle showers simulated using Geant4
- Primary particle energy 10-500 GeV (electrons)
  - Event → 25 x 25 x 25 image → 15, 625 cells
    - 200,000 events
- Detector response as 3D images
  - Highly segmented (pixelized)
    - critical for particle identification and energy determination
  - Highly sparse
    - only ~20% cells with energy deposition
  - Large dynamic range
    - > seven orders of magnitude





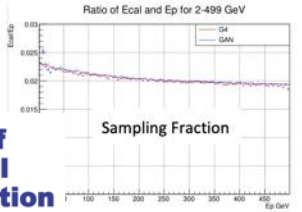
http://cds.cern.ch/record/2254048

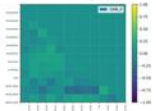
http://clicdp.web.cern.ch/

Distributed Training of Generative Adversarial Networks for Fast Simulation

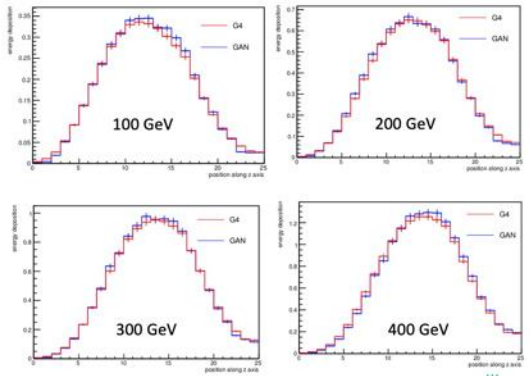
Transfer Learning

- Training for 2-500 GeV spectrum
  - Starting from pretrained weights (trained for 100-200 GeV)

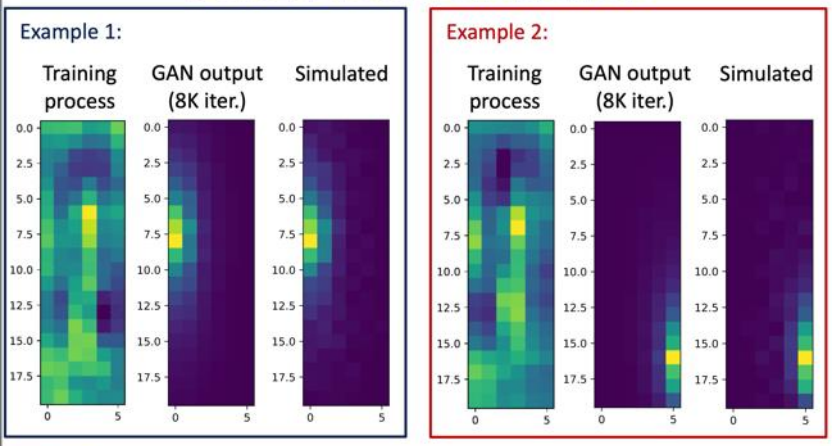




Shower shapes in the longitudinal direction for Different Primary energies



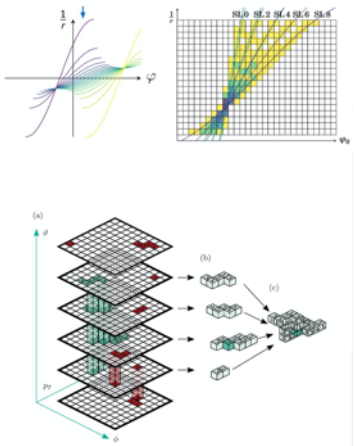
Other Machine Learning Applications: DUNE Photon Simulation



Online Estimation of Particle Track Parameters based on Neural Networks for the Belle II Trigger System

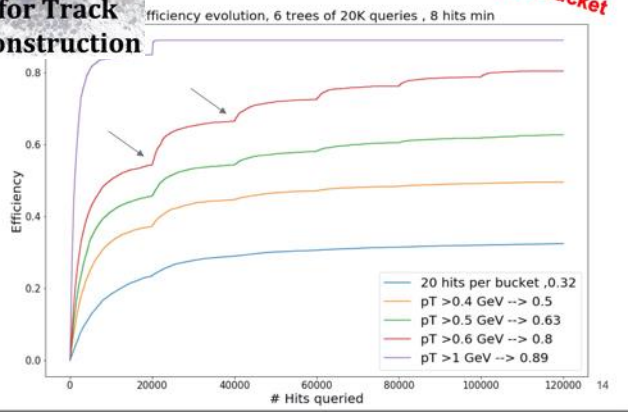
Outlook – Upgrade

- Operation using networks trained with experiment data
  - Improving resolution further to achieve better trigger system
- Advanced Preprocessing -> 3D-Hough Tracker with Bayesian Weights
  - Finds 99% of tracks within the detector
  - 2 cm theoretical resolution
- Future Upgrade using new FPGAs
  - Prototype implemented currently
  - Multi-Track NNT on same FPGA



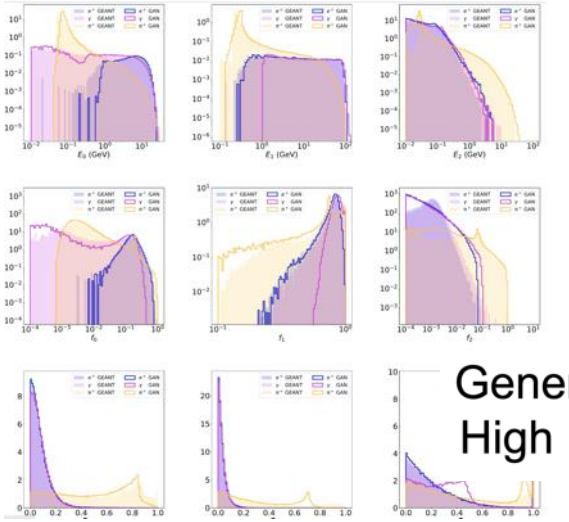
Hashing & Metric Learning for Track Reconstruction

20 hits per bucket



Rebuilding the data structure, after removal of found particles,improves the efficiency

CaloGAN validation



Physics distributions: shower shapes

=

1D marginals of data distribution

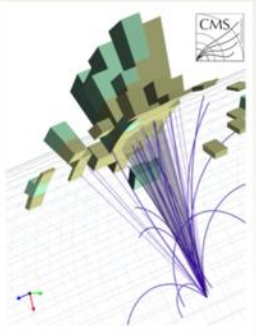
Challenge to reach needed accuracy

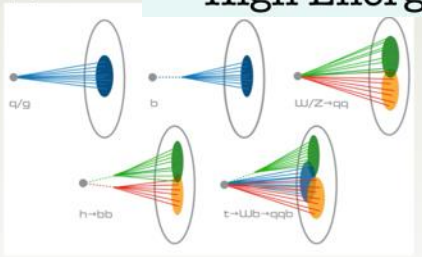
Generative models in High Energy Physics

[1705.02355]

Advances in Classification and Regression in High Energy Physics

- Jets are a good 'playground' for DNN solutions
  - Jet classification crucial for a plethora of physics
  - Complex
  - Hard to find simple analytic solutions
  - Involve many detector components, so 'messy' objects





- Jets consist of a varying number of particles, each with different properties
- The correlation between the particles can be crucial to classify the jet

Operating the LHC

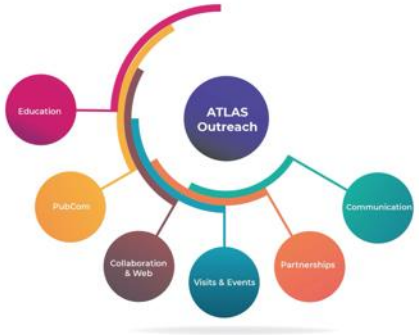
Machine learning in accelerator physics:
Applications at the CERN Large Hadron Collider

Large-Scale Scientific endeavours
the production and dissemination of advance computer science's knowledge

8

ATLAS Outreach, CEVALE2VE.org, ICTP-PWF program work on making HEP and Computer Science accessible.

Today's presentation is about the **efforts in the use of Open Access & Source for data samples, analysis frameworks and Software** related tools developed **for HEP training and education in standard university and student computers in the region**

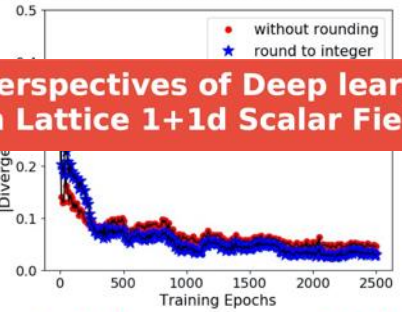


GAN – generate proper configurations

The divergence condition get learned automatically:

'Physical' configs can be generated

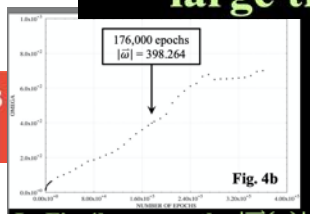
$$\nabla \cdot k(n) = \sum_{\nu} [k_\nu(n) - k_\nu(n-\hat{\nu})] = 0$$

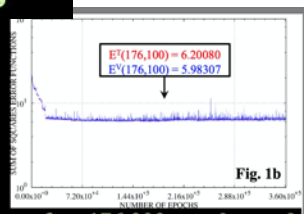


Automatically capture the implicit physical constraint!

Perspectives of Deep learning techniques in Lattice 1+1d Scalar Field Theory

Regularization methods vs large training sets





In Fig 4b, we see the $|\bar{\omega}(\rho)|$ value keeps on growing after 176,000 epochs, and the training and validation error curves remain close together, Fig. 1b, indicating the absence of data overfitting and overtraining.

Noise complexity grows

Ideal BCs complexity remains fixed

Dataset 45,100 BC

Ideal BCs complexity << noise complexity

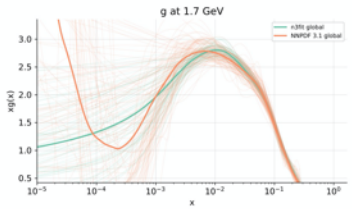
Dataset 451,000 BC

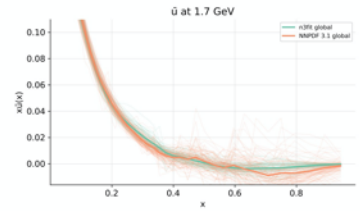
With 451,000 BCs, the ANN complexity will be used to extract the true features present in the BCs rather than noise.

36

Result comparison: replica-by-replica

Comparison, with the same selection of





- ✓ region
- ✓ More replicas satisfy post-fit requirements
- ✓ Compatibility of old and new results

Which translates to

- ✓ Many more studies can be performed at the same cost → more science
- ✓ Leading to a more accurate PDF determination

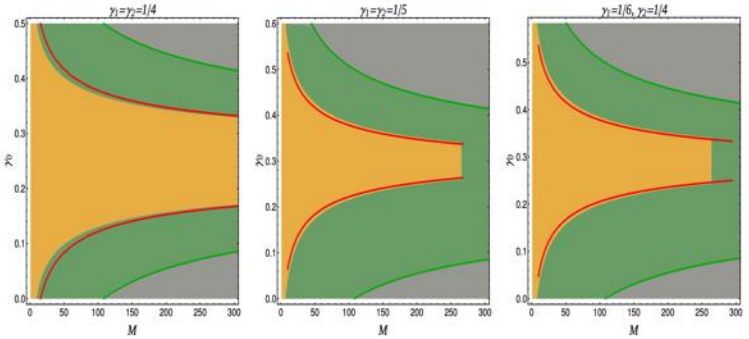
Studying the parton content of the proton with deep learning models

Randomness Characterization through Bayesian Model Selection

Comparing with Borel's Normality bounds

A sequence $\hat{s}$ is Borel-Normal if:

$$\left|\gamma_j^{(\beta)} - \frac{1}{2^\beta}\right| < \sqrt{\frac{\log_2 M}{M}}, \qquad \beta \leq \log_2 \log_2 M$$



Region where $\mathcal{M}_{\rm sym}$ is the likeliest , region allowed by Borel Normality test, and Approximated bounds obtained by an expansion of $\log \mathcal{M}_{\rm sym}/\mathcal{M}_{\alpha_2}$.

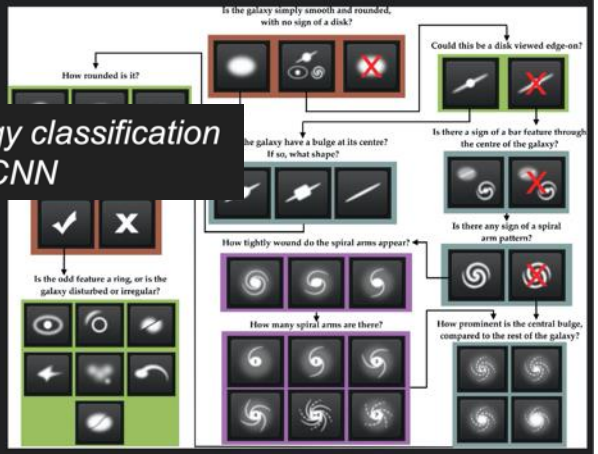
Galaxy Morphology classification using CNN

SDSS classified galaxies:

~ 300 000

carried out by thousands of "normal" people, and then a probability is given based on the number of votes.

https://zoo4.galaxyzoo.org

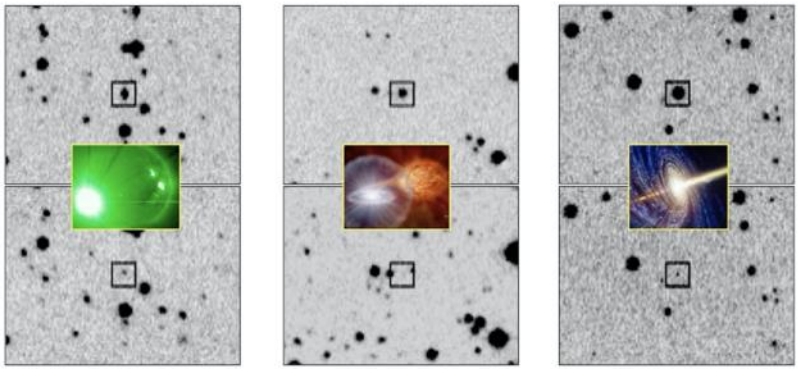


Willett et al. 2013

J. Antonio Vazquez-Mata ● jvazquez@astro.unam.mx

Astronomy in the Era of Big Data and Machine-Assisted Discovery

Automated Classification of Transients

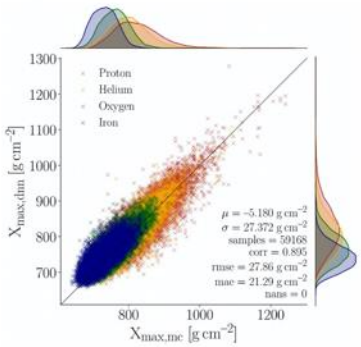


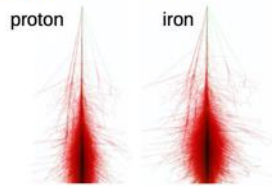
Vastly different physical phenomena, yet they look the same!
Which ones are the most interesting and worthy of follow-up?

Rapid, automated transient classification is a critical need!

Deep Learning for Cosmic-Ray Observatories

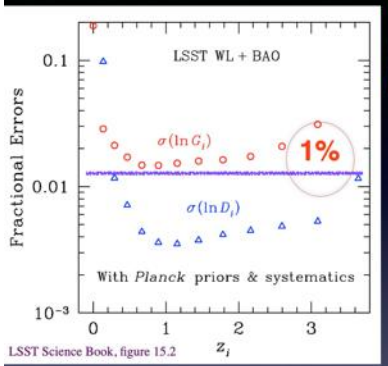
Reconstruction of Shower Maximum





- ✔Successful shower maximum reconstruction
  - Shows expected separation of elements
  - Resolution < 30 g/cm²
  - Absolute bias of ~ 5 g/cm²
  - Significant improvement to previous methods

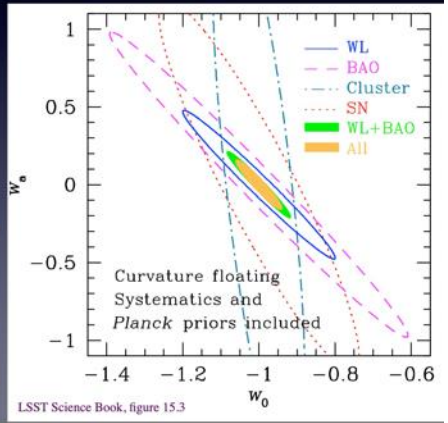
Cosmology with LSST: high precision measurements



LSST Science Book, figure 15.2

"Was Einstein right?"

- Multiple probes is the key!



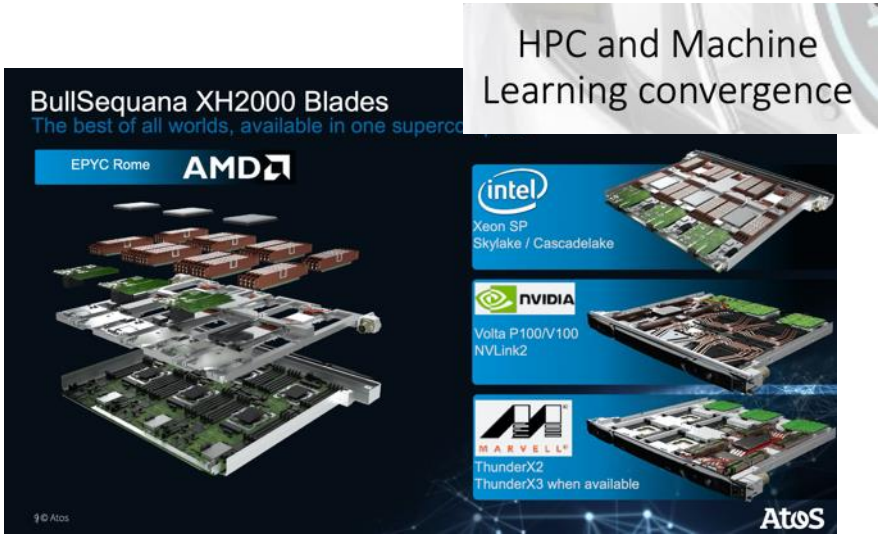
LSST Science Book, figure 15.3

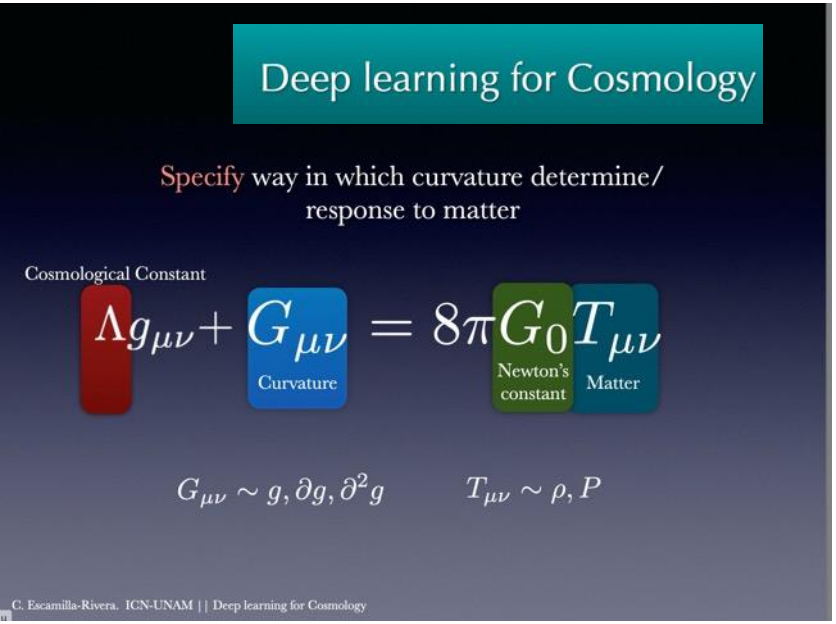
Upcoming Big Data and AI Challenges in Astronomy

By simultaneously measuring growth and curvature, LSST will tell us whether the recent acceleration is due to dark energy or modified gravity.

ML/DL and HPC


The Amazon ML Stack
ML for everyone: different users require different tools
AI SERVICES
ML SERVICES
AMAZON SAGEMAKER
ML FRAMEWORKS & INFRASTRUCTURE
© 2019, Amazon Web Services, Inc. or its Affiliates. All rights reserved. Amazon Trademark
aws


HPC and Machine Learning convergence
BullSequana XH2000 Blades
EPYC Rome
AMD
intel
Xeon SP
Skylake / Cascadelake
NVIDIA
Volta P100/V100
NVLink2
MARVELL
ThunderX2
ThunderX3 when available
Atos

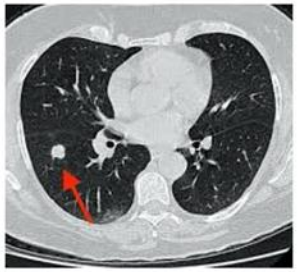

Deep learning for Cosmology
Specify way in which curvature determine/ response to matter
Cosmological Constant
$\Lambda g_{\mu\nu} + G_{\mu\nu} = 8\pi G_0 T_{\mu\nu}$
Curvature
Newton's constant
Matter
$G_{\mu\nu} \sim g, \partial g, \partial^2 g$
$T_{\mu\nu} \sim \rho, P$
C. Escamilla-Rivera. ICN-UNAM | | Deep learning for Cosmology

Health & Society

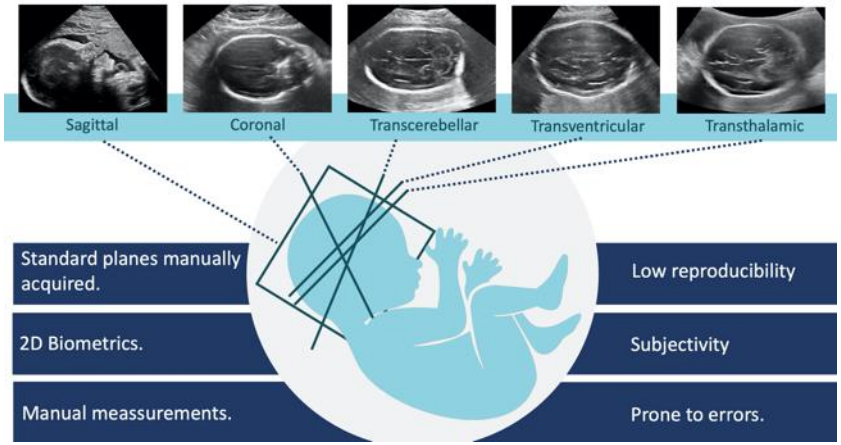
A machine learning approach for the feature extraction of pulmonary nodules

Cecilia Loeza, Rajesh Roshan

- Classification of pulmonary nodules in benign and malignant with different machine learning techniques using chest CT scans.



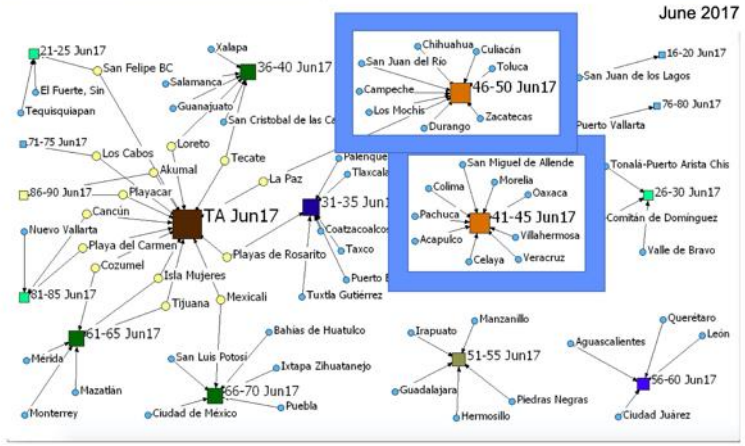
Variational Autoencoders for Medical Imaging



Network science to analyse insecurity perturbations that undermine Mexican tourism industry

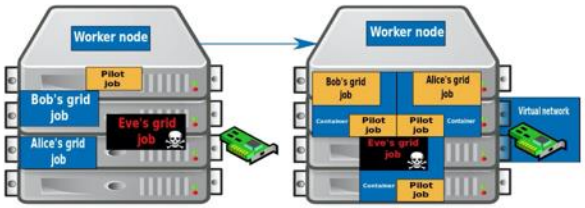
Findings

June 2017

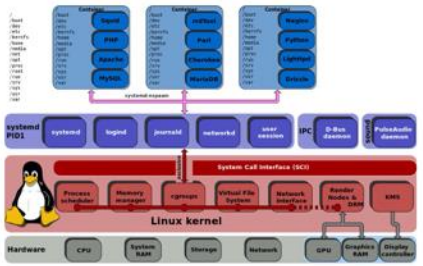


Proposed solution: 1. Grid Job execution and network isolation

Grid jobs are executed inside containers for isolation among the underlying system and other jobs.



Linux containers are a lightweight alternative to virtual machines. Processes are executed over the same kernel.



Arhuaco: Deep Learning and Isolation Based Intrusion Detection in High Energy Physics

The Ethical ”Vexata Quaestio” & methodology

Ethics & AI

Ethical Values

Trustworthy AI
The AI4EU approach

- Autonomy
- Beneficence
- Non-maleficence
- Justice
- Fidelity

- Think for a moment, *how* might these principles relate to research?

ia@cs.upc.edu

Regulating Emerging Technologies: Opportunities and Challenges for Latin America

FACT:

Machine Learning classification algorithms are created to discriminate*.

* *Discriminate*
From latin *discriminare*. tr. To select by excluding.

Practical Machine Ethics

Yandex SCHOOL OF DATA ANALYSIS

From online education to gig science and augmented intelligence

Project big lifecycle (co-researcl

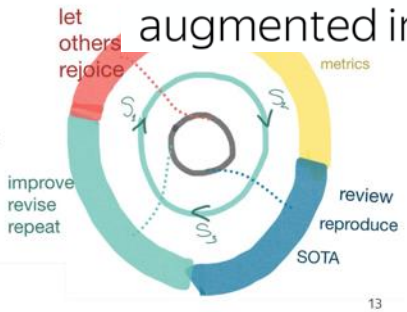
Inception stage

Learning: doing papers review, study of «state-of-the-art» methods

individual solution: iterative search, model understanding and improvement

Writing final report, slides and paper

NB: iterations may be nested on every stage.

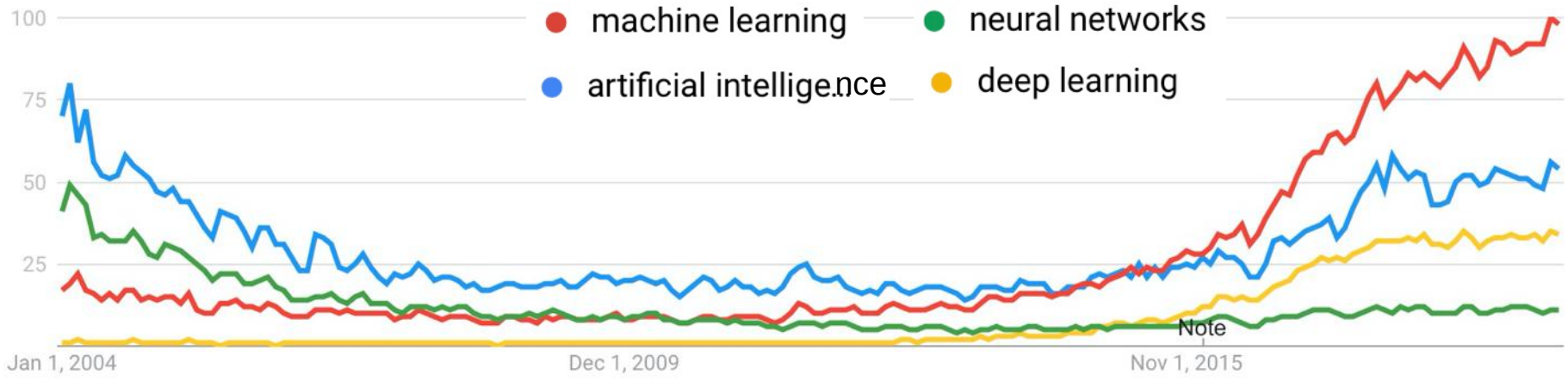


Andrey Ustyuzhanin

Take home messages

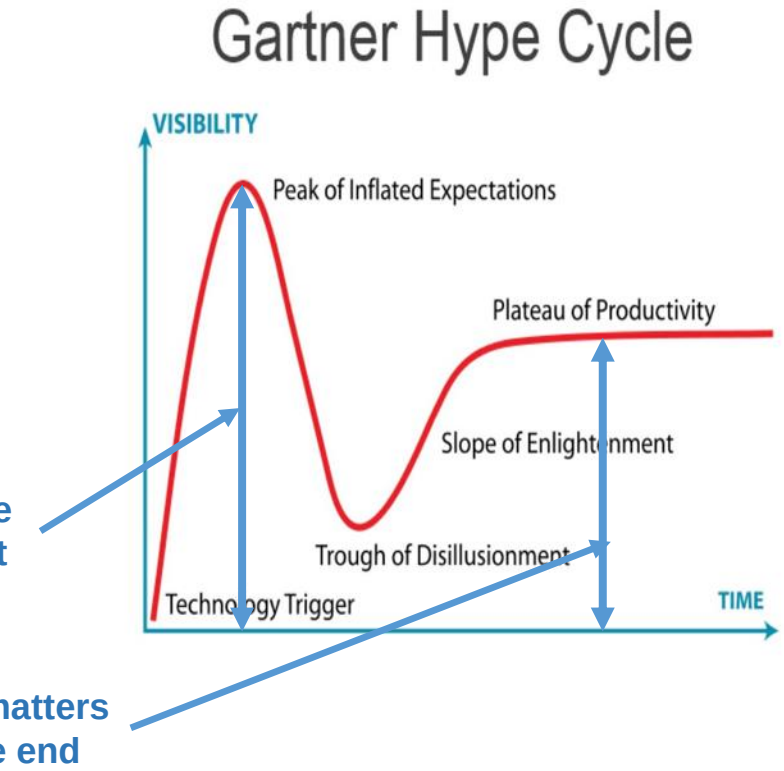
my own bias

The big picture



The Big Picture

- AI is NOT a new concept, in fact it is as old as the computers
  - c.f. The Touring challenge
- What makes it “resurgent” every so many years are the breakthrough in computing power
- However this time the step in available computing performance is really making the difference
- And the hype does help



AI is the ideal companion to Big Data

- I think we are just at the beginning of the exploitation of AI for Big Data
- The ability to use highly non-linear algorithms to find complex patterns is invaluable
- GAN of various types offer a model-less way of simulate
- In some sense DL is “fitting on steroids”, but tons of them…
- … and quantity has a quality of its own

AI for science

- The ability to “fuse” different sources of information and find patterns is also very important
- Astronomy is one of the foremost beneficiary, but biology, social sciences, earth sciences are following quickly
- I believe the word “revolution” may be well justified

AI & science

- AI will NOT replace science or scientific models
- Without a model, AI results may be meaningless and useless
- But AI can help building and validating models
- Model-less and model science can live side-by-side and cross-fertilize each others

AI for Industry

- Needless to say AI is finding a choice place in industrial applications of all sorts
- However in a large number of cases interpretability is necessary
- Forward AI provides correlation, not necessarily causation

AI providers

- We see a flurry of *program suites that do it all* by several vendors
- Standards are sorely missing and
  - On one side there is a lot of “packaging” of few basic algorithms
  - On the other, everybody seems to feel the need to write his/her DL code
- The market is rather confusing, but the smoke will settle at some point

AI & Society

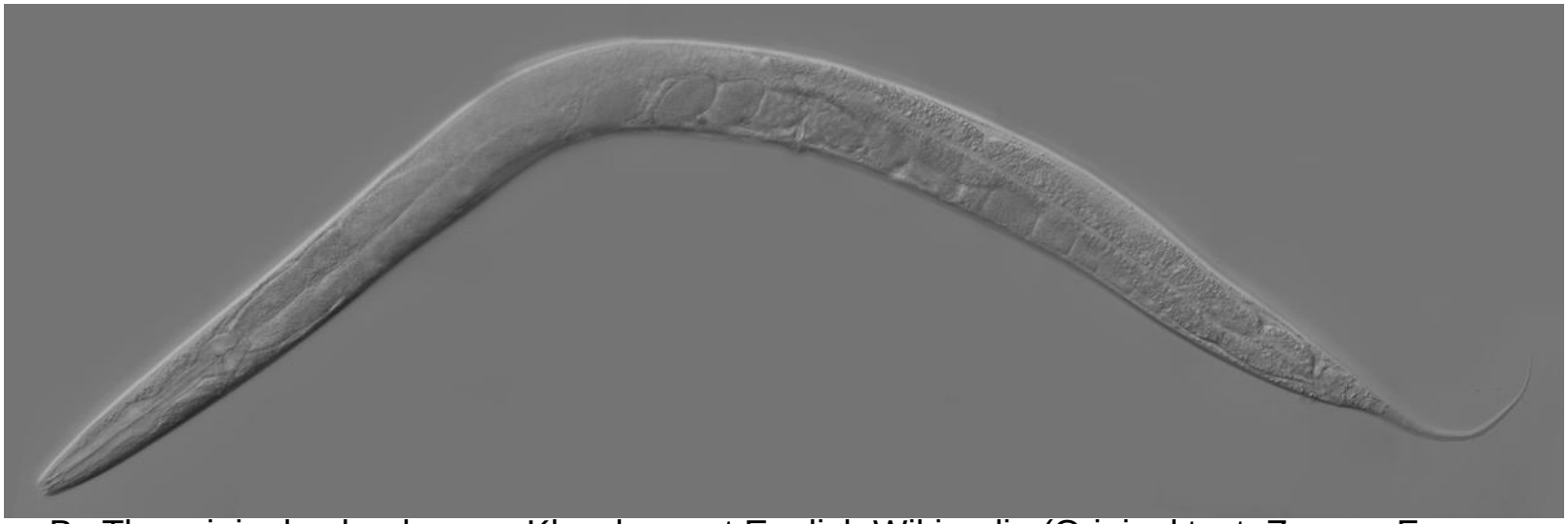
- AI may play a revolutionary role in dealing with societal issues such as health care, urban planning and optimization, environmental planning, transportation, prevention, education…
- However interpretability seems here to be an essential element in all cases where humans are involved

AI & Skynet

- We have seen two very different point of views
  - “Don’t you worry, it will never happen”
  - “If your iPhone is not where you put it last night, you are REALLY in trouble”
- The trouble here is that we do not know what consciousness is
- And Touring does not help here

AI & Skynet

- True any computer is much less complex than the human brain
- But not much less complex than a fruit fly, or, say, a Caenorhabditis elegans…



By The original uploader was Kbradnam at English Wikipedia.(Original text: Zeynep F. Altun, Editor of www.wormatlas.org) - Transferred from en.wikipedia to Commons.(Original text: Donated by Zeynep F. Altun),

AI & Ethics

- AI is opening a new discussion on ethics
- A ubiquitary and pervasive technology needs a solid ethical foundation
- … but it is just ≥6000 years that we are discussing about it
- AI may be a potent stimulant to further our understanding of ethics
- But I think it will still be an ethical nightmare

AI & Law

- Here again we swing between extremes
  - There is no more law in the robot than in the horse
  - A whole new legal framework is needed
- I would err on the side of the first statement
- However, again, AI may prompt us to give a fresh look to old concept

AI & Quantum

- The merging of two such fast developing might seem ambitious
- However the large parameter space of Quantum Computing and its ability to perform Gibbs sampling on this parameter space are a match made in heaven
- A QSVM algorithms for Higgs search has been presented in this conference (IBM talk)

Conclusions

- AI is probably here to stay this time
- Industry and science are, for the moment, making the most of it
- Probably the most appealing opportunities are in the social / environmental realm (health care being high on the list)
- But this is also where the legal and ethical issues are more relevant
- The “market” is confusing, but it always is with new technologies

Post conclusions

Thanks again for the beautifully organized workshop!

And Now This

The gender of the robot

- A Spanish teacher was explaining to her class that in Spanish, unlike English, nouns are designated as either masculine or feminine.
- House for instance is feminine, *la casa;* pencil however is masculine *el lapiz*
- The teacher split her class into two groups and asked them to decide whether a computer was male or female

The gender of the robot

- The men's group decided that computer was definitely female, *la computadora* because
  1. No one but their creator understands their logic
  2. The native language they use to communicate with other computers is incomprehensible to everyone else
  3. Even the smallest mistakes are stored in long memory for possible retrieval
  4. As soon as you make a commitment to one, you will find yourself spending half your paycheck on accessories for it

The gender of the robot

- The women's group, however concluded that computers should be masculine *el computador* because.
  1. In order to do anything with them, you have to turn them on
  2. They have a lot of data but still can't think for themselves
  3. They are supposed to help you solve problems, but half the time they ARE the the problem
  4. As soon as you commit to one , you realise that if you had waited a little longer, you could have obtained a better model .