CERN

Modelling of Complex Systems with Data-Driven Models

Christoph Obermair

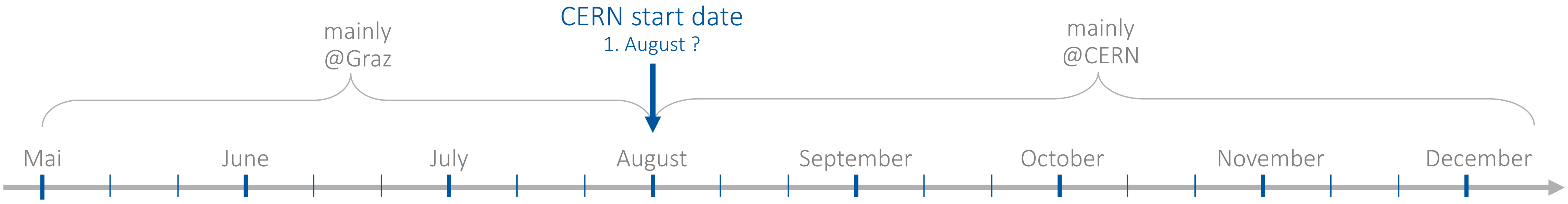
Presentation Outline

1. Motivation
2. Overview of data-driven models
3. Data-driven models in the PE section
4. Existing data-driven models
5. Conclusion

1. Motivation

Motivation for this presentations: trigger discussions, set expectations, gather ideas, extend scope of view

My preliminary timeline:



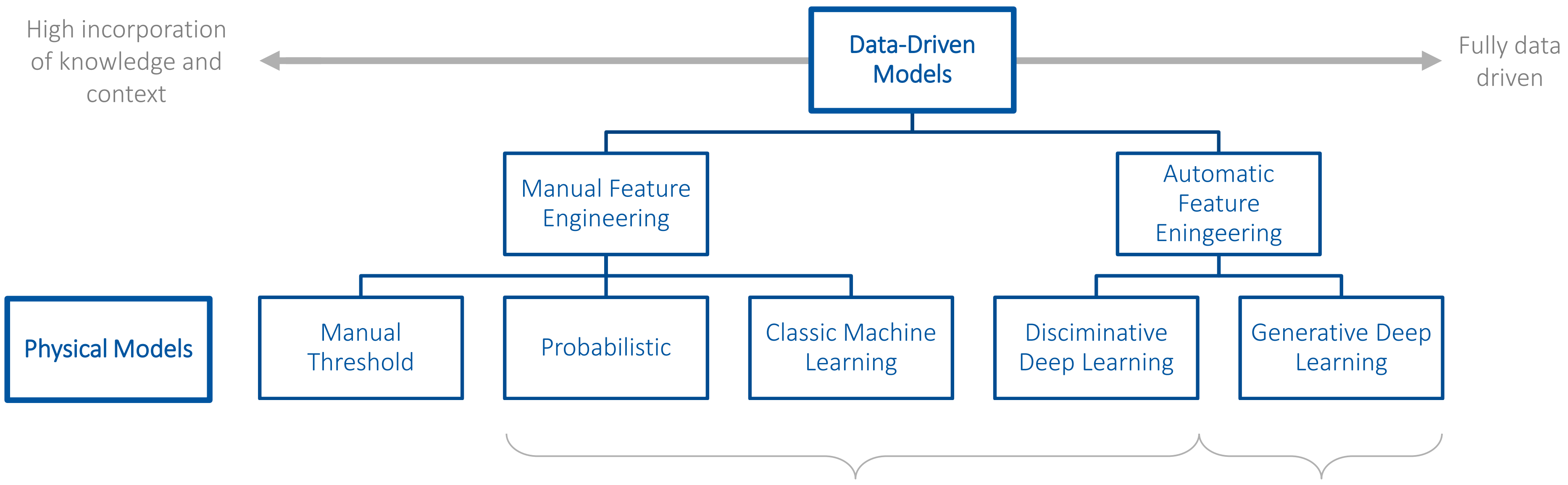
Initial Research Graz:

- General research
- Explore multivariate time series classification algorithms with class imbalance and data limitation on synthetic dataset
- Explore existing data analysis projects in PE section
- Generate catalog of datasets

CERN:

- Explore further data driven models
- Build ensemble of suitable algorithms
- Discuss further approach

2. Overview of Data-Driven Models



High incorporation of knowledge and context

Fully data driven

Data-Driven Models

Manual Feature Engineering

Automatic Feature Eningeering

Physical Models

Manual Threshold

Probabilistic

Classic Machine Learning

Discriminative Deep Learning

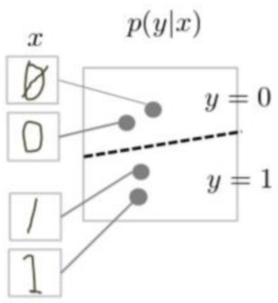
Generative Deep Learning

Ensemble Methods:

- Combination of different models (i.e. hybrid models [1])
- Wisdom of the crowd: Each classifier will contribute to a better output (as long as it is better than guessing)

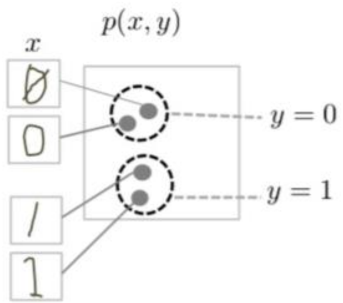
Discriminative:

- Output: Conditional probability: p(Output | Input)
- Threshold is adjusted based on model



Generative:

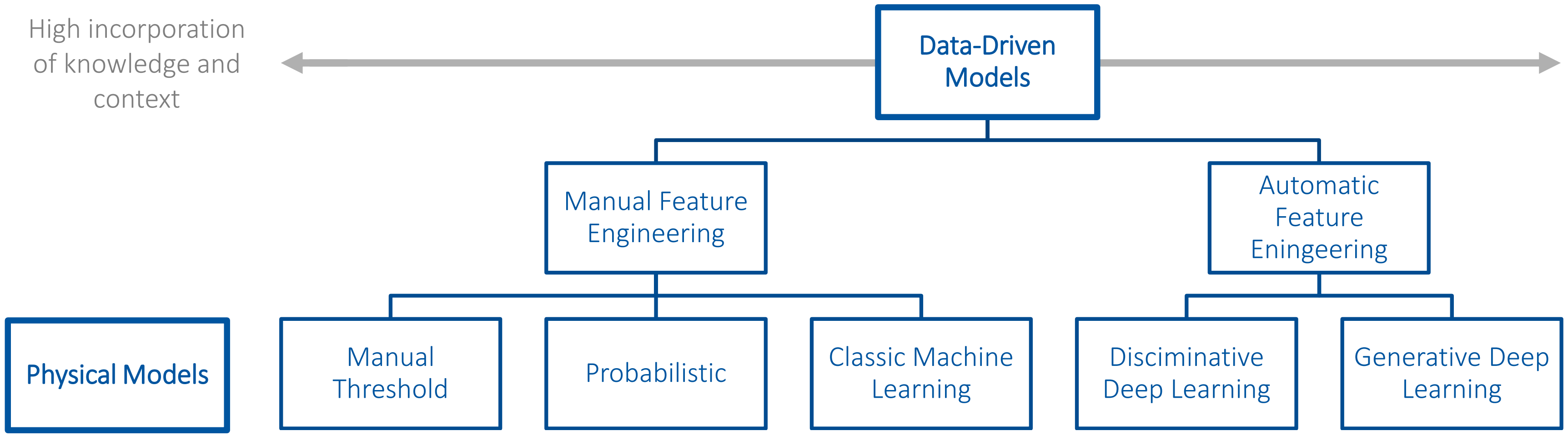
- Output: Joint probability p(Output, Input)
- Clustering as a Prestep



[1] Christoph Obermair. “ Extension of Signal Monitoring Applications with Machine Learning”. TU Graz, 2020

Content from M. Maciejewski and H. Fawaz et al., „Deep learning for time series classification: a review“, 2019

2. Overview of Data-Driven Models



High incorporation of knowledge and context

Fully data driven

Data-Driven Models

Manual Feature Engineering

Automatic Feature Eningeering

Physical Models

Manual Threshold

Probabilistic

Classic Machine Learning

Discriminative Deep Learning

Generative Deep Learning

Approach 1:

Find novel result for given use case

1. Study use case
2. Develop new model for use case (go left to right)
3. Compare

Requirement:

Deep understanding of given use case

Approach 2:

Find novel model, suitable to many use cases

1. Study existing models
2. Develop new model
3. Compare new model with existing model

Requirement:

Data sets to measure performance e.g.: $D = \{(x_1, y_1), \ldots, (x_N, y_N)\}$

Content from M. Maciejewski and H. Fawaz et al., „Deep learning for time series classification: a review“, 2019

Approach 1:
Find novel result for given use case

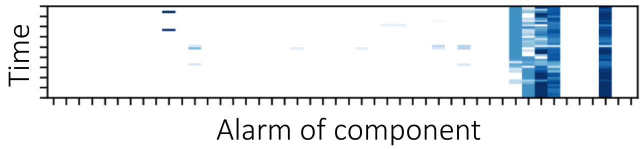
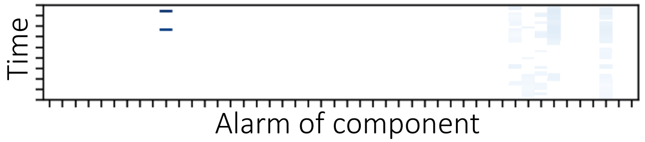
3. Data-driven Models in the PE Section

a. Failure Mechanism Analysis
b. Quench Heater Analysis
c. UFO Dynamic Studies
d. BLM Data Analysis
e. EE-Switches
f. Busbar Modelling
g. AvailSim4

a. Failure Mechanism Analysis

- Input data: List of alarms
  - Main feature: Alarm priority ∈ $[0,1,2,3]$ of failing component
  - Priority defined by FAULT_FAMILY/_MEMBER_/CODE
- Current Goal:
  - Predict priority 3, given past priorities
- Used data:
  - 8 Power converter signals, 8 Beam destinations signals, 27 Interlock signals
  - Data size: per component about 5-15 priority 3 events, much more alarms with lower priority
- Current Approach:
  - Use discriminative deep learning (classic machine learning as reference)
- Difficulties:
  - No detection of cascade failures due to logging frequency (all alarms occur at same timestamp)
- Future Goals:
  - Use additional component signals
  - Find architectural dependencies (which component is the root cause?)

Overview approach

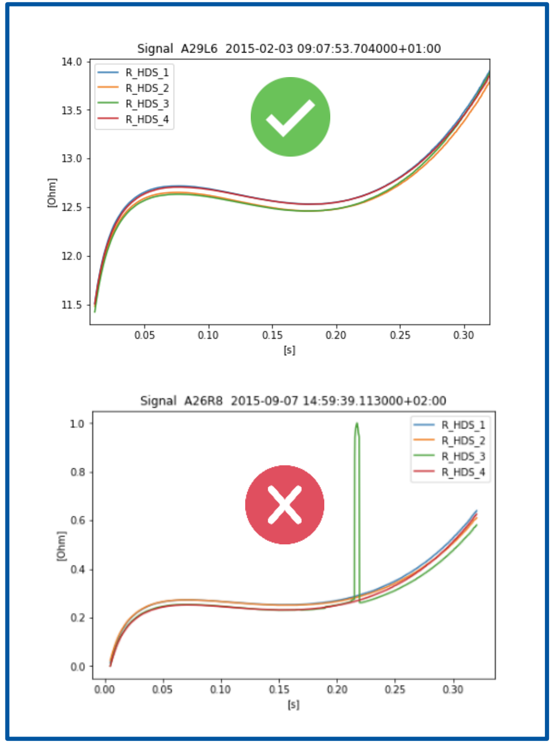
Input	
Prediction	There will be a priority 3 alarm in the future
Reason / Precursor	

L. Felsberger et al., ”Analyzing Failure Mechanism in Complex Infrastructures”, 2020

b. Quench Heater analysis

- Input data: Electrical signals
  - Voltage, current discharges of QH
  - Extracted features: resistance, min, max, characteristic time etc.
- Current Goal: Classify discharge, given past discharges
- Used data:
  - 1232 dipole magnets with 4 QH each
  - Data size: 1-10 discharges per QH, ~16000 data points per discharge
- Current Approach:
  - Use classic machine learning (threshold based classification as reference)
- Difficulties:
  - Classification is as good as input features
  - High additional complexity, low additional benefit
- Future Goals:
  - Failure prediction (similar to Failure Mechanism Analysis)

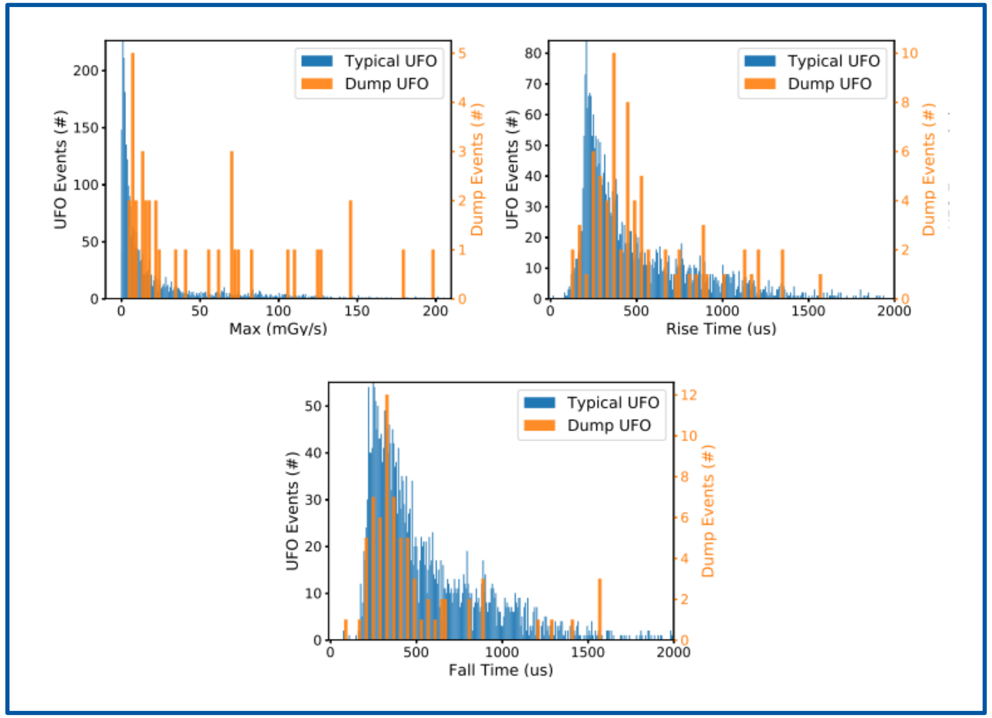
Resistance during QH discharge



c. UFO Dynamic Studies

- Input data: BLM data (triggered by UFO Busters)
- Current Goal:
  - Classify and understand UFOs with physical models
- Used data:
  - 300 000 triggers during Run 2
  - of which 3000 have more than 5 sample points
  - of which 100 are during beam dump (manually analyzed, aborted signal)
- Current Approach:
  - Physical models, Probabilistic
- Difficulties:
  - Further data-driven models could provide additional information
- Future Goals:
  - Cluster UFO signals to detect possible new classes
  - Extract patterns of dump UFOs

UFO histograms



P. Belanger “Update on Ufo dynamics studies“, CERN, 2019

d. BLM Data Analysis

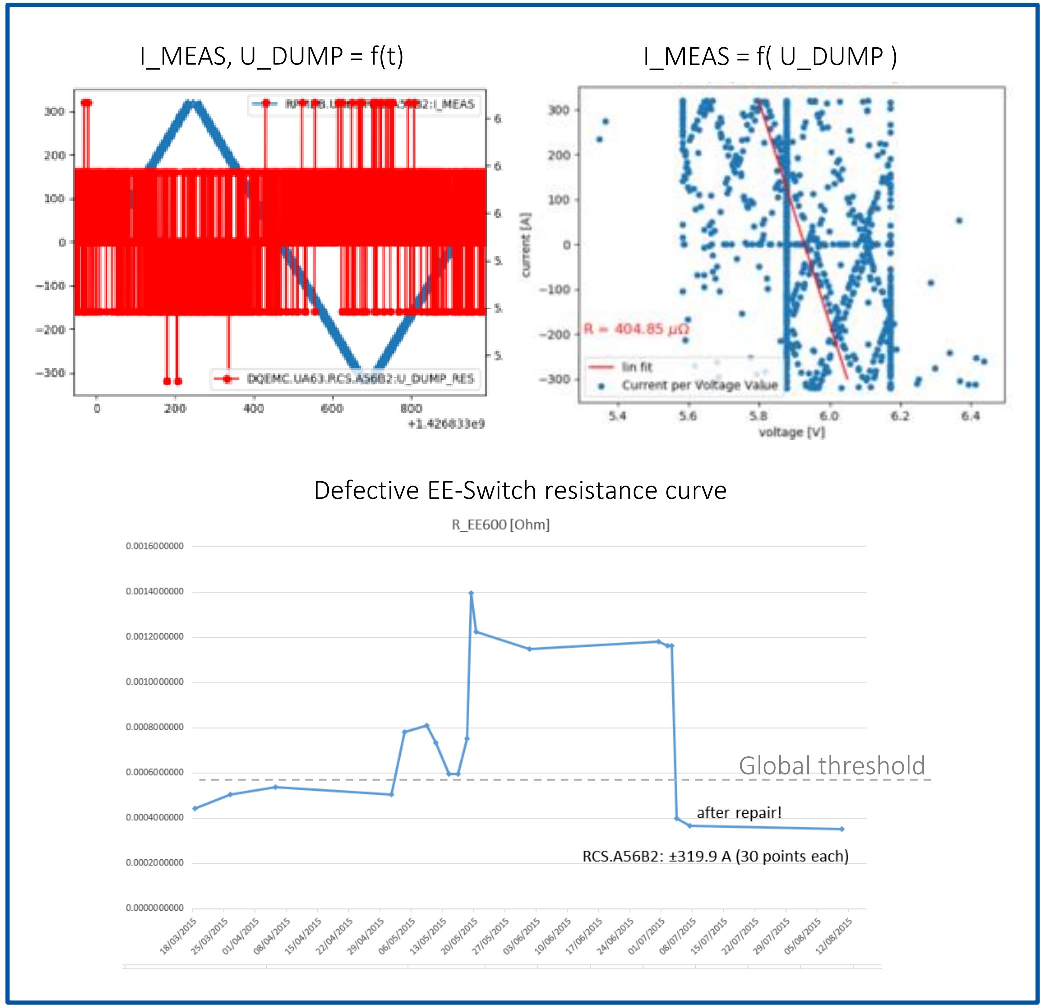
- Input data: BLM data, context data (e.g beam energy, beam intensity)
- Current Goal:
  - Find patterns and correlations with data driven threshold models (physical model as reference)
- Used data:
  - ~ 1500-6000 “high energy” dumps in Run 2
  - ~ 4000 BLMs
  - ~ 96000 Data points per BLM per Dump (different resolutions e.g 1s with 40µs sampling, 501s with 1,3s sampling)
- Current Approach:
  - Calculate simple statistical properties of BLM data (e.g. min, max, std, kurtosis) compare to
- Difficulties:
  - Context data changes frequently
- Future Goals:
  - Find patters and correlations with machine learning (physical and threshold model as reference)

Y. Nie et al., “Concept of Beam-Related Machine Protection for the Future Circular Collider”, IPAC,2019

e. EE -Switches

- Input data: Electrical signals
  - Voltage, current discharges of ramp cycle
  - Extracted features: resistance
- Current Goal:
  - Classify ramp cycle, given past ramp cycles
- Used data:
  - 752 ramp-cicles
  - 202 EE-Switches
- Current Approach:
  - threshold based classification (global threshold: healthy < 550μΩ)
- Difficulies:
  - Resistance calculation is prone due to low resolution of voltage measurement
- Future Goals:
  - Improve resistance calculation
  - Cluster ramp cycles to find individual threshold for each EE-Switch

Electrical signals EE-switches



I_MEAS, U_DUMP = f(t)

I_MEAS = f(U_DUMP)

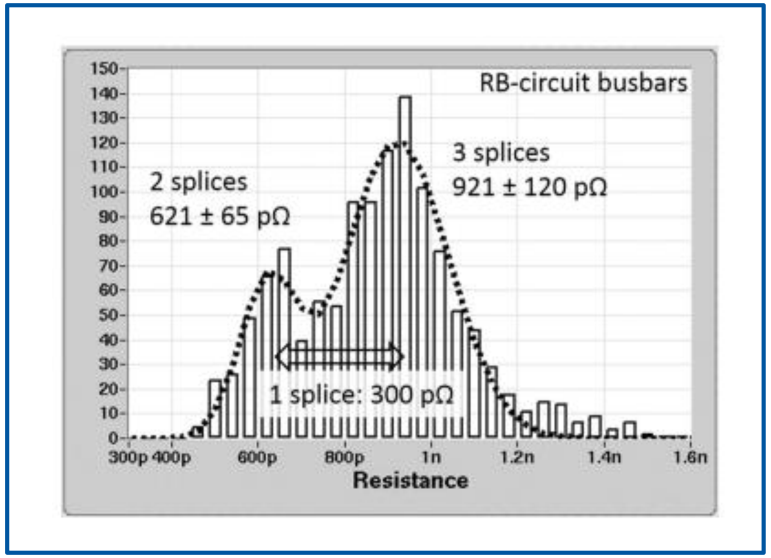
Defective EE-Switch resistance curve

A. Muller, “Machine Learning on Data of 600A Energy Extraction Switches“

f. Busbar Modelling

- Input data: Electrical signals
  - Voltage, current during ramp ups
  - Extracted features: resistance, SNR
- Current Goal:
  - Probabilistic analysis of BB
- Used data:
  - ~2500 ramp ups
  - 8 circuits with 154 dipole busbars
- Current Approach:
  - Probabilistic anomalie detection
- Difficulies:
  - Probabilistic approaches are not capable to detect patterns
- Future Goals:
  - Find patterns in BB resistance growth across different circuits

Histogram of splice RB splice resistance



Z. Charifoulline et al., “Resistance of Splices in the LHC Main Superconducting Magnet Circuits at 1.9 K“, 2018

g. AvailSim4

- Input data: List of failure modes, architectural dependencies
- Current Goal:
  - Investigate fault effects
- Used data:
  - 100 – 10000 Signals
  - Fault + component + impact on other components
- Current Approach:
  - Fault tree analysis based on Monte Carlo simulations
- Difficulies:
  - Model complexity, computational cost
- Future Goals:
  - Furhter explore data-driven models (i.e. surrogate models)

Niemi A. et al. “Availability modelling approach for future circular colliders based on the LHC operation experience”, CERN, 2016

Data-Driven Models in the PE Section

→ repetitive goals, data types (time series (TS)), and difficulties

Project	Failure Mechanism Analysis	QH Discharges	EE Switch	UFO	BLM	AvailSim4	Busbar Modelling
Contact	L. Felsberger T. Cartier-Michaud A. Apollonio A. Muller B. Todd	C. Obermair M. Maciejewski Z. Charifoulline	A. Muller M. Maciejewski T. Cartier-Michaud L. Felsberger	D. Wollman C. Wiesner	C. Wiesner T. Cartier-Michaud A. Apollonio	T. Cartier-Michaud A. Apollonio	M. Maciejewski Z. Charifoulline
Input data	List of alarms (TS)	Electrical signals (TS)	Electrical signals (TS)	BLM signals (TS)	BLM signals (TS)	List of failure modes(TS)	Electrical signals (TS)
Current goal	Prediction with discriminative deep learning	Classification with classic ML	Classification with manual threshold	Find patterns and corelations with physical models	Find patterns and corelations with manual threshold	Fault tree analysis based on Monte Carlo simulations	Probabilistic analysis of BB resistance
Data size	8 PC Signals 8 Beam destinations 27 Interlock	$u, i \in \mathbb{R}^{N\times C\times D}$ N … samples (~16000) C … circuits (4) D … events (3246)	$u, i \in \mathbb{R}^{N\times C\times D}$ N … samples C … circuits (202) D … events (752)	$BL \in \mathbb{R}^{N\times C\times D}$ N … samples (>5) C … BLMs (~4000)? D … events (~3000)	$BL \in \mathbb{R}^{N\times C\times D}$ N … samples (96k) C … BLMs (~4000) D … events (~6000)	100 – 10000 Signals: fault + component + impact on other components	$u, i \in \mathbb{R}^{N\times C\times D}$ N … samples C … RB BB (8*154)? D … events (~2500)
Difficulties	Data Inconsistency Data Quality Data Limitation	Data Inconsistency Data Quality Data Limitation	Data Inconsistency Data Quality Data Limitation	Data Inconsistency Data Quality Data Limitation	Data Inconsistency Data Quality	Model Complexity	Data Inconsistency Data Quality
Future goal	Forecasting, find conditional dependencies	Forecasting	Anomalie detection, clustering	Clustering, data augmentation	Find patterns and corelations with ML	Investigate fault effects with ML	Learn sequence of BB resistances with ML

Table contributed by T. Cartier-Michaud, extentended by C. Obermair

Existing Difficulties

→ repetitive goals, data types (time series (TS)), and difficulties

→ LHC Signal Monitoring Project offers solution to many difficulties → perfect environment for data analysis

Difficulties	Explanation	Solution
a.) Data Acquisition	query data	LHC-SM API (for PM, NXCALS)
b.) Data Inconsistency	e.g. through frequent maintenance actions	explore context data (e.g. from NXCALS)
c.) Data Quality	e.g. no failure precoursors, data is representation of consequence not root cause	explore additional data from different data bases (e.g. NXCALS in addition to LASER)
d.) ML Algorithms	find best algorithm to archive goal	build ensemble of suitable algorithms for PE section
e.) Data Limitation	not enough data available	combination with physical models, transfer learning

Approach 2:
Find novel model, suitable to many use cases

Approach 2:
Find novel model, suitable to many use cases

4. Existing Data-Driven Models

a. Data-Driven Models for Data Limitation

b. Overview of Existing Meta Learning Algorithms

H. Fawaz et al., „Deep learning for time series classification: a review“, 2019

a. Data-Driven Models for Data Limitation

- A child can generalize the concept of a giraffe from a single picture in a book
- Machine learning algorithms typically needs a lot of data to learn from

  → Astonishing process in recent years

Approaches:

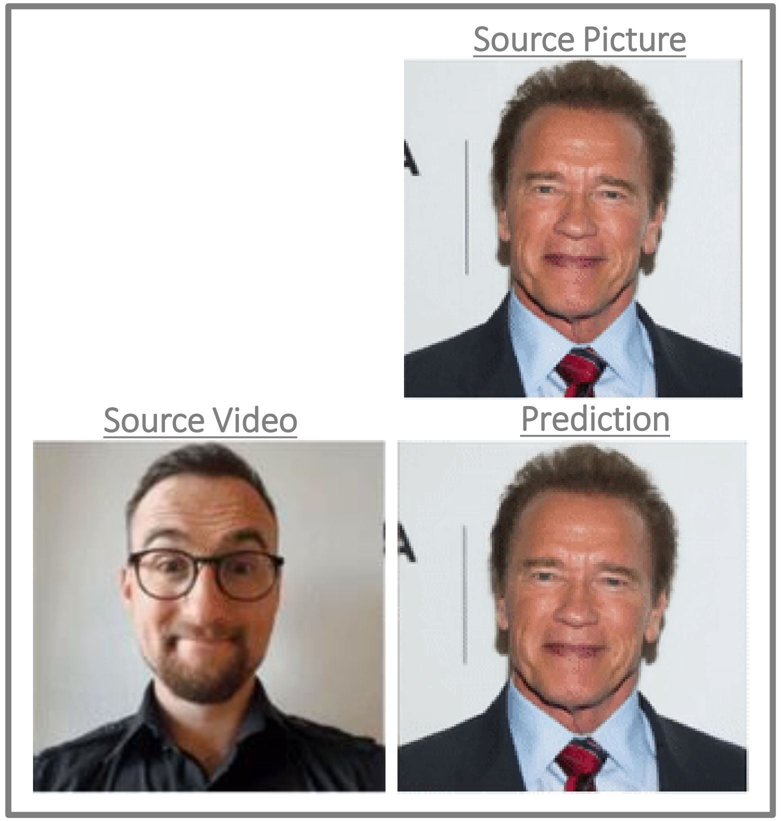
- Combination of existing models with data-driven models
- Transfer Learning: e.g. meta learning for few-shot classification:

  → gain knowledge while solving a task

  → apply knowledge to related task
- Class Imbalance: data limitation in on class (e.g. lack of failure cases)

  → re-sampling

  → re-weighting



A. Siarohin et al., “First Order Motion Model for Image Animation”, NIPS, 2019

b. Overview of Existing Meta Learning Algorithms

i. Metric based meta learning

→ Learn a metric to distinguishing between different classes

ii. Optimization and initialization based meta learning

→ Improve weight optimization or initialization during training.

iii. Model based meta learning

→ Improve internal structure of model

iv. Hallucination based meta learning

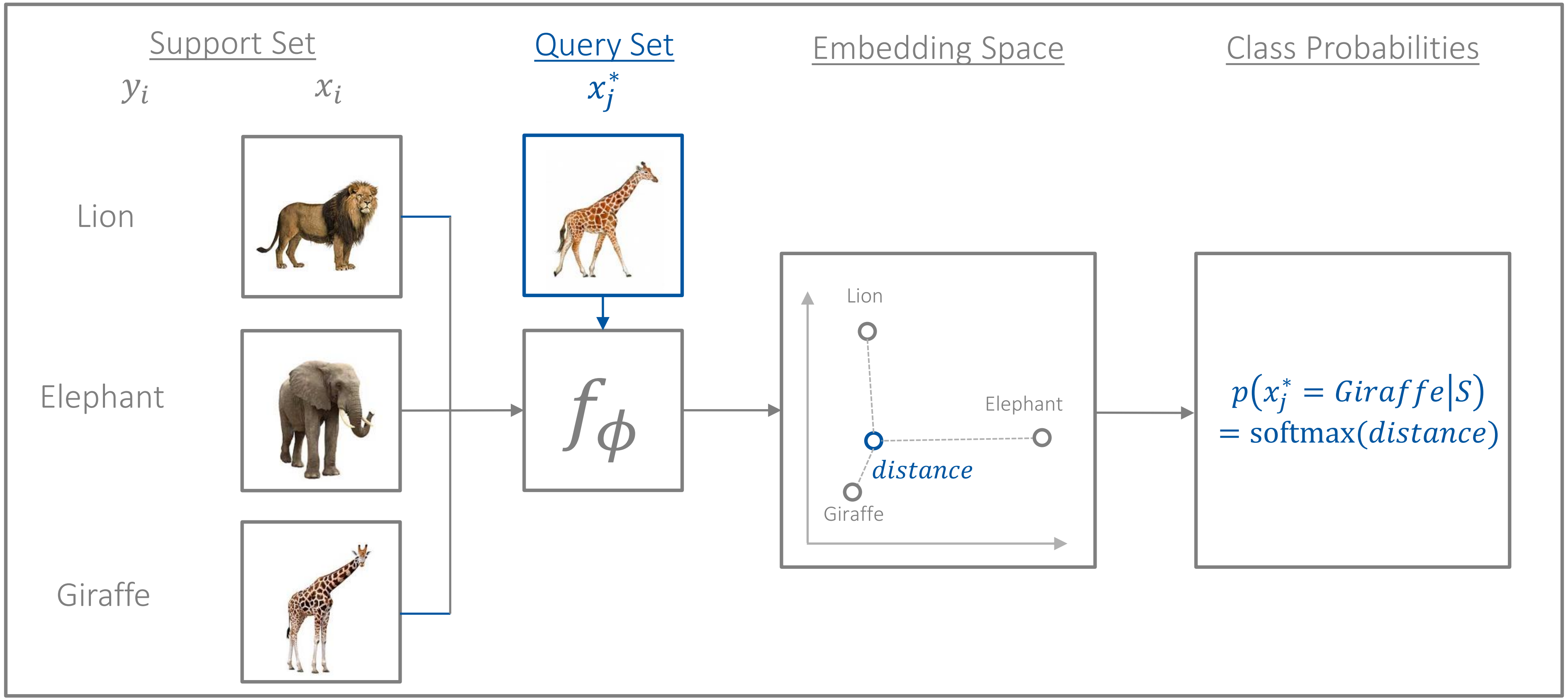
→ Generate additional data

Notation few shot learning:

$S = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{N}$ … Support Set for training

$Q = \{(\boldsymbol{x}_j^*, y_j^*)\}_{j=1}^{T}$ … Query Set for testing

i. Metric Based Meta Learning



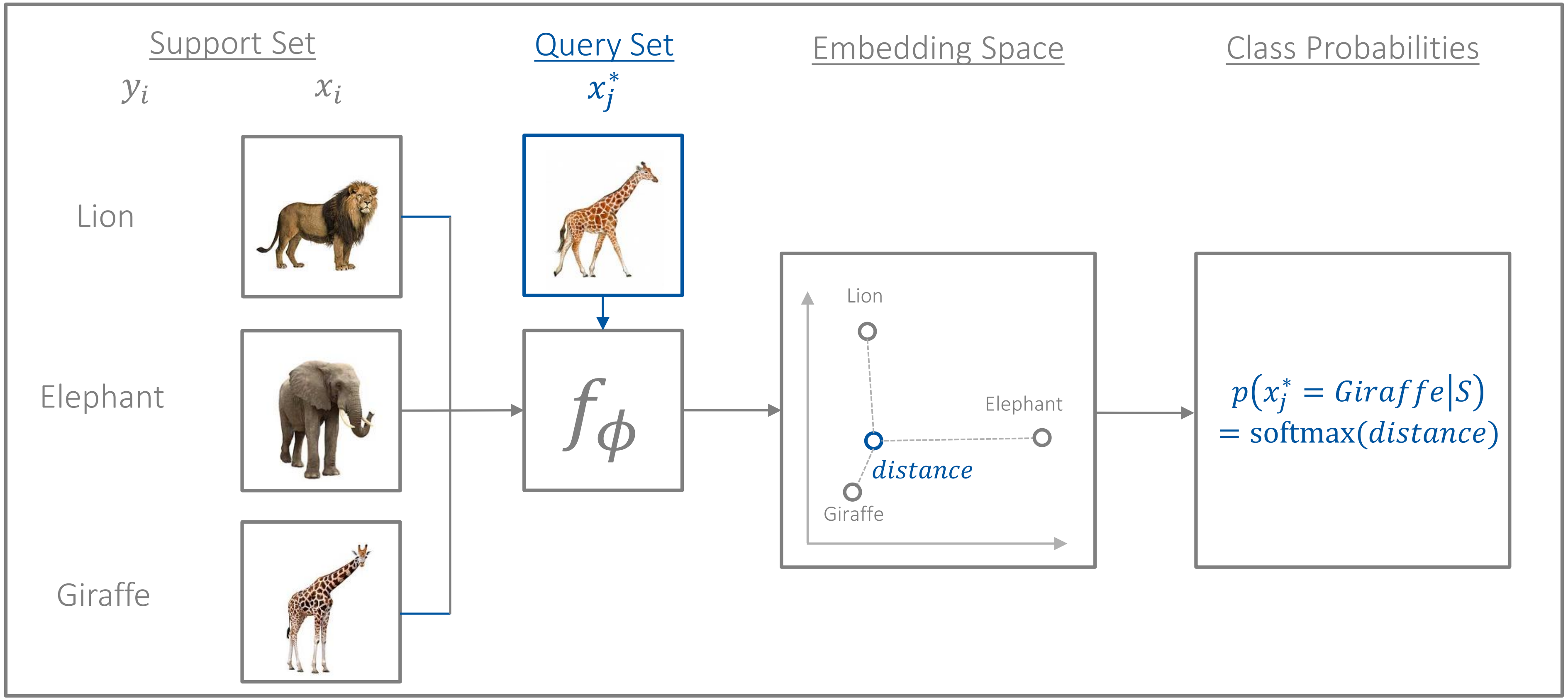
a. Generate an embedding with a model $f_\phi\colon \mathbb{R}^D \rightarrow \mathbb{R}^M$ (i.e. automated feature extraction)

b. Learn a metric to distinguishing between different classes in embedding space

c. Optimize model parameter ϕ with gradient decent: $\mathrm{L}(\phi) = -\log(p(x_j^* = Giraffe|S))$

J. Snell et al.,"Prototypical Networks for Few-shot Learning", NIPS, 2017

i. Metric Based Meta Learning



Metric based meta learning for time series?

→ Possible field of research

J. Snell et al.,''Prototypical Networks for Few-shot Learning'', NIPS, 2017

J. Y. Franceschi et al., „Unsupervised Scalable Representation Learning for Multivariate Time Series“, NIPS, 2019

5. Conclusion

Approach 1:
Find novel result for given use case

- The goal is to analyze data (not necessarily with machine learning)
- Many existing use cases in PE section
- LHC Signal Monitoring Project offers great data analysis environment

Approach 2:
Find novel model, suitable to many use cases

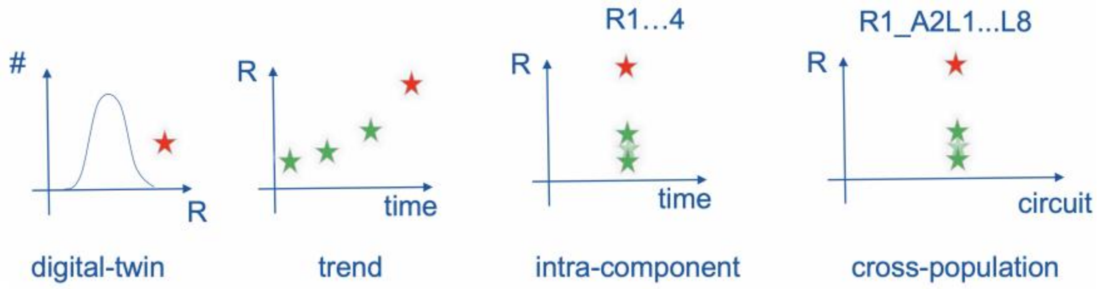
- There already exist many models suitable for difficulties of PE section
- In order to use them: Catalog of data sets necessary
- Possible research direction: Metric based meta learning for time series

Notes

CERN

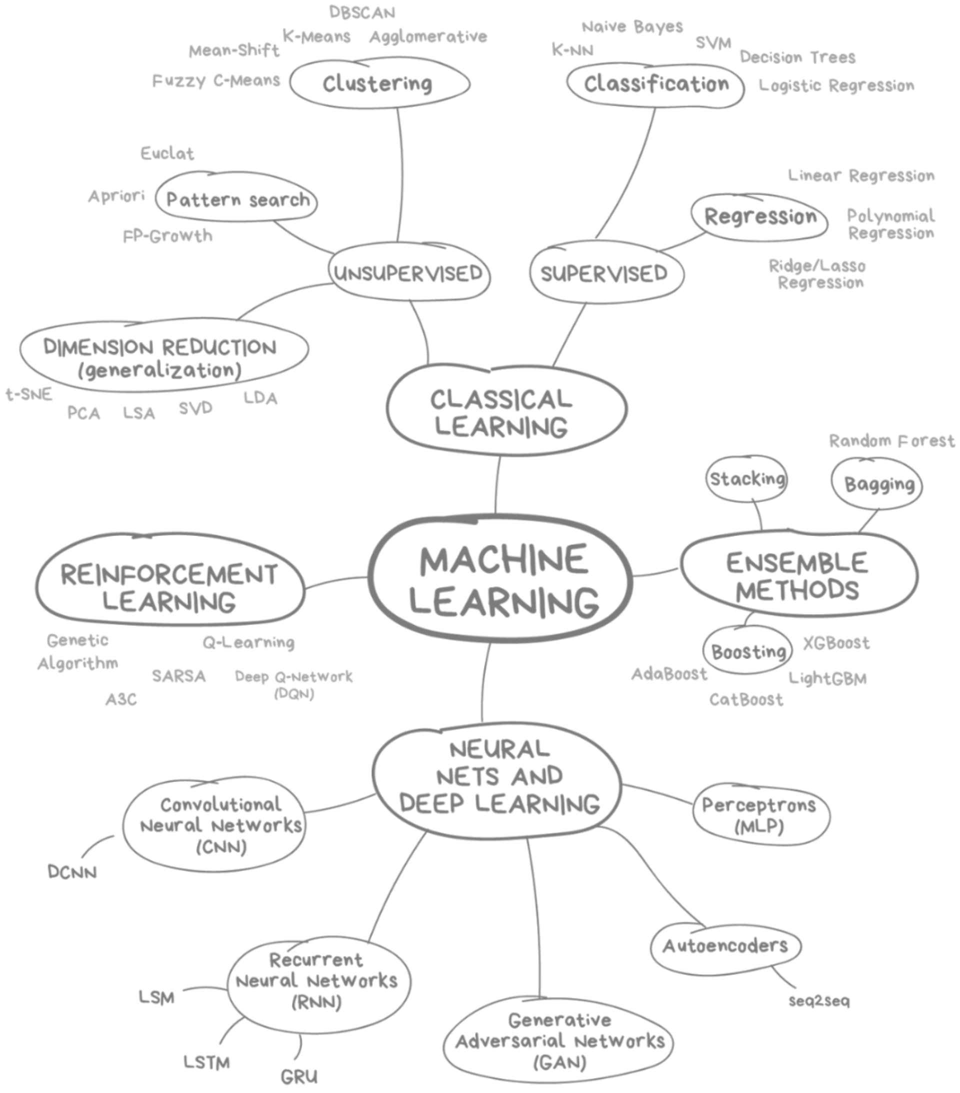
Existing ML Algorithms

- Lots of existing open source library's
  - Gather best algorithms suited for PE projects
- Initially with standard datasets (e.g. MNIST –easy to understand), supplemented with CERN datasets
- Structure:
  1. Input: e.g. Time series
  2. Goal: e.g. Classification
  3. Algorithm: e.g. RNN, DCNN, SVM
  4. Extension: e.g. Ensemble, Meta learning
  5. Testing: e.g. Cross-Validation



digital-twin trend intra-component cross-population

contributed by M. Maciejewski



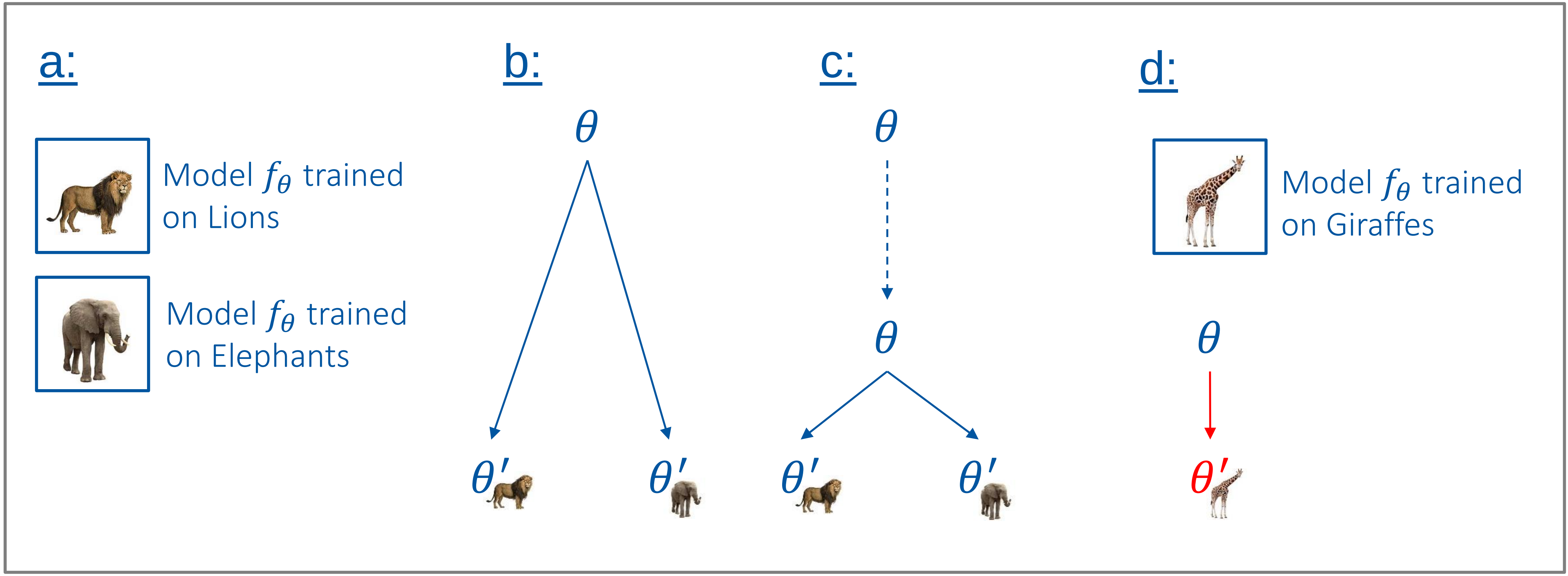
ii. Optimization and initialization based meta learning

a. A model f_θ with the same initial values θ is trained on different tasks T_i

b. Initial values are optimized with gradient decent: $\theta_i' = \theta - \alpha \nabla_\theta L(f_\theta)$

c. Initial values θ from a.) are adjusted: $\theta \leftarrow \theta_i' - \beta \nabla_\theta \sum L(f_{\theta_i})$

d. Fast convergence of new task



a:

Model f_θ trained on Lions

Model f_θ trained on Elephants

Model f_θ trained on Giraffes

C. Finn, P. Abbeel, and S. Levine, “Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks,” 2017

iii. Model based meta learning

Networks with memory capacity (e.g. RNN with Long Short-Term Memory):

→ Quickly adapts to new information without forgetting main information

→ Memory-Augmented Neural Networks use this behavior for meta learning

→ Adaptation to never bevor seen classes in a few shots.

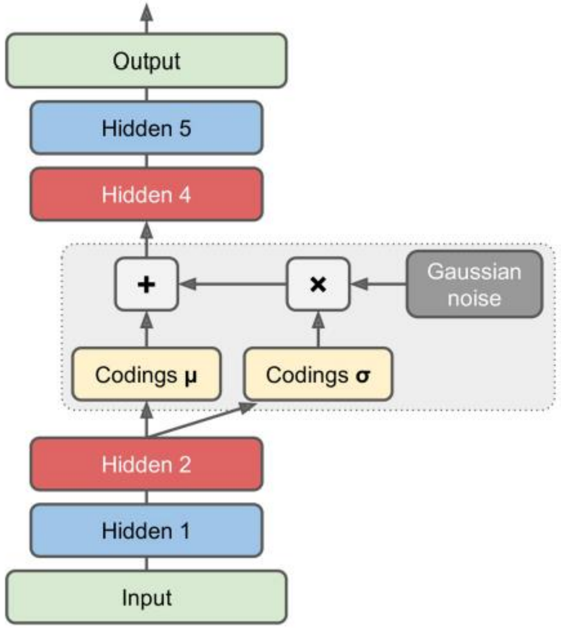
A. Santoro, et al.“One-shot Learning with Memory-Augmented Neural Networks,” May 2016
S. Hochreiter and J. Schmidhuber, Long Short-Term Memory. 1997.

iv. Hallucination based meta learning

The goal is to train a data generator which is able produce artificial data.

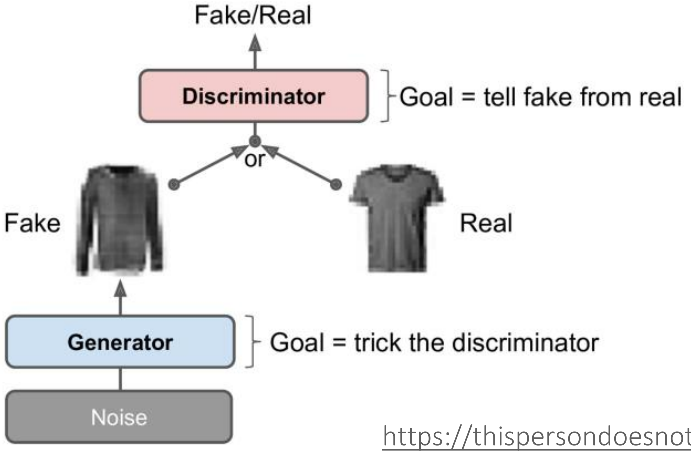
Variational Autoencoders

1. Reduce Dimension
2. Add noise
3. Increase Dimension



Generative Adversarial Networks (GANs)

1. Generator produces artificial data
2. Discriminator tries to classify between real or fake data



https://thispersondoesnotexist.com/

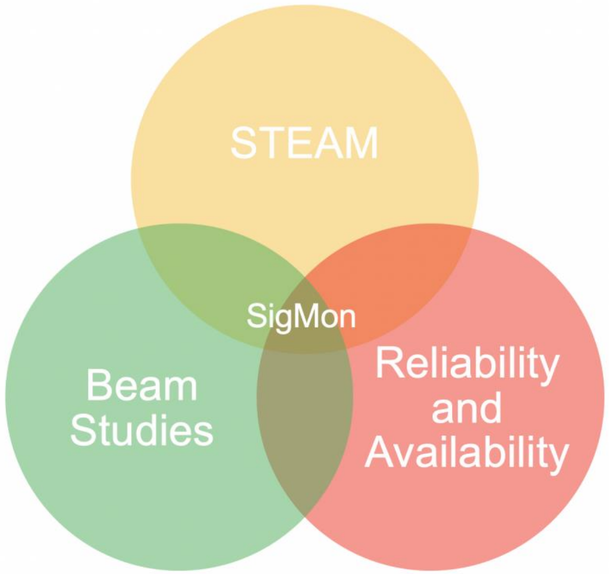
D. P. Kingma and M. Welling, “Auto-Encoding Variational Bayes,” May 2014

Géron, Aurélien,"Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow”, 2019

I. Goodfellow, et. al., “Generative Adversarial Nets,” 2014

Solutions to Data Acquisition

- LHC Signal Monitoring Project:
  - Query Data from PM, NXCALS, (CALS)
  - General way to query data (pyeDSL)
  - Execution pipeline
  - Good documentation
  - Library for:
    - Preprocessing algorithms
    - Feature engineering algorithms
  - ...


STEAM
SigMon
Beam Studies
Reliability and Availability