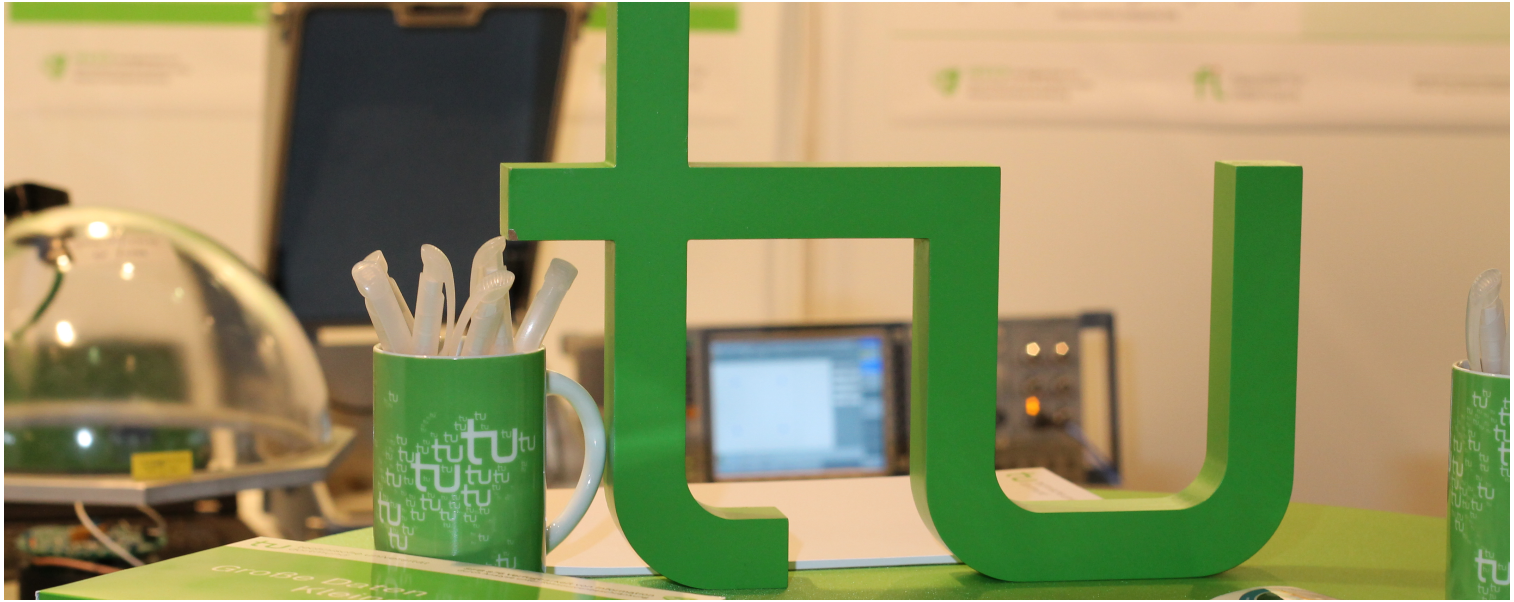
technische universität dortmund

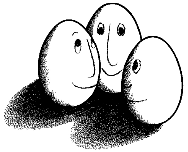
ML2R

SFB 876 Providing Information by Resource-Constrained Data Analysis



Machine Learning and Physics

Katharina Morik, Artificial Intelligence , TU Dortmund



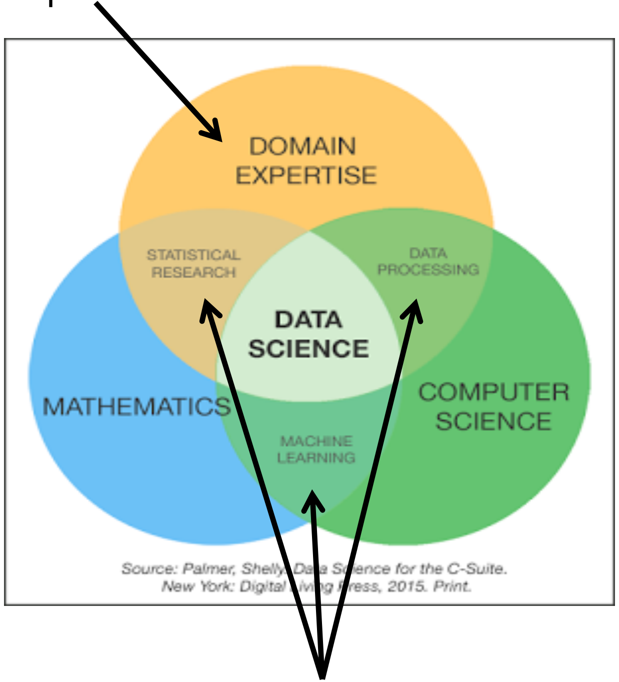
Overview

- AI and ML
  - History of AI
  - Machine learning (ML)
  - Illustration of ML research
- Physics and ML
  - Ice Cube, FACT, CTA, SKA
  - Deep learning on FACT
  - SVM on FACT
  - Unfolding on FACT
  - FPGA for FACT

Crab nebular around crab pulsar
Mixed optical and X-Ray spectrum

The overall picture

Experiments



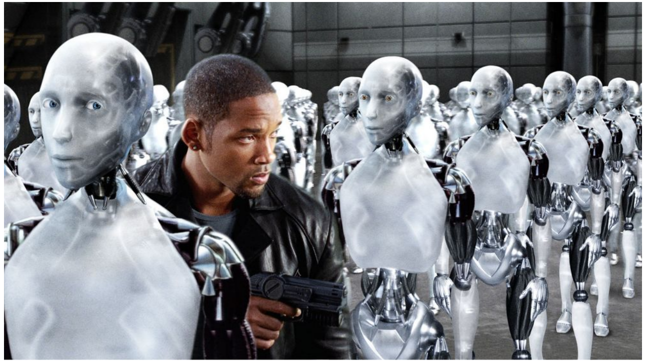
Analytics

- Scientific experiments
  - produce big data
  - incorporate results.
- Big data analytics
  - delivers views of the data, summaries, predictions
  - controls noise and error.
- Foundations
  - Sciences
  - Computer architectures
    - GPU, FPGA, Multicore,...
  - Software frameworks
    - Streams, Hadoop, SQL,...
  - Algorithms
  - Statistical models

Artificial Intelligence

- AI is about *computer* behavior that – if performed by *humans* – would be considered intelligent.
- It is about a task – humans perform many tasks!
- What we consider "intelligent":
  - Calculation, memorizing is no longer considered intelligent. Because *computers* are good at it?
  - Tying shoelaces is very difficult for robots. Are we as *humans* proud of this skill?

- AI since 1956 (USA), 1976 (Germany)
- AI as a model of cognition
- AI as a performer
- AI as an analogy or for science fiction



Artificial Intelligence classics

- Planning
- Logical Inference, Reasoning
- Knowledge Representation
- Natural Language Processing
- Machine Learning
- Multi-Agent Systems
- Robotics
- Cognition
- Vision Understanding, Computer Graphics
- Games

1996 DeepBlue defeats Kasparow – AI without machine learning

Machine Learning is

- Part of computer science with applications in all other areas.
- Based on data
  - Well acquired data (Excel)
  - Given data bases (SQL)
  - Big data (Hadoop distributed file system HDFS)
  - Structured data (graphs, facts)
- Implemented on an architecture
- Delivering an action.

Classification

- Given:
  examples $\{(x_1, y_1), (x_2, y_2)\dots\}$
- Find:
  predictions $f: X \rightarrow Y$,

  such that the quality $Q(Y, f(X))$ is maximized.

- Risk:

$$R(Y, f(X)) = \sum_{i=1}^{N} Q(y_i, x_i) p_i$$

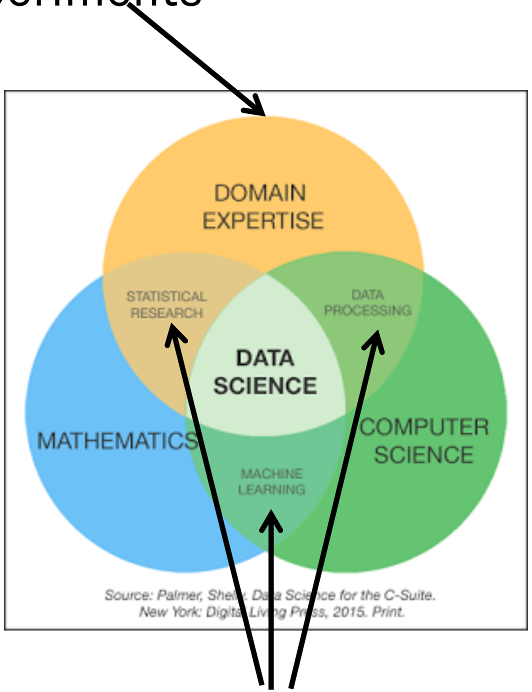
Classes of machine learning algorithms

- Induction of Decision Trees
- Neural Networks, Convolutional Neural Networks (deep learning)
- Support Vector Machines
- Clustering
- Probabilistic Graphical Models
- Frequent Itemset Mining
- Reinforcement, Q Learning
- Time Series Classification, Clustering, Prediction

2016 alphaGo defeats Lee Sedol – AI with machine learning

Machine learning

Experiments

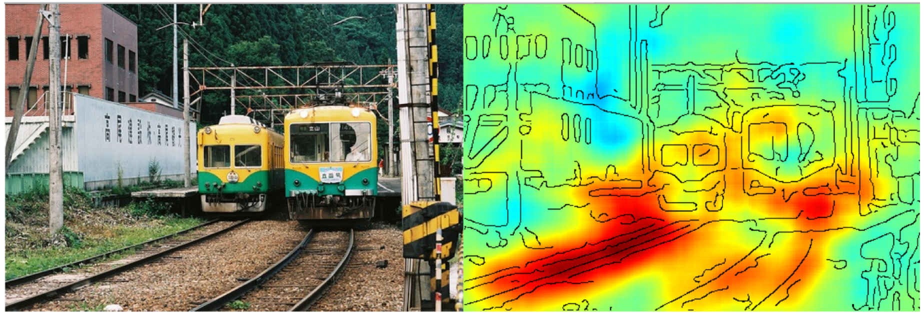


Analytics

- What makes computers learn?
- Tight bounds for
  - Correctness
    - Precision, recall, AUC, …
  - Efficiency
    - Runtime, memory, communication, energy
  - Robustness
  - Fairness
- Inspection of data, interaction with experts

Deep learning

- Machine learning algorithms need a sound theoretical basis.
  - Proven tight quality guarantees.
  - Robust against minor changes in the input data.
  - Verified combination of methods.
- Applying machine learning requires to know the theoretical foundations of machine learning.



Klaus-Robert Müller (TU Berlin), Wojciech Samek (HHI, Berlin) developed a certification method for Deep Learning. Here, it shows by heat maps that the classification is based on features of the environment, not of the object. This phenomenon is well known for DeepNeural Networks. DOI: 10.1038/s41467-019-08987-4

Graphical Models

- Graph G=(V,E)
- *Sufficient statistic* implicitly mapping joint vertex assignment into vector space
  $\phi(\vec{x}): \mathcal{X} \longrightarrow \mathcal{R}^d$
- *Parameter vector* to be learned:
  θ in $\mathcal{R}^d$
- Log *partition function*:
  $$A(\vec{\theta}) = \ln \sum_i \exp\left(\langle \vec{\theta}, \phi(\vec{x}_i) \rangle\right)$$

CRF:

$$p(\vec{y}|\vec{x}) = \frac{1}{Z(\vec{\theta},\vec{x})} \exp\left(\sum_i \theta_i \phi_i(\vec{y},\vec{x})\right) \quad \left| \frac{1}{a} = \exp(-\ln a) \right.$$

$$= \exp\left[\left(\sum_i \theta_i \phi_i(\vec{y},\vec{x})\right) - \ln Z(\vec{\theta},\vec{x})\right]$$

$$= \exp\left[\langle \vec{\theta}, \phi(\vec{y},\vec{x}) \rangle - A(\vec{\theta})\right]$$

MRF:

$$p(\vec{x}) = \frac{1}{Z(\vec{\theta})} \exp\left(\sum_i \theta_i \phi_i(\vec{x})\right)$$

$$= \exp\left[\langle \vec{\theta}, \phi(\vec{x}) \rangle - A(\vec{\theta})\right]$$

aka Information Field Theory (Torsten Ensslin)

Probabilistic inference

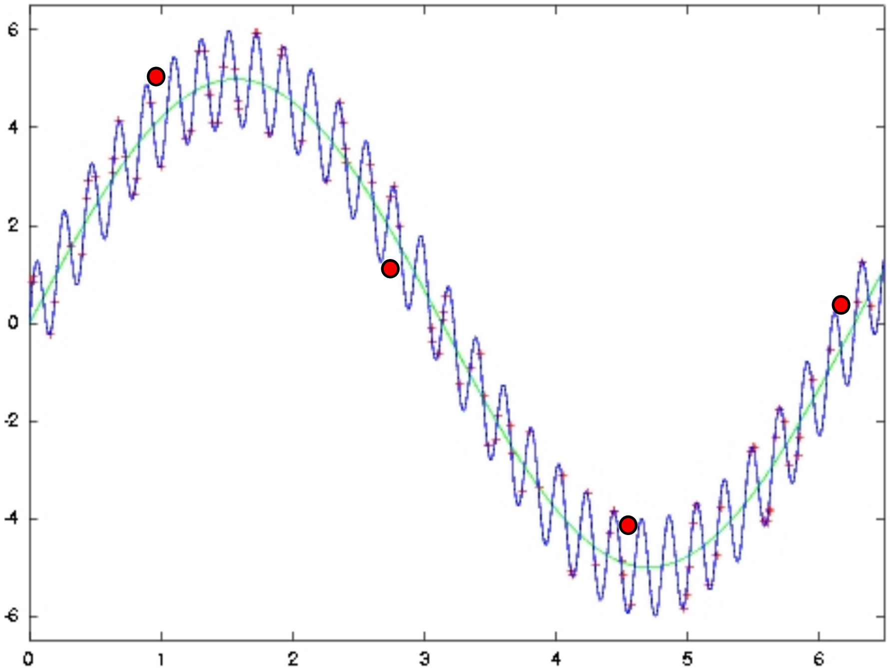
- Marginal probabilities
$$p\left(X_i = x_i | X_j = x_j\right)$$
- Partition function
$$Z(\theta) = \int \psi(x) d\, \nu(x)$$
- Maximum A Posteriori (MAP) state
$$\max_{x \in \aleph} p(x)$$
- Maximum likelihood parameter
$$\max_{\theta \in \Re^d} \prod_{x \in D} p_\theta(x)$$

- Complexity of computing the partition function depends on hardness of the integration. Closed form known for Gaussian, Poisson, Laplace, Weibull, ... distributions.
- For discrete random variables, belief propagation is efficient for tree structured graphs.

Regularization, Reparametrization

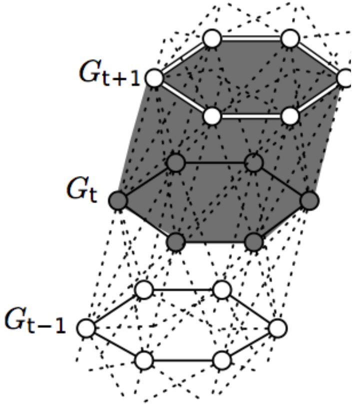
- Regularizations decrease the model complexity.
  - $L_{1,2}$ against overfitting
- Reparametrizations map the parameters to another space
  - Vector $\Delta \in \mathcal{R}^k$ is a reparametrization of θ
  - $\eta: \mathcal{R}^k \to \mathcal{R}^d$ and $\eta(\Delta) = \theta$

Attention! Changing the number of parameters changes the independence structure.
Look there!



Spatio-temporal random fields

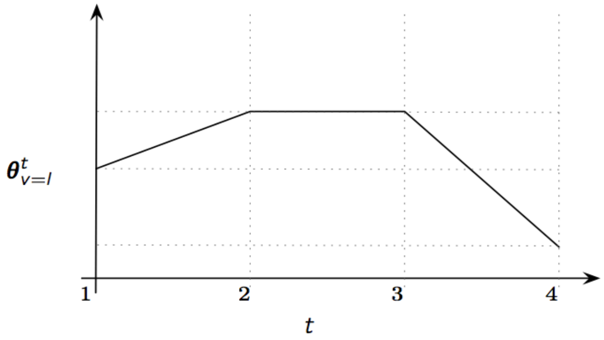
- The spatio-temporal graph is trained to predict each node's maximum a posteriori probability with the marginal probabilities.
  - Generative model predicting all nodes.
- Dimension
  $T \times |V_0| \times |\mathcal{X}| +$
  $[(T-1)(|V_0|+3|E_0|)+ |E_0|] \times |\mathcal{X}|^2$
- If edges in some subset represent similar relations and have a common state space, then they may share parameters.
- **Proof** (Piatkowski 2018, p. 83 ff.)
  - Distance between true θ and estimated $\eta(\Delta)$ is bounded
  - Sparsity in the estimate implies redundancy in the true parameter
- Keeping the quality,
  regularization and reparametrization saves memory and learning becomes faster.



Reparametrization compresses the model

- Reparametrize model
  $\Delta_t \approx \theta_{t+1} - \theta_t$
  Δ regularized by L1, L2 norm
- Quality is not at all less than MRF, 4NN.
- Learning is faster.
- There are not many changes over time. Model is highly compressed.

Assumption: Smoothness over time



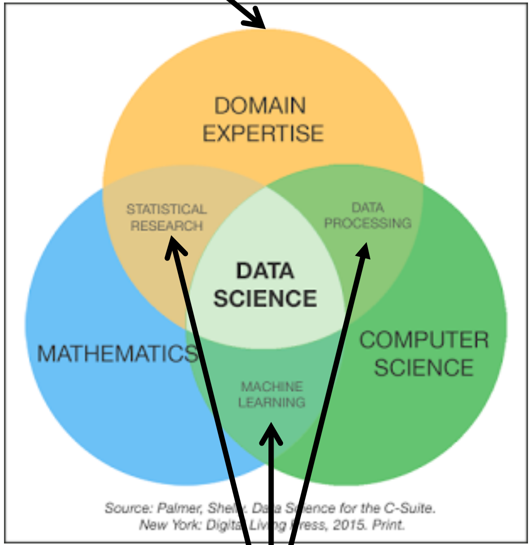
Idea:

- Remove the near zero slopes, while retaining the performance

Several other proofs on error bounds available in literature.
For proofs regarding resource constraints s. Piatkowski 2018

Illustration of what machine learning research looks like

Experiments



Analytics

Probabilistic graphical models

- Tight bounds for quality, while reducing
  - Memory
  - Energy
  - Runtime
- Theoretically well-based, not heuristic,
- Carefully implemented,
- Empirically tested on several data sets.

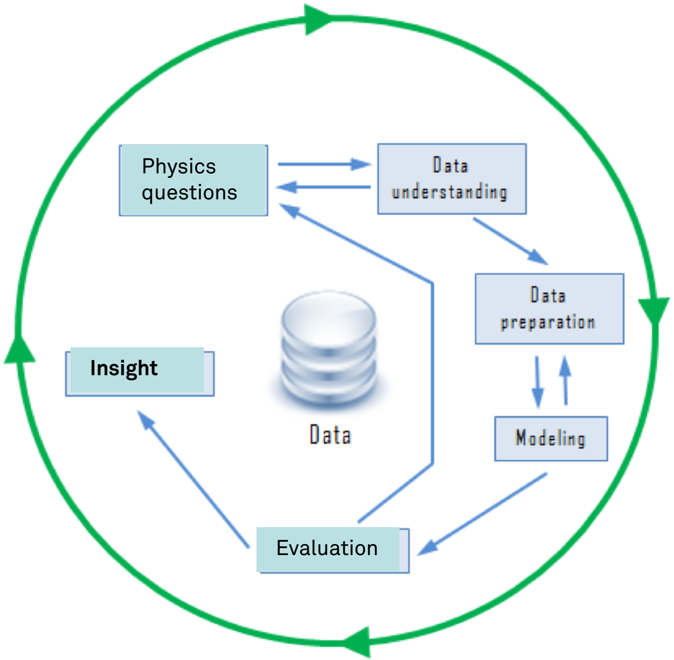
Overview

- AI and ML
  - History of AI
  - Machine learning (ML)
  - Illustration of ML research
- Physics and ML
  - Ice Cube, FACT, CTA, SKA
  - Deep learning on FACT
  - SVM on FACT
  - Unfolding on FACT
  - FPGA for FACT

Crab nebular around crab pulsar
Mixed optical and X-Ray spectrum

Astrophysics

- Explanation of dark matter and the apparent lack of antimatter in the universe
- Find experimental indications for physics beyond the Standard Model
- Astroparticle physics:
  - creation by dark matter annihilations;
  - journey of high energy particles interstellar and intergalactic space.

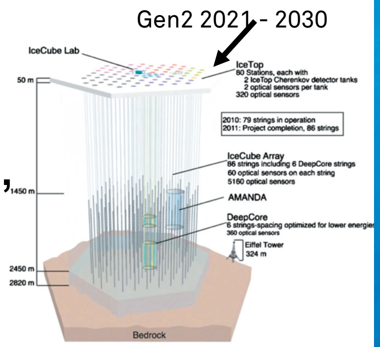


IceCube

Extremely high-volume data 1 TB per day

Data analysis

- form the trace of the neutrino,
- separate it from other particles,
- estimate the energy



Breakthrough 2018

22.September 2017, a very high-energy neutrino is detected by IceCube. It points at the blazar TXS 0506+056, 4 billion light years from Earth.

Gamma ray observatories identify the same source of high-energy gamma rays.

technische universität dortmund

ML2R

SFB 876 Providing Information by Resource-Constrained Data Analysis

Interdisciplinary, empirical work for theory development

arXiv:1409.4535v1 [astro-ph.HE] 16 Sep 2014

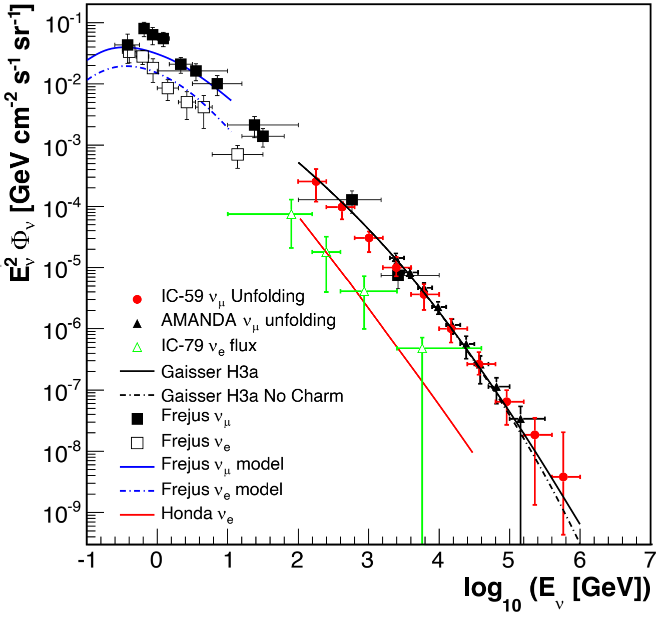
Eur. Phys. J. C manuscript No.
(will be inserted by the editor)

Development of a General Analysis and Unfolding Scheme and its Application to Measure the Energy Spectrum of Atmospheric Neutrinos with IceCube

IceCube Collaboration: M. G. Aartsen[2], M. Ackermann[46], J. Adams[15], J. A. Aguilar[24], M. Ahlers[29], M. Ahrens[37], D. Altmann[23], T. Anderson[43], C. Arguelles[29], T. C. Arlen[43], J. Auffenberg[1], X. Bai[35], S. W. Barwick[26], V. Baum[30], J. J. Beatty[17,18], J. Becker Tjus[10], K.-H. Becker[45], S. BenZvi[29], P. Berghaus[46], D. Berley[16], E. Bernardini[46], A. Bernhard[32], D. Z. Besson[27], G. Binder[8,7], D. Bindig[45], M. Bissok[1], E. Blaufuss[16], J. Blumenthal[1], D. J. Boersma[44], C. Bohm[37], F. Bos[10], D. Bose[39], S. Böser[11], O. Botner[44], L. Brayeur[13], H. P. Bretz[46], A. M. Brown[15], J. Casey[5], M. Casier[13], E. Cheung[16], D. Chirkin[29], A. Christov[24], B. Christy[16], K. Clark[40], L. Classen[23], F. Clevermann[20], S. Coenders[32], D. F. Cowen[43,42], A. H. Cruz Silva[46], M. Danninger[37], J. Daughhetee[5], J. C. Davis[17], M. Day[29], J. P. A. M. deAndré[43], C. De Clercq[13], S. De Ridder[25], P. Desiati[29], K. D. de Vries[13], M. de With[9], T. DeYoung[43,a], J. C. Díaz-Vélez[29], M. Dunkman[43], R. Eagan[43], B. Eberhardt[30], B. Eichmann[10], J. Eisch[29], S. Euler[44], P. A. Evenson[33], O. Fadiran[29], A. R. Fazely[6], A. Fedynitch[10], J. Feintzeig[29], J. Felde[16], T. Feusels[25], K. Filimonov[7], C. Finley[37], T. Fischer-Wasels[45], S. Flis[37], A. Franckowiak[11], K. Frantzen[20], T. Fuchs[20], T. K. Gaisser[33], R. Gaior[14], J. Gallagher[28], L. Gerhardt[8,7], D. Gier[1], L. Gladstone[29], T. Glüsenkamp[46], A. Goldschmidt[8], G. Golup[13], J. G. Gonzalez[33], J. A. Goodman[16], D. Góra[46], D. Grant[22], P. Gretskov[1], J. C. Groh[43], A. Groß[32], C. Ha[8,7], C. Haack[1], A. Haj Ismail[25], P. Hallen[1], A. Hallgren[44], F. Halzen[29], K. Hanson[12], D. Hebecker[11], D. Heereman[12], D. Heinen[1], K. Helbing[45], R. Hellauer[16], D. Hellwig[1], S. Hickford[15], G. C. Hill[2], K. D. Hoffman[16], R. Hoffmann[45], A. Homeier[11], K. Hoshina[29,b], F. Huang[43], W. Huelsnitz[16], P. O. Hulth[37], K. Hultqvist[37], S. Hussain[33], A. Ishihara[14], E. Jacobi[46], J. Jacobsen[29], K. Jagielski[1], G. S. Japaridze[4], K. Jero[29], O. Jlelati[25], M. Jurkovic[32], B. Kaminsky[46], A. Kappes[23], T. Karg[46], A. Karle[29], M. Kauer[29], A. Keivani[43], J. L. Kelley[29], A. Kheirandish[29], J. Kiryluk[38], J. Kläs[45], S. R. Klein[8,7], J. H. Köhne[20], G. Kohnen[31], H. Kolanoski[9], A. Koob[1], L. Köpke[30], C. Kopper[29], S. Kopper[45], D. J. Koskinen[19], M. Kowalski[11], A. Kriesten[1], K. Krings[1], G. Kroll[30], M. Kroll[10], J. Kunnen[13], N. Kurahashi[29], T. Kuwabara[14], M. Labare[25], D. T. Larsen[29], M. J. Larson[19], M. Lesiak-Bzdak[38], M. Leuermann[1], J. Leute[32], J. Lünemann[30], J. Madsen[36], G. Maggi[13], R. Maruyama[29], K. Mase[14], H. S. Matis[8], R. Maunu[16], F. McNally[29], K. Meagher[16], M. Medici[19], A. Meli[25], T. Meures[12], S. Miarecki[8,7], E. Middell[46], E. Middlemas[29], N. Milke[20], J. Miller[13], L. Mohrmann[46], T. Montaruli[24], R. Morse[29], R. Nahnhauer[46], U. Naumann[45], H. Niederhausen[38], S. C. Nowicki[22], D. R. Nygren[8], A. Obertacke[45], S. Odrowski[22], A. Olivas[16], A. Omairat[45], A. O'Murchadha[12], T. Palczewski[41], L. Paul[1], Ö. Penek[1], J. A. Pepper[41], C. Pérez de los Heros[44], C. Pfendner[17], D. Pieloth[20], E. Pinat[12], J. Posselt[45], P. B. Price[7], G. T. Przybylski[8], J. Pütz[1], M. Quinnan[43], L. Rädel[1], M. Rameez[24], K. Rawlins[3], P. Redl[16], I. Rees[29], R. Reimann[1], M. Relich[14], E. Resconi[32], W. Rhode[20], M. Richman[16],

2

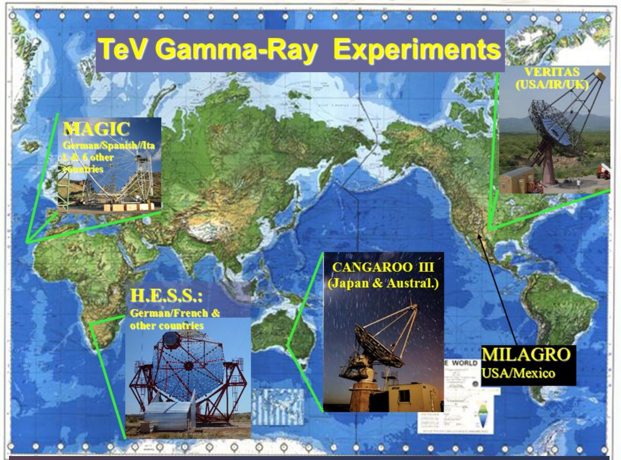
B. Riedel[29], S. Robertson[2], J. P. Rodrigues[29], M. Rongen[1], C. Rott[39], T. Ruhe[20], B. Ruzybayev[33], D. Ryckbosch[25], S. M. Saba[10], H.-G. Sander[30], J. Sandroos[19], M. Santander[29], S. Sarkar[19,34], K. Schatto[30], F. Scheriau[20], T. Schmidt[16], M. Schmitz[20], S. Schoenen[1], S. Schöneberg[10], A. Schönwald[46], A. Schukraft[1], L. Schulte[11], O. Schulz[32], D. Seckel[33], Y. Sestayo[32], S. Seunarine[36], R. Shanidze[46], M. W. E. Smith[43], D. Soldin[45], G. M. Spiczak[36], C. Spiering[46], M. Stamatikos[17,c], T. Stanev[33], N. A. Stanisha[43], A. Stasik[11], T. Stezelberger[8], R. G. Stokstad[8], A. Stößl[46], E. A. Strahler[13], R. Ström[44], N. L. Strotjohann[11], G. W. Sullivan[16], H. Taavola[44], I. Taboada[5], A. Tamburro[33], A. Tepe[45], S. Ter-Antonyan[6], A. Terliuk[46], G. Tešić[43], S. Tilav[33], P. A. Toale[41], M. N. Tobin[29], D. Tosi[29], M. Tselengidou[23], E. Unger[44], M. Usner[11], S. Vallecorsa[24], N. van Eijndhoven[13], J. Vandenbroucke[29], J. van Santen[29], M. Vehring[1], M. Voge[11], M. Vraeghe[25], C. Walck[37], M. Wallraff[1], Ch. Weaver[29], M. Wellons[29], C. Wendt[29], S. Westerhoff[29], B. J. Whelan[2], N. Whitehorn[29], C. Wichary[1], K. Wiebe[30], C. H. Wiebusch[1], D. R. Williams[41], H. Wissing[16], M. Wolf[37], T. R. Wood[22], K. Woschnagg[7], D. L. Xu[41], X. W. Xu[6], J.P. Yanez[46], G. Yodh[26], S. Yoshida[14], P. Zarzhitsky[41], J. Ziemann[20], S. Zierke[1], M. Zoll[37] and K. Morik[21]



The senses of astrophysics

Observatories of many kinds all around the earth.

- Gravitational wave observatories
  LIGO Washington, Lousiana; VIRGO Italy
- IceCube at the south pole looks for neutrinos, Gen2 – 100 TB /day
- Cherenkov telescopes around the world: MAGIC, H.E.S.S., FACT – 1 TB / day
- CTA North 20 telescopes, South 100 telescopes – 100 TB /day
- Square kilometer array (SKA) Australia, South Africa– 100 PB / day



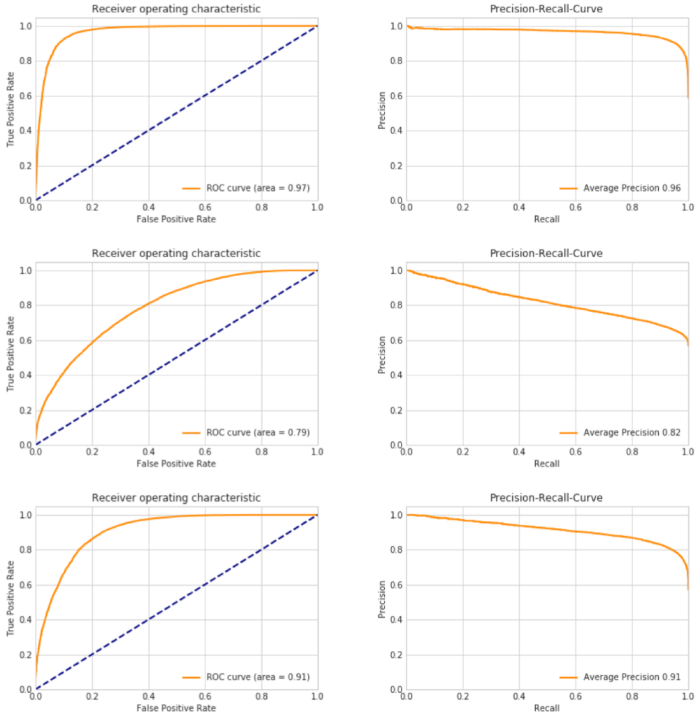
Overview

- AI and ML
  - History of AI
  - Machine learning (ML)
  - Illustration of ML research
- Physics and ML
  - Ice Cube, FACT, CTA, SKA
  - Deep learning on FACT
  - SVM on FACT
  - Unfolding on FACT
  - FPGA for FACT

Crab nebular around crab pulsar
Mixed optical and X-Ray spectrum

Machine learning analysis – deep learning on FACT

- Designing the architecture
  - 2D-CNN on cleaned images (width, height, RGB) for signal-hadron-separation
  - 3D-CNN including time on raw data (some frames over time at each position) end-to-end learning
- Training, parameter optimization
  - On ground truth, on aggregated data, on every second frame (sliced)
- Problem:
  No theoretically well based guarantees!



2D binary

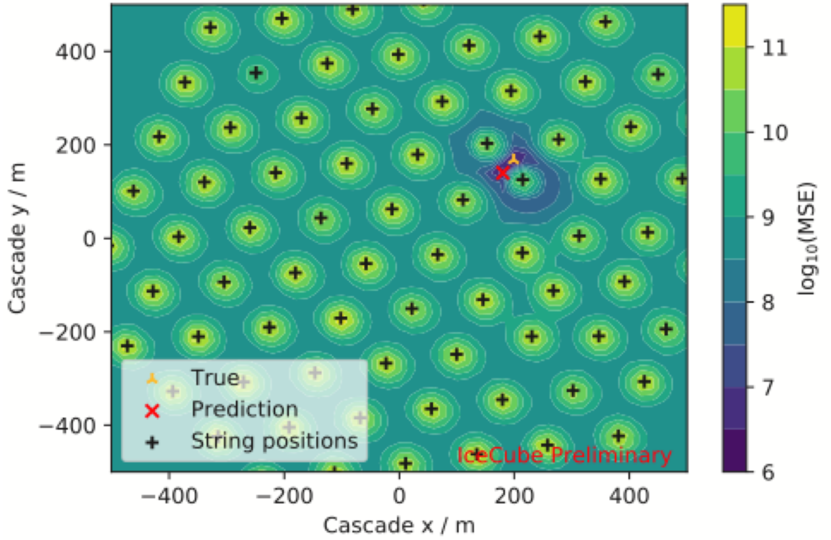
2D mean

3D sliced

Stefan Rötner „Deep Learning on Raw Telescope Data“

Machine learning analysis – deep learning on IceCube

- Reconstruction of events by CNN
  - alerts to telescopes around the world in real-time
  - Muon energy resolution
- Problem:
  no use of prior knowledge (as is done by probabilistic models)
- Prediction of waveform at each DOM by generative neural network
  - Train on all 7 parameters
  - Fix 5, predict two



z,t, azimuth, zenith, energy fixed

IceCube Collaboration, Mirco Hünnefeld „Reconstruction Techniques in IceCube…“

Overview

- AI and ML
  - History of AI
  - Machine learning (ML)
  - Illustration of ML research
- Physics and ML
  - Ice Cube, FACT, CTA, SKA
  - Deep learning on FACT
  - SVM on FACT
  - Unfolding on FACT
  - FPGA for FACT

Crab nebular around crab pulsar
Mixed optical and X-Ray spectrum

Machine learning analysis – SVM on FACT

Support vector machine (SVM) determines an optimal separating hyperplane, proofs given.

- Optimization in O(n^3)

  $W(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i}^{n} \sum_{j}^{n} y_i y_j \alpha_i \alpha_j \langle x_i, x_j \rangle$

- Kernel functions, RBF, arccos

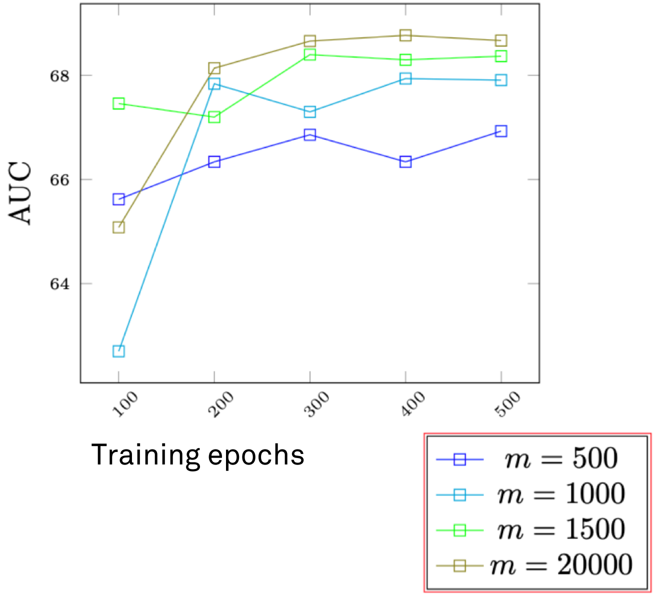
  $$\theta = arccos\left(\frac{x\,y}{\|x\|\|y\|}\right)$$

  $$J_1(\theta) = sin\theta + (\pi - \theta)cos\theta$$

- Nyström approximation of the kernel matrix for less memory and runtime O(nm^2)
  Split n X n kernel matrix into
  m X m, n X m -- m X n need not be calculated

FACT gamma separation

- 170 000 examples



Martin Senz „Effiziente Kernel Basierte Klassifikation von Teleskop-Daten…“

Overview

- AI and ML
  - History of AI
  - Machine learning (ML)
  - Illustration of ML research
- Physics and ML
  - Ice Cube, FACT, CTA, SKA
  - Deep learning on FACT
  - SVM on FACT
  - Unfolding on FACT
  - FPGA for FACT

Crab nebular around crab pulsar
Mixed optical and X-Ray spectrum

Machine learning analysis – unfolding on FACT

- Energy distribution of gamma particles

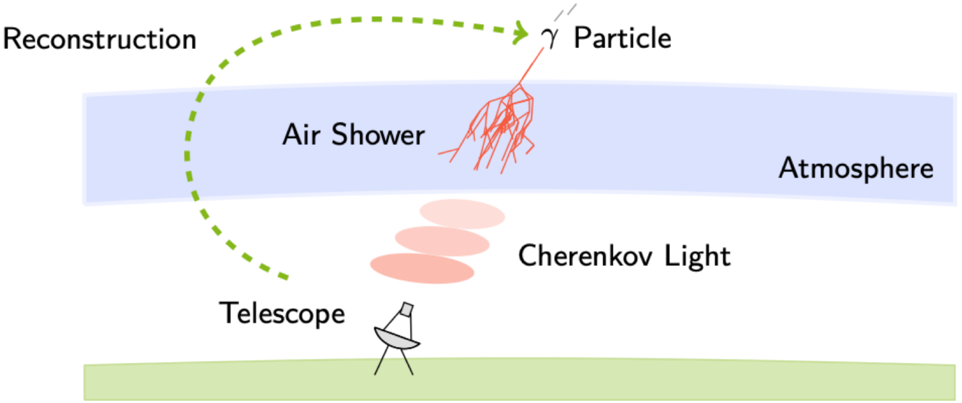
  $f : Y \to \Re$

- is reconstructed from measurements

  $g : X \to \Re$

- Detector matrix $R_{i,j}$

$$\int_{Y_i} \int_{X_j} R(x|y)\, dx\, dy$$



Detector matrix cannot be inverted

$$\hat{f}_{impossible} = R^{-1} g$$

Iterative Bayesian unfolding (IBU)

- Bayesian approach
  - Estimate target density using Bayes law.
- The prior is not accurate.
- Iterating replacements of the prior with the estimate of the previous iteration.
- Problem: if the prior is updated too often, the result diverges from good estimate.
- Stopping criterion needed!

prior

$$\hat{P}(Y=i \mid X=j) = \frac{\hat{P}(X=j \mid Y=i)\hat{P}(Y=i)}{\sum_{i'=1}^{J} \hat{P}(X=j \mid Y=i)\hat{P}(Y=i)}$$

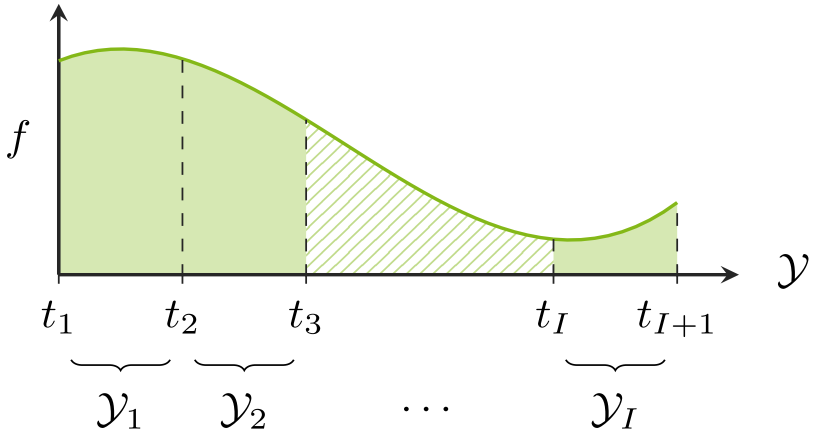
$$\hat{f}_i = \sum_{j=1}^{J} \hat{P}(Y=i \mid X=j) g_j$$

$$\hat{f}_i^{(k)} = \sum_{J=1}^{J} \frac{R_{ij}\hat{f}_i^{(k-1)}}{\sum_{i'=1}^{I} R_{i'j}\hat{f}_{i'}^{(k-1)}} g_j$$

DSEA – Deconvolution as classification

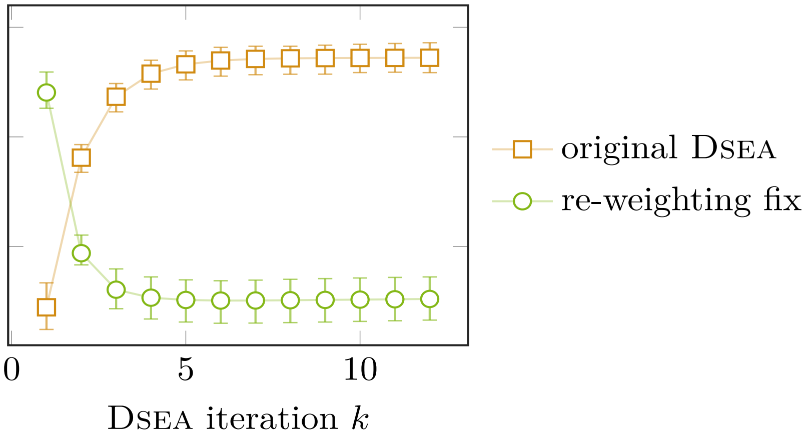
- Turning deconvolution into a classification problem
- Discretize the target energy, consider it a class label.
- Receive the histogram from simulation data.
- Learn a classifier with confidence.
- Output confidence for each example.

$$f_i = P(Y = i) = \int_{Y_i} f(y)\,dy \quad \forall 1 \le i \le I$$



Enhancements DSEA+

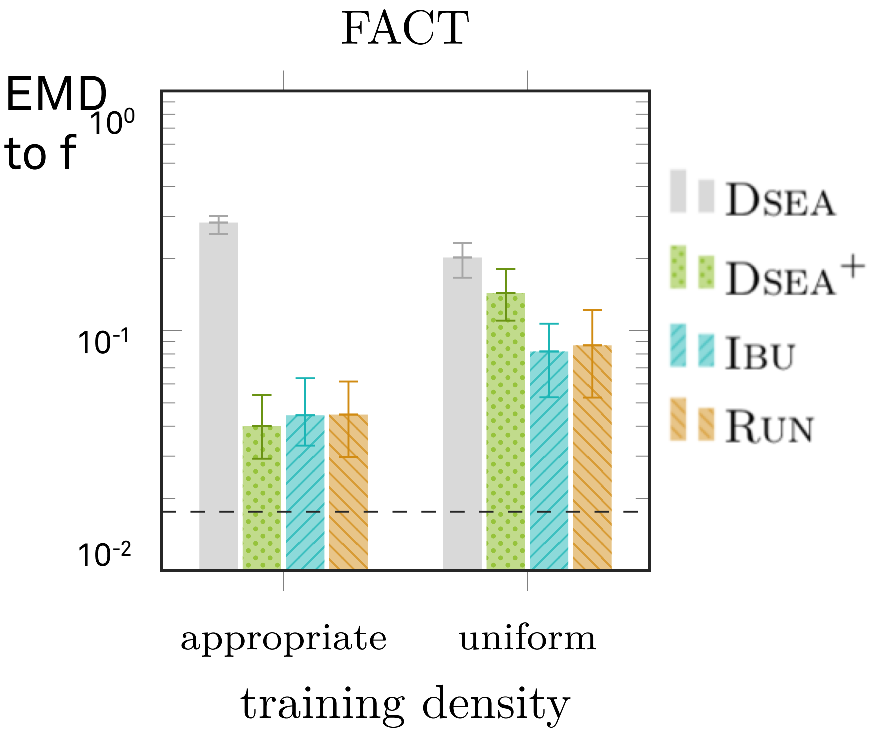
- DSEA diverges from the optimal estimate in further iterations.
- DSEA+ re-weights the examples in each iteration by the ratio between the estimated probability and the probability of the unweighted training set.
- DSEA+ speeds up by adaptive step size.



Earth Mover Distance (EMD): cost of transforming one histogram to another

DSEA+ Results

- The accuracy of DSEA , Ibu, and Run, Earth Mover’s Distance between the unfolding results and the corresponding true densities: the lower the better.
- Each bar is obtained with the best configuration evaluated on 20 bootstrap samples.
- The FACT training set is too small for uniform sub-sampling. Results are better for balanced classes.



Overview

- AI and ML
  - History of AI
  - Machine learning (ML)
  - Illustration of ML research
- Physics and ML
  - Ice Cube, FACT, CTA, SKA
  - Deep learning on FACT
  - SVM on FACT
  - Unfolding on FACT
  - FPGA for FACT

Crab nebular around crab pulsar
Mixed optical and X-Ray spectrum

Machine learning analysis – Random Forest on FACT

- Filtering sensor data when they arrive.
- Random Forest on FPGA or CPU – what is faster?
- FPGAs can even compute all comparisons in 1 clock cycle and then traverse the tree using the precomputed.

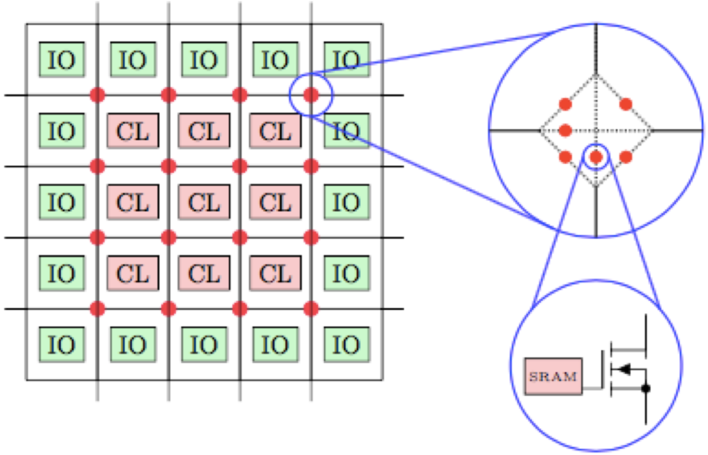


Fig. 4. Signal routing in FPGAs. Each crossing has 6 transistor attached to it, which control each of the lanes. Transistors are programmed using a 1 bit SRAM cell.

Comparison FPGA, ARM

- Concerning energy consumption FPGA outperforms CPU.
- For small decision trees, FPGAs are fine.
- Disjunctive normal form representation of trees did not pay off.
- New implementation with unrolling and precomputing gives excellent throughput on ARM.

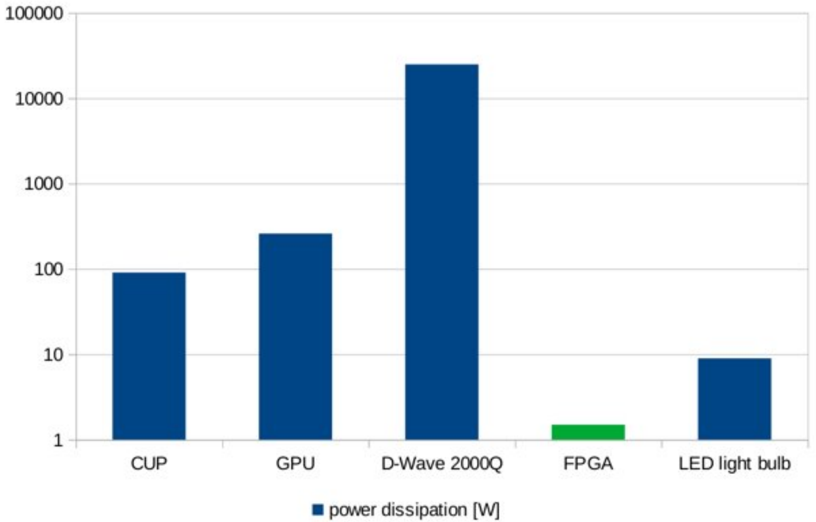
TABLE I
THROUGHPUT COMPARISON FOR DIFFERENT IMPLEMENTATIONS OF RANDOM FORESTS AND DECISION TREES ON THE ZEDBOARD.

	FPGA $\left[\frac{\text{elem}}{\text{ms}}\right]$	ARM $\left[\frac{\text{elem}}{\text{ms}}\right]$
If-Tree	$1480 \pm 2.7 \cdot 10^{-9}$	29000 ± 0.0027
If-Tree-Forest	-	$780 \pm 2.7 \cdot 10^{-9}$
Native-Tree	1170 ± 0.00034	14500 ± 0.00054
Native-Tree-Forest	-	$460 \pm 4.9 \cdot 10^{-9}$
DNF-Tree	$1100 \pm 4.7 \cdot 10^{-10}$	$1900 \pm 4.9 \cdot 10^{-9}$
DNF-Tree-Forest	-	$30 \pm 1.4 \cdot 10^{-13}$

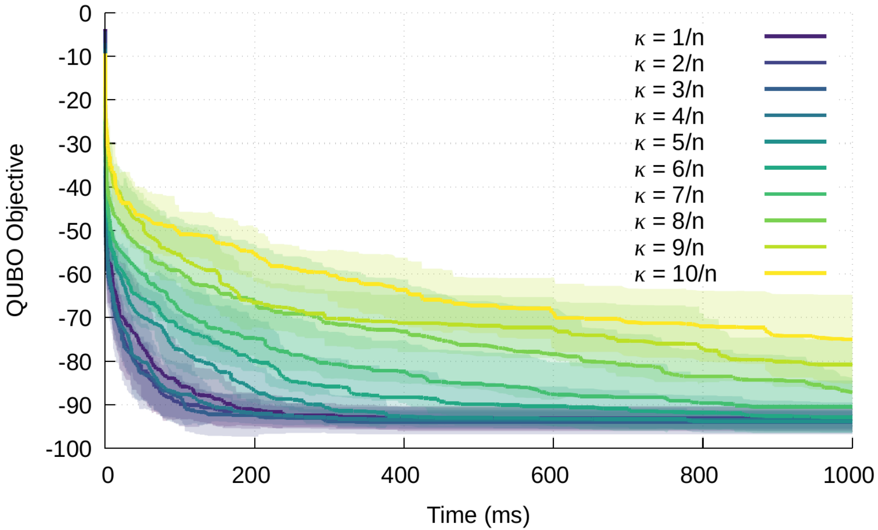
S. Buschjäger, K.Morik, *Decision tree and random forest implementations for fast filtering of sensor data*

Direct implementation of machine learning on FPGA

Convergence of an MRF MAP estimation problem on 500 points of a 23-dimensional dataset (*mushroom*) using an FPGA implementation that applies Evolutionary Optimization, comparing multiple per-bit mutation probabilities κ.



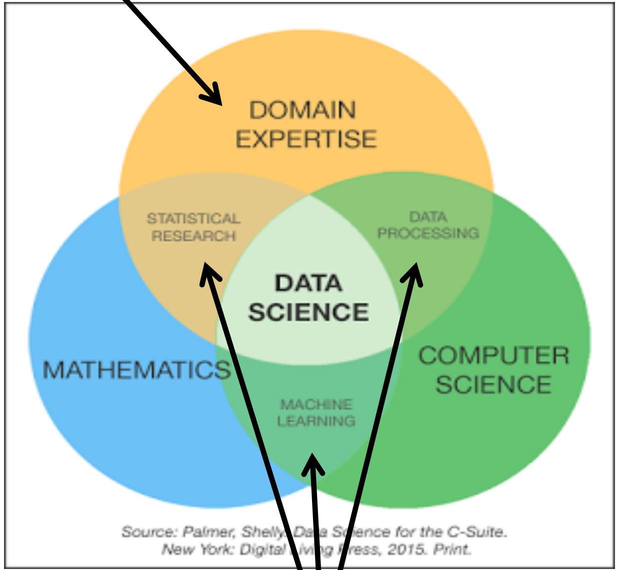
Comparison of power dissipation for multiple modern hardware components typically used for Machine Learning (+ light bulb for reference, log. scale)



S.Mücke, *Evolutionäre Optimierung pseudoboolescher Funktionen auf FPGA*

Machine learning is an instrument of science

Experiments



Source: Palmer, Shelly. Data Science for the C-Suite. New York: Digital Living Press, 2015. Print.

Analytics

- Sometimes the ground truth cannot be grasped by a human expert.
- It is hidden
  - in large volume, large dimension data
  - in noise
  - in nanoscopic or faraway structures.
- **Machine learning digs into data from detectors lifting insights that were out of reach before.**
- Learned models must come with proven guarantees of their properties.

- **Still a lot of work to do!**

Summary

- AI and ML
  - What machine learning (ML) is
  - Many classes of algorithms
  - What ML research does: graphical models
- Physics and ML
  - Experiments: Ice Cube, FACT, CTA, SKA
  - Some learning algorithms on FACT data:
    - Deep learning
    - SVM with Nyström approximation
    - Unfolding as classification
    - Random Forest, QUBO on FPGA,

Crab nebular around crab pulsar
Mixed optical and X-Ray spectrum

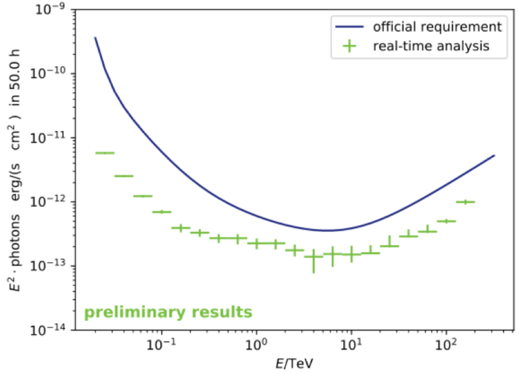
Project C3
Multi-level statistical analysis of high-frequency spatio-temporal process data
Prof. Dr. Katharina Morik, Prof. Dr. Dr. Wolfgang Rhode, Dr. Tim Ruhe

Future work in SFB 876

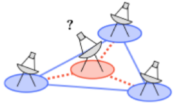
Distributed Learning for Telescope Arrays

CTA [Brügge et al. 2017]

- Transfer FACT experience to CTA
- Prototype already meets current CTA real-time requirements
- **Horizontal scalability** through Apache Flink for future requirements
- Recording starts by the end of 2018



- Observations may be **incomplete**
  - Broken camera pixels
  - Sky partially obstructed by clouds
- PGM can fill missing information (A1)



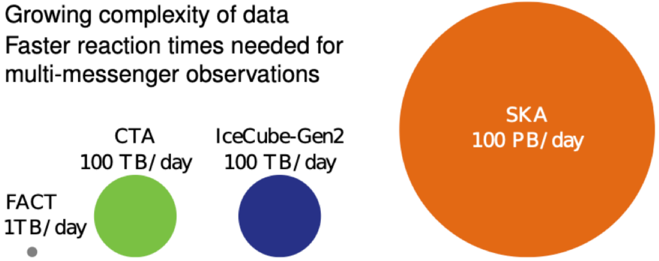
$$\mathbb{P}(T_1 \mid T_2 \cap T_3) = \frac{\mathbb{P}(T_1 \cap T_2 \cap T_3)}{\mathbb{P}(T_2 \cap T_3)}$$

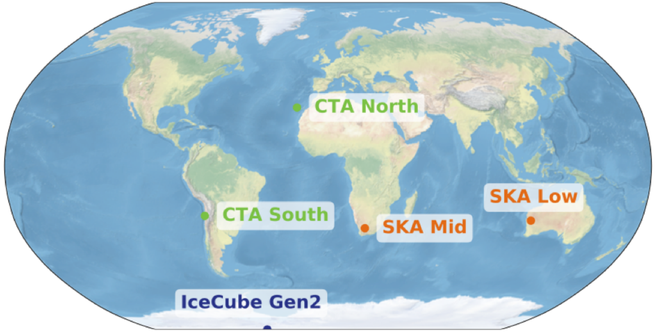
SKA

- From streams of events to a continuous data stream

Challenges Ahead

- Next generation of experiments now being planned and built
- Detector arrays instead of single telescopes
- Dramatic increase in data volumes
- Growing complexity of data
- Faster reaction times needed for multi-messenger observations





References

- V. Blobel, *An unfolding method for high energy physics experiments*, Adv. Stat. Techniques in Particle Physics, 2002.
- C.Bockermann, K. Brügge, J. Buss, A. Egorov, K.Morik, W. Rhode, T.Ruhe, *Online Analysis of High-Volume Data Streams in Astroparticle Physics,* ECML PKDD 2015
- Brügge, K., Egorov, A., Bockermann, C., Morik, K. and Rhode, W. *Distributed Real-Time Data Stream Analysis for CTA* ADASS XXVII, 2017
- M. Bunse, N. Piatkowski, T. Ruhe W. Rhode, K. Morik, *Unification of Deconvolution Algorithms for Cherenkov Astronomy*, IEEE Int. Conf. Data Science and Advanced Analytics, 2018
- S. Buschjäger, K.Morik, *Decision tree and random forest implementations for fast filtering of sensor data,* IEEE Transaction on Circuits and Systems, issue 99, 2017
- G. D'Agostini, *Improved iterative Bayesian unfolding*, arXiv:1010.0632, 2010.
- IceCube Collaboration and M. Hünnefeld, *Reconstruction techniques in IceCube using convolutional and generative neural networks*, EPJ Web of Conferences 207, 2019
- IceCube Collaboration and K. Morik, *Development of a general analysis and unfolding scheme and its application to measure the energy spectrum of atmospheric neutrinos with IceCube*, The European Physical Journal C, Volume 75, 2015.
- G. Montavon, W. Samek, K-R. Müller, *Methods for interpreting and understanding deep neural networks,* Digital Signal Processing 73, 2018.
- S.Mücke, *Evolutionäre Optimierung pseudoboolescher Funktionen auf FPGA*, Master thesis TU Dortmund 2019
- Piatkowski, N. *Exponential Families on Resource-Constrained Systems,* Ph. D. thesis, TU Dortmund 2018

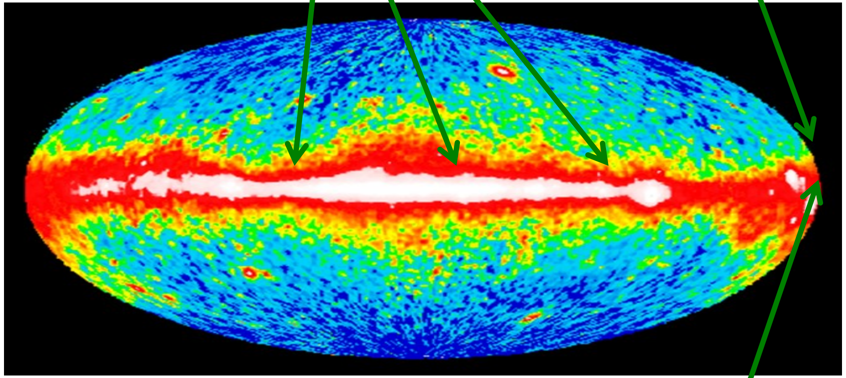
References

- Rötner, S. *Deep learning on raw telescope data,* Diploma Thesis TU Dortmund 2017
- T. Ruhe, M. Börner, M. Wornowizki, T. Voigt, W. Rhode, K. Morik, *Mining for spectra – the Dortmund spectrum estimation algorithm*, ADASS XXVI, 2016.
- Senz, M. *Effiziente kernelbasierte Klassifikation von Teleskop-Daten durch die Anwendung der Nyström-Approximation,* Master Thesis TU Dortmund 2018

The gamma ray sky

Milky way

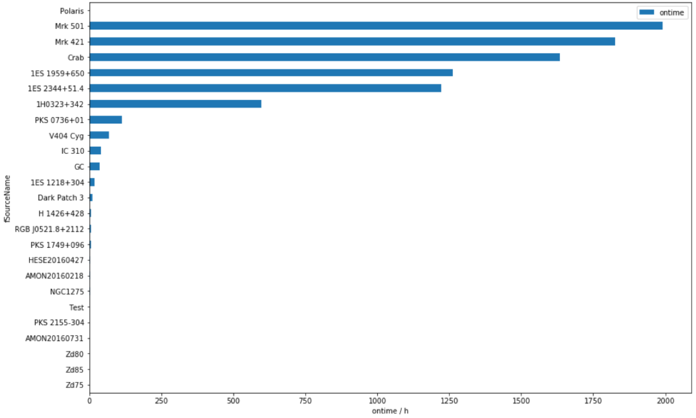
Geminga pulsar



seen by the satellite EGRET

Crab (SNR)

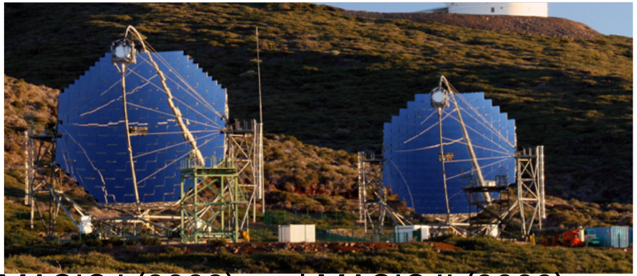
- Measurement hours of sources of gamma rays measured by FACT



hours

Skewed distribution challenging data analysis

- Calibration, cleaning
- Feature extraction
- Signal separation
- Energy estimation
- A simulator provides labeled observations.
- Gamma rays of high energy are rare events as opposed to hadrons, ratio 1 to 1000.



MAGIC I (2003) and MAGIC II (2009)
La Palma, Roque de los Muchachos
FACT telescope, same type, same place

Bockermann, Christian and Brügge, Kai and Buss, Jens and Egorov, Alexey and Morik,Katharina and Rhode, Wolfgang and Ruhe, Tim
“Online Analysis of High-Volume Data Streams in Astroparticle Physics”
Best Paper Award ECML PKDD 2015