Taming the Data Divide to Enable AI-Driven Networks

Alessandro Finamore
Principal engineer, AI4NET Datacom lab
Huawei Technologies Research Center Paris
alessandro.finamore@huawei.com
africanio
The talk in 1 slide

1. “Our” challenges surrounding data and replicability
2. How other communities face those challenges

I have a lot of Qs ... but not many As
Part I
Introduction

Part II
Data-sharing & reproducibility

Part III
Easing the data-dependency
Part I
Introduction

Part II
Data-sharing & reproducibility

Part III
Easing the data-dependency
Rewinding time

IC0703 - Data Traffic Monitoring and Analysis: theory, techniques, tools and applications for the future networks (2008 – 2012)

Sep, 2008 – TMA meeting @ Samos

My “first contact” with the European traffic measurements research community
...since Samos’08

Cooperation with Italian ISP on traffic monitoring

Malware traffic analysis

Country-scale mobile network analytics

Customer satisfaction modeling via BigData and Machine learning

AI technologies applied to traffic monitoring

Telcos data as the common denominator
Huawei R&D in a nutshell

- 180,000 Employees
- 80,000 R&D employees
- 170+ Countries
- 15+ Oversea R&D Centers
- No. 70 Interbrand's Top100 Global Brands
- No. 83 Fortune Global 500

~15% of revenue re-invested in R&D
~30/70% long/short-term split
~2000 researchers in European Research Institute (ERI)
~150 researchers in Paris RC
~30 network researchers (aka DataCom Lab)
AI-assisted networking @ HUAWEI

evolved cloud-native 3-tier architecture

NAIE
Offline training
Training, data aggregation, and model generalization

iMaster NAIE
Cloud platform OSS/Third-party app

iMaster NCE
Network-wide analysis, inference & closed-loop optimization

Controller & Analyzer

NCE/MAE

Controller

Analyzer

Cloud platform

OSS/Third-party app

Controller

Model-driven telemetry

NETCONF/YANG

Cross-vendor southbound API

Huawei devices

Vendor B’s devices

Air/Net
Cloud/HiSec Engines

Engines

Measurement, edge inference & real-time decision-making

Schedulers
AI-aware job schedulers

Training:
Federated Learning, Distributed training

General:
Multi-vendor graph/models, Transfer learn

Specific:
Deep Models Quantization & Distillation

Control:
large-scale, data-driven, explainable deep RL

Deploy:
Cloud vs edge vs fog vs mist ...

XAI O&M:
Unsupervised Fault detection, Semi-supervised repair

Real-time:
inference & control

Model self-awareness, Self-supervised learning

Continuous learning, Few-shot learning

Offshore global

Online global

Realtime local

Offline global

Opportunities & challenges

CPUs
GPUs
TPUs
Ascend310
Ascend910

ARM

Ascend310

TPU

AI4Net

AR

TNSM’22 : Landing AI on Networks: An equipment vendor viewpoint on Autonomous Driving Networks

**Roadmap to the Edge Intelligence**

- **5G Edge**
  - First commercial 5G MEC Deployments

- **Edge AI**
  - AI features brought to each edge node (useful to learn and to share models with other edge nodes).

- **Distributed AI**
  - AI algorithms distributed in a network of edge devices, providing scalability and reliability.

- **Secure and private**
  - Secure edge systems that ensure user privacy and keep information secure.

- **Nanophotonic technologies**
  - Nanophotonic fabrics will perform complex matrix operations.

- **6G**
  - First deployments of a new generation of Edge AI

- **Pre-trained Edge**
  - Pre-trained AI models used for processing data at the edge

- **Dedicated Hardware**
  - Specialized edge devices capable of performing AI computation

- **Learning-driven Communication**
  - Complex wireless communication systems managed by edge intelligence

- **Real-time training**
  - New distributed algorithms that make it possible to build models almost in real-time

**ARTIFICIAL INTELLIGENCE CHIPSETS MARKET**

- **$8.14 Billion**
  - 2019

- **$108.85 Billion**
  - 2027

- **CAGR 38.9%**

**ARTIFICIAL INTELLIGENCE [AI] IN BUSINESS**

**MARKET TRENDS**

- Popularity of Artificial Intelligence & Machine Learning
- Rising Adoption of the Cloud-based Solutions
- Emergence of Quantum Computing

**AI TRENDS IN VARIOUS SECTORS**

- [Link](https://www.fortunebusinessinsights.com/infographics/artificial-intelligence-ai-chipsets-market-104500)
- [Link](https://www.hackerearth.com/blog/developers/applications-of-artificial-intelligence/)
...but are we ready for such AI data-driven networks?

https://xkcd.com/1838/
People smarter than me say...

The network will be programmed by many, operated by a few

Nick McKeown – Stanford University

With a panel of graduate student discussants from around the world.

« Machine learning is very good at understanding and predicting the behavior of systems we do not understand [...] but networking is mostly about implementing something according to a “model” we already know »

My take

AI is an opportunity BUT if/what we need to and where/how to integrate AI in networks is still largely an ongoing debate
Standards are in their infancy
AI and networks: a multi-faceted relationship

- Computational budget
- Resources Orchestration
- Cloud-vs-edge displacement
- Intent & Configuration
- Explainability
- Meta-learning & Generalization
- Data sharing
- Low latency
Computational budget
Resources Orchestration
Cloud-vs-edge displacement
Computational budget
Intent & Configuration
Computer network
Explainability
Low latency
Data sharing
Meta-learning & Generalization
AI and networks: a multi-faced relationship
Part I
Introduction

Part II
Data-sharing & reproducibility

Part III
Easing the data-dependency
Part I

Introduction

Part II

Data-sharing & reproducibility

In a broad sense, but in particular fine-grained logs across planes/layers

Part III

Easing the data-dependency
“We” know how to build platforms
The top-2 (known) problems in our community

Public data-access & Measurements longevity

...and AI is just renewing this (old) divide

https://xkcd.com/2582/
The increasing data divide

An Effort to Democratize Networking Research in the Era of AI/ML

- Arpitha Gupta
  - UCSB
  - Santa Barbara, CA

- Chris Mac-Keiser
  - NKU
  - Newtown, KY

- Walter Willinger
  - NKU
  - Newtown, KY

[... publicly available network measurements in support of network automation tasks are rare, not necessarily representative, often a by-product of some other measurement activities [...]

Recommendation: treat university campus networks are real production environments

Data-driven Networking Research: models for academic collaboration with industry (a Google point of view)

- Jeffrey C. Mogul
  - UC Berkeley
  - Berkeley, CA

- Christopher Dietz
  - Google

- John Wilkes
  - Philips Gil

- Annet Veelenturf
  - Google

[... We encourage academic researchers to focus less on “can we obtain network-related data from Google?” and more on “how can we do more collaborative, data-driven networking research with Google?” [...]

Recommendation: setup ad-hoc collaboration between Companies (Google) and researchers

Workshop on Overcoming Measurement Barriers to Internet Research (WOMBIR 2021) Final Report

- kc claffy
  - claffy@shop.org

- David Clark
  - MIT
  - dclark@mit.edu

- Mattija Juker
  - University of Dusseldorf
  - juker@uni-due.de

- Ellen Zegura
  - Georgia Tech
  - ezegura@is.gatech.edu

- Fabio E. Bustamante
  - University of Miami
  - fabio@cs.miami.edu

[... data sets require longitudinal (long-term, ongoing) data collection and sharing, support for which is more challenging for Internet research than other fields [...]

But often an employee of the associated company is an author on the paper, which triggers concerns regarding scientific objectivity [...]

Recommendations:

- New model for cross-sectors collaborations
- Fund longitudinal measurements platforms
- Annual state-of-the-Internet report/conference
- Data sharing code of conduct

[Note: Data is made available in curated repositories, or otherwise provided in ways that allow adequate access for legitimate scientific research.

- Access requires registration with data owner and legitimate research use.
- Standard anonymization methods are used where needed.
- Recipients agree to not report names.
- Recipients agree that they will not de-anonymize data.
- Recipients can publish analysis and data examples necessary to review research.
- Recipients agree to use accepted protocols when handling sensitive data, such as security vulnerabilities or data on human subjects.
- Recipients agree to cite the repository and provide publications back to repository.
- Repositories can create certified products developed by researchers.

Table X: Codes of conduct have been developed that enable responsible sharing of data in ways that protect stakeholders while allowing research [20, 22].}
Deeper roots than the mere lack of incentives

1. Hard constraints (privacy/business concerns)
2. Soft constraints (scale, operational risk, staff time)
3. Anonymization is not a panacea
4. Data/code cleansing for public release is time-consuming
...no, but wait a sec

① We do have **Best Dataset awards** @ TMA, IMC, PAM

Yes, and that’s **Awesome**

...but very **opportunistic and tied to specific studies**

② We also have **badges**

Yes, so let’s talk about those
Are we improving at replicability? 

**Functional**
The artifacts associated with the research are found to be documented, consistent, complete, exercisable, and include appropriate evidence of verification and validation.

**Reusable**
The artifacts associated with the paper are of a quality that significantly exceeds minimal functionality.

**Available**
Author-created artifacts relevant to this paper have been placed on a publically accessible archival repository. [...] they need not be complete in the sense described above.

**Reproduced**
Results of the paper have been obtained in a subsequent study by a person or team other than the authors, using, in part, artifacts provided by the author.

**Replicated**
Results of the paper have been independently obtained in a subsequent study by a person or team other than the authors, without the use of author-supplied artifacts.

---

**Venue** | **2018** | **2021**
--- | --- | ---
IMC | 16% (7/43) | 0% *
CoNEXT | 38% (12/31) | 38% (14/36)
SIGCOMM | 17% (7/40) | 48% (28/58)

* Community award implied a dataset release (but apparently no badge was assigned)
However badges...

1. Do not imply (data) longevity
   *(single snapshots of specific moment in time)*

2. Do not imply (data) generalization
   *(focus on specific problems)*
Going beyond “badges”

All CS conferences, and for sure all the experimental ones @ACMSIGCOMM @usenix, should have a “replication” track to build confidence in the scientific merit of our results.

Fabian E. Bustamante  
@bustamantef

Oliver Hohlfeld @ohohlfeld  ·  Jun 9
Replying to @bustamantef @ACMSIGCOMM and @usenix
CS largely lacks behind other disciplines. E.g., meta-analysis plays a fundamental role in medicine-i.e. statistically combining data from multiple studies on a particular topic. Single efficacy studies are often too small to reliably assess risks. CS is not yet at this point.
Going beyond “badges”

All CS conferences, and for sure all the experimental ones @ACMSIGCOMM @usenix, should have a “replication” track to build confidence in the scientific merit of our results.

My take

- We already do replicability when doing the state-of-the-art comparison
- Yet, we lack venues to foster those discussions
Meanwhile, in other communities

Long tradition for artifacts (SIGMOD since 2008, VLDB since 2012)
Authors submits artifacts evaluated by a committee

Yearly reproducibility challenge
“Crowdsourced”: select a paper, reproduce it, and submit a report

Data-sharing is a core value for earth & space-related sciences
Sharing for the benefit of the research community and humanity

Sara Issaoun PyCon22 keynote: “Imaging a black hole with the event horizon telescope”
Netflix documentary – *Black holes The edge of all we know*
How AI communities foster debate around data

NeurIPS 2021 Datasets and Benchmarks Track

The pre-proceedings are now available! See the NeurIPS Online Proceedings page.

Quickly find papers in the virtual conference: click on the paper in the Accepted Paper List.

We are immensely grateful for the tremendous contributions of the 33 area chairs and 548 reviewers to make this new endeavor a success.

The Datasets and Benchmarks track serves as a novel venue for high-quality publications, talks, and posters on highly valuable machine learning datasets and benchmarks, as well as a forum for discussions on how to improve dataset development. Datasets and benchmarks are crucial for the development of machine learning methods, but also require their own publishing and reviewing guidelines. For instance, datasets can often not be reviewed in a double-blind fashion, and hence full anonymization will not be required. On the other hand, they do require additional specific checks, such as a proper description of how the data was collected, whether they show intrinsic bias, and whether they will remain accessible.

CRITERIA. We are aiming for an equally stringent review as the main conference, yet better suited to datasets and benchmarks. Submissions to this track will be reviewed according to a set of criteria and best practices specifically designed for datasets and benchmarks, as described below. A key criterion is accessibility: datasets should be available and accessible, i.e., the data can be found and obtained without a personal request to the PI, and any required code should be open source. Next to a scientific paper, authors should also submit supplementary materials such as detail on how the data was collected and organized, what kind of information it contains, how it should be used ethically and responsibly, as well as how it will be made available and maintained.

RELATIONSHIP TO NEURIPS. Submissions to the track will be part of the main NeurIPS conference, presented alongside the main conference papers. Accepted papers will be officially published in associated proceedings clearly linked to, yet separate from, the NeurIPS proceedings. The proceedings will be called Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks and they will be hosted on the NeurIPS website next to the main NeurIPS proceedings. We will maintain a page on the NeurIPS website with all accepted datasets and additional information.

174 papers in the program!!!

SIGCOMM’21 58
INFOCOM’21 251
NeurIPS’21 2,300
If I had three wishes for the genie

1. More “code challenges”
   They can be occasion to release data and put focus on specific problems

2. Create one permanent replicability track/workshop
   - Decouple study state-of-the-art from promoting new ideas
   - Foster data/code sharing for the benefit of the community

3. Federate universities/research centers for data access/sharing
   Break the barrier of 1-to-1 cooperations

The data divide affects the whole measurements community
   AI-driven measurement methods is just exacerbating it
...anything else to mitigate the data divide?

What if we reduce the dependency from data?
Part I
Introduction

Part II
Data-sharing & reproducibility

Part III
Easing the data-dependency
The infinite data loop in (AI-based) monitoring

How to reduce data-dependency
- Continuous learning
- Few shot learning
- Self-supervised learning, repr. learning

How to reduce on-device costs
- Real-time inference
- Distributed agents coord.
The infinite data loop in (AI-based) monitoring

How to reduce data-dependency
- Continuous learning
- Few shot learning
- Self-supervised learning, repr. learning

How to reduce on-device costs
- Real-time inference
- Distributed agents coord.
Few-Shots Learning (FSL): A definition

“Learning from a limited number of examples with supervised information”

Popular Classes (many samples)  Rare classes (few samples)

ELI5: Use the accumulated knowledge to solve a new problem using few examples as reference
The need for FSL

“People learning often generalize successfully from just a single example ...

... people learn richer representations than machines do, using them for a wider range of functions, including creating new abstract categories of objects based on existing categories”

Joshua Tenenbaum
Cognitive Scientist
MacArthur Fellow (2019)

Network traffic is imbalanced by nature
“Traffic is neither rack-local nor all-to-all; locality depends upon the service”

Massive (manual) labeling is hard
“Even the most well-known hand labeled datasets [ImageNet] have label error rates of at least 5%”

Pervasive Label Errors in Test Sets Destabilize Machine Learning Benchmarks (arxiv’21)
https://labelerrors.com/

Empirical evidence

Biology

Massive (manual) labeling is hard
“Even the most well-known hand labeled datasets [ImageNet] have label error rates of at least 5%”

Pervasive Label Errors in Test Sets Destabilize Machine Learning Benchmarks (arxiv’21)
https://labelerrors.com/
Let’s consider a practical use-case
Deep Learning and Zero-Day Traffic Classification: Lessons learned from a commercial-grade dataset

Lixuan Yang, Alessandro Finamore, Feng Jun, Dario Rossi
Huawei Technologies, France

A traffic classifier covering 200 classes (4x the literature)

Long-term goal: A classifier covering O(1,000) classes
This talk: A toy-case example

...BTW, we are working internally so to release an anonymized version of the dataset to the community
Modeling Toy-case via a CNN (\textit{monolithic approach})

Dataset (\textit{pkt-size + direction of first 10 pkts})

- 45 classes with $\geq 10$ samples
- Only 10 classes with $> 10^4$ samples

![Graph showing relationship between samples and F1 score]
Modeling Toy-case via a CNN (monolithic approach)

Dataset (\textit{pkt-size + direction of first 10 pkts})

- 45 classes with $\geq$ 10 samples
- Only 10 classes with $> 10k$ samples

- Extreme imbalance : $\rho = 7,000$
- Weighted f1 score : 97.8%
- Macro f1 score : 49%
- 19 classes (42%) with F1 score = 0
Why learning from a few examples is hard?

*FSL and Empirical risk*
Why learning from a few examples is hard?

*FSL and Empirical risk*

\[ \theta_{\text{optim}} \]

- A problem has an optimal solution
Why learning from a few examples is hard?

**FSL and Empirical risk**

- A problem has an optimal solution
- The hypothesis space constrains the best solution that can be found
Why learning from a few examples is hard?

FSL and *Empirical risk*

- A problem has an optimal solution
- The hypothesis space constrains the best solution that can be found
Why learning from a few examples is hard?

*FSL and Empirical risk*

- A problem has an optimal solution
- The hypothesis space constrains the best solution that can be found
- The search is further constrained by the data available (the more the data, the better)
Why learning from a few examples is hard?

**FSL and Empirical risk**

- A problem has an optimal solution
- The hypothesis space constrains the best solution that can be found
- The search is further constrained by the data available (the more the data, the better)
Why learning from a few examples is hard?

FSL and Empirical risk

A problem has an optimal solution

The hypothesis space constrains the best solution that can be found

The search is further constrained by the data available (the more the data, the better)

FSL problems have high empirical risk
FSL methods taxonomy

How to handle empirical risk

**Data:** learn to **augment**
“Hallucinate” the train set by introducing new (synthetic) data

**Model:** learn to **compare**
FSL models derive from baseline models (new classes “compare” with baseline ones)

**Algorithm:** learn to **initialize**
Learn a generalized model from which is easy to derive FSL models
FSL methods taxonomy

How to handle empirical risk

Data: learn to augment
“Hallucinate” the train set by introducing new (synthetic) data

Model: learn to compare
FSL models derive from baseline models (new classes “compare” with baseline ones)

Algorithm: learn to initialize
Learn a generalized model from which is easy to derive FSL models
FSL methods taxonomy

How to handle empirical risk

**Data:** learn to augment

“Hallucinate” the train set by introducing new (synthetic) data

**Model:** learn to compare

FSL models derive from baseline models (new classes “compare” with baseline ones)

**Algorithm:** learn to initialize

Learn a generalized model from which is easy to derive FSL models
FSL methods taxonomy

How to handle empirical risk

**Data:** learn to **augment**
“Hallucinate” the train set by introducing new (synthetic) data

**Model:** learn to **compare**
FSL models derive from baseline models (new classes “compare” with baseline ones)

**Algorithm:** learn to **initialize**
Learn a generalized model from which is easy to derive FSL models

---

\[ \theta_{\text{optim}} \]
\[ \theta_{\text{hypot space}} \]
\[ \theta_{\text{small data}} \] + hallucinations

---

\[ \theta_{\text{optim}} \]
\[ \theta_{\text{hypot space}} \]
FSL methods taxonomy

How to handle empirical risk

Data: learn to **augment**
“Hallucinate” the train set by introducing new (synthetic) data

Model: learn to **compare**
FSL models derive from baseline models
(new classes “compare” with baseline ones)

Algorithm: learn to **initialize**
Learn a generalized model from which is easy to derive FSL models
FSL methods taxonomy

How to handle empirical risk

**Data:** learn to **augment**
“Hallucinate” the train set by introducing new (synthetic) data

**Model:** learn to **compare**
FSL models derive from baseline models (new classes “compare” with baseline ones)

**Algorithm:** learn to **initialize**
Learn a generalized model from which is easy to derive FSL models
FSL methods taxonomy

How to handle empirical risk

Data: learn to augment
"Hallucinate" the train set by introducing new (synthetic) data
Extra knowledge = Generative model

Model: learn to compare
FSL models derive from baseline models (new classes “compare” with baseline ones)
Extra knowledge = a pre-trained model

Algorithm: learn to initialize
Learn a generalized model from which is easy to derive FSL models
Extra knowledge = a pre-trained model

The need for supplementary knowledge

Unexperienced + Very wise = Happy
**FSL methods taxonomy**

*How to handle empirical risk*

**Data:** learn to **augment**

"Hallucinate" the train set by introducing new (synthetic) data

Extra knowledge = Generative model

**Model:** learn to **compare**

FSL models derive from baseline models (new classes “compare” with baseline ones)

Extra knowledge = a pre-trained model

**Algorithm:** learn to **initialize**

Learn a generalized model from which is easy to derive FSL models

Extra knowledge = a pre-trained model

---

The need for supplementary knowledge

---

**Meta-Learning** methods extract such supplementary knowledge
Meta-learning: “learning to learn”

“The goal of the trained model is to quickly learn a new task from a small amount of new data, and the model is trained by the meta-learner to be able to learn on a large number of different tasks.”

Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks (ICML’17)

The “unit of learning” are tasks NOT samples

Aim for a higher level of abstraction/generalization
# Traditional model training

<table>
<thead>
<tr>
<th>classes</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
</table>

Dataset

- A: [Data points]
- B: [Data points]
- C: [Data points]
- D: [Data points]
- E: [Data points]
- F: [Data points]
Traditional model training

Dataset

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Validation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Classes

- A
- B
- C
- D
- E
- F

Partition by samples

Train

Parameters update

Validation

Overfitting and checkpointing

Test

Final performance
Traditional model training

Dataset

<table>
<thead>
<tr>
<th>classes</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>train</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>val</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Train

Validation

Test

partition by samples

Parameters update

Overfitting and checkpointing

Final performance

epochs

1

1

epochs
Traditional model training

Dataset classes
A B C D E F

Train
Validation
Test

Parameters update
Overfitting and checkpointing
Final performance

partition by samples

epochs

train
test
val
train
test
val
train
test
val
train
test
val

Final performance
Traditional model training

- **Classes**: A, B, C, D, E, F
- **Dataset**
- **Train**, **Validation**, **Test**
- **Partition by samples**
- **Parameters update**
- **Overfitting and checkpointing**
- **Final performance**
- **Epochs**

**Best yet**
Traditional model training

Dataset

<table>
<thead>
<tr>
<th>classes</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
</table>

Train

Validation

Test

partition by samples

Parameters update

Overfitting and checkpointing

Final performance

epochs

1

2

1

2

best yet
Traditional model training

Dataset

classes A B C D E F

Train

Validation

Test

partition by samples

Parameters update

Overfitting and checkpointing

Final performance

epochs

1 2 n

Best yet

n

61
Traditional model training

Dataset

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>train</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>val</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**classes** A B C D E F

partition by samples

Train

Validation

Test

Parameters update

Overfitting and checkpointing

Final performance

epochs

1 2 n
Traditional model training

Core principles

1. **One task**: *the model learns from subset of samples from **ALL classes***

2. **Multiple scans**: *the model learns by iterating across **ALL samples***

Meta-Learning violates both of those principles
Model training via meta-learning
Model training via meta-learning

Classes:
- A
- B
- C
- D
- E
- F
- G
- H

Full dataset:

Partition by samples classes:
- Train
- Validation
- Test
Model training via meta-learning

N = 2 classes
K = 1 support
Q = 5 query
Model training via meta-learning

Full dataset

Train

Validation

Test

1. Some classes/samples are hold back

N = 2 classes
K = 1 support
Q = 5 query
Model training via meta-learning

### Full dataset

- **N = 2 classes**
- **K = 1 support**
- **Q = 5 query**

### Partition by samples classes

- **Train**
- **Validation**
- **Test**

1. Some classes/samples are hold back
2. Update params based on how the embedding of support generalize for the query
Model training via meta-learning

Full dataset

Train

Validation

Test

\[ N = 2 \text{ classes} \]
\[ K = 1 \text{ support} \]
\[ Q = 5 \text{ query} \]

Some classes/samples are hold back

Update params based on how the embedding of support generalize for the query

As episodes progress, the model gets “wiser and wiser”
Model training via meta-learning

- **Full dataset**
- **Train**
- **Validation**
- **Test**

**partition by samples classes**

N = 2 classes
K = 1 support
Q = 5 query

1. Some classes/samples are hold back
2. Update params based on how the embedding of support generalize for the query
3. As episodes progress, the model gets “wiser and wiser”
4. Validation against classes NEVER seen at training

Validation against classes NEVER seen at training
Model training via meta-learning

N = 2 classes
K = 1 support
Q = 5 query

epochs
epochs

partition by samples classes
Model training via meta-learning

Full dataset

Train

Validation

Test

N = 2 classes
K = 1 support
Q = 5 query

partition by samples classes

Validation against classes NEVER seen at training or validation
FSL in action (1/2)

Monolithic approach
*Model everything at once*

- **FSL** Protonet (NIPS’17)
  - 5-shots

BaselineCNN

- Extract knowledge from popular classes
- Reuse knowledge on unpopular classes

#samples vs. classes

20 classes vs. 25 classes

#samples vs. #samples

f1 score vs. samples
FSL in action (2/2)

Not too bad performance considering we use 5-shots: good extrapolation power

Protonet model is 50x smaller than BaselineCNN

FSL significantly better than alternatives

...yet arguably far from being production ready
Conclusions

1. Many open challenges surrounding the data divide
   *Face them together as a community*

2. Opportunities not fully explored offered by AI
   *Few-shot learning, Continuous learning, etc.*
Datacom AI4NET Lab - Recent pointers

[TNSM-22] D. Rossi, L. Zhang
Landing AI on Networks: An equipment vendor viewpoint on Autonomous Driving Networks

Accelerating Deep Learning Classification with Error-controlled Approximate-key Caching

[IEEE Network-21] L. Yang, D. Rossi
Quality monitoring and assessment of deployed Deep Learning models for Network AIOps

[ICML-UDL-21] L. Yang, D. Rossi,
Thinkback: Task Specific Out-of-Distribution Detection

FENXI: Fast In-Network Analytics

A First Look at Class Incremental Learning in Deep Learning Mobile Traffic

Alessandro Finamore
Principal engineer, AI4NET Datacom lab
Huawei Technologies Research Center Paris
alessandro.finamore@huawei.com
afinamore.io