Machine Learning Engineering

(and Sustainability)

Prof. Dr. Michael Granitzer
Chair of Data Science
University of Passau

contributions from Thomas Weißgerber, Christofer Fellicious, Mehdi Ben Amor, Lorenz Wendlinger

Virtual Summer School
25.06.2020, Online

Research has been partially funded by the INTERREG project DataKMU

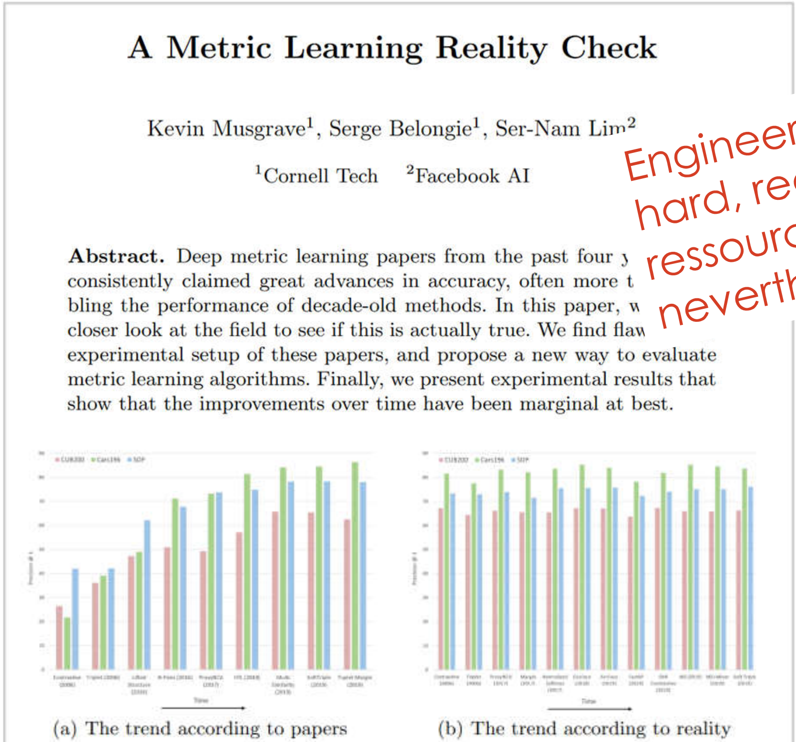
What constitutes a good Machine Learning solution?

1. Machine Learning Algorithm
2. Dataset
3. Hyperparameter Tuning
4. Evaluation Metric

Often overlooked: The engineering process (or how we bring 1-4 together)

- Software Engineering
- Experimental Setup
- Data Management

Some Examples of Engineering Impact



A Metric Learning Reality Check

Kevin Musgrave[1], Serge Belongie[1], Ser-Nam Lim[2]

[1]Cornell Tech [2]Facebook AI

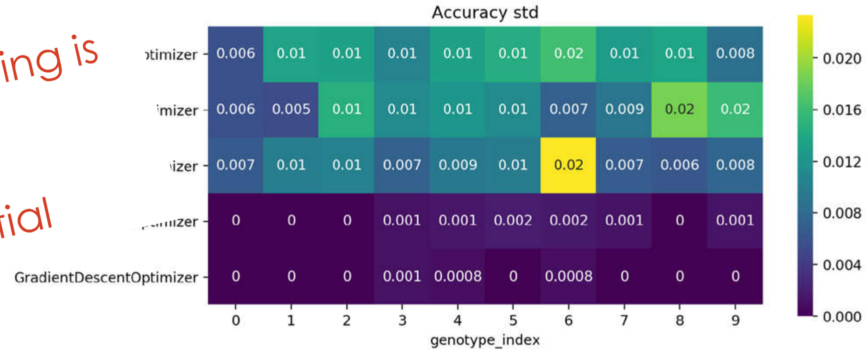
Abstract. Deep metric learning papers from the past four y consistently claimed great advances in accuracy, often more t bling the performance of decade-old methods. In this paper, w closer look at the field to see if this is actually true. We find flaw experimental setup of these papers, and propose a new way to evaluate metric learning algorithms. Finally, we present experimental results that show that the improvements over time have been marginal at best.

(a) The trend according to papers (b) The trend according to reality

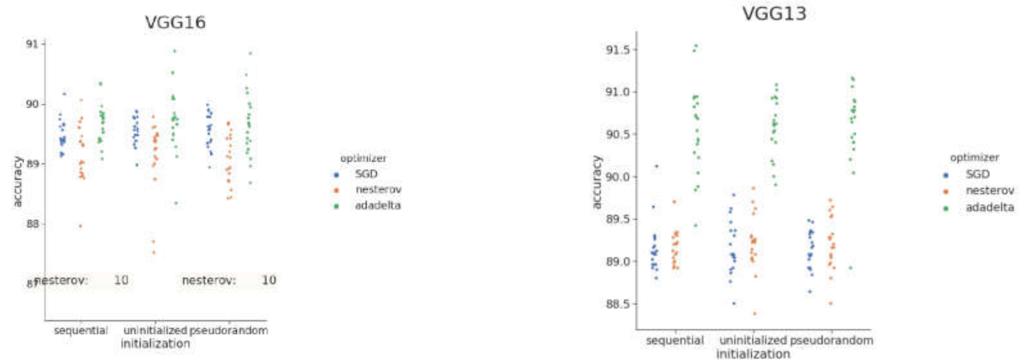
Engineering and testing is hard, requires lots of ressources, but is nevertheless essential

Proper engineering, fair comparison

„Engineering > Algorithm"



Influence of non-derteministic GPU execution



Optimisation Parameters + Random Seeds gives ~ ±1%

Fellicious, Weißgerber, Granitzer, Effects of random seeds on the accuracy of Convolutional Neural Networks, LOD 2020

Why is ML Engineering complex?

Some Factors

- Dataset Creation (train/test/validation/**development** sets)
- (Big) Data Management
- Potentially infinite model choices and hyperparameter tuning
- Moving Targets / Changing Requirements / Learning on the go
- Complex Technical Infrastructure
- Demanding Skillset of Practioniers / Interdisciplinary Teams (Coding, Math, Communication, Domain Knowledge etc.
- Reporting Results (e.g. in Papers)

Our Wish List

What we want to have…

- Logging strategies
- Robust and standardized testing
- DevOPs – Deployment, Data Management
- Dont Repeat Yourself - Reuse existing experimental results
- Collaboration and Communication Support (beyond ipyn's)

What we dont want to have…

- Additional effort in
  - Coding
  - Deployment
- Setting up additional environments or tools
- Change underlying software development and business processes

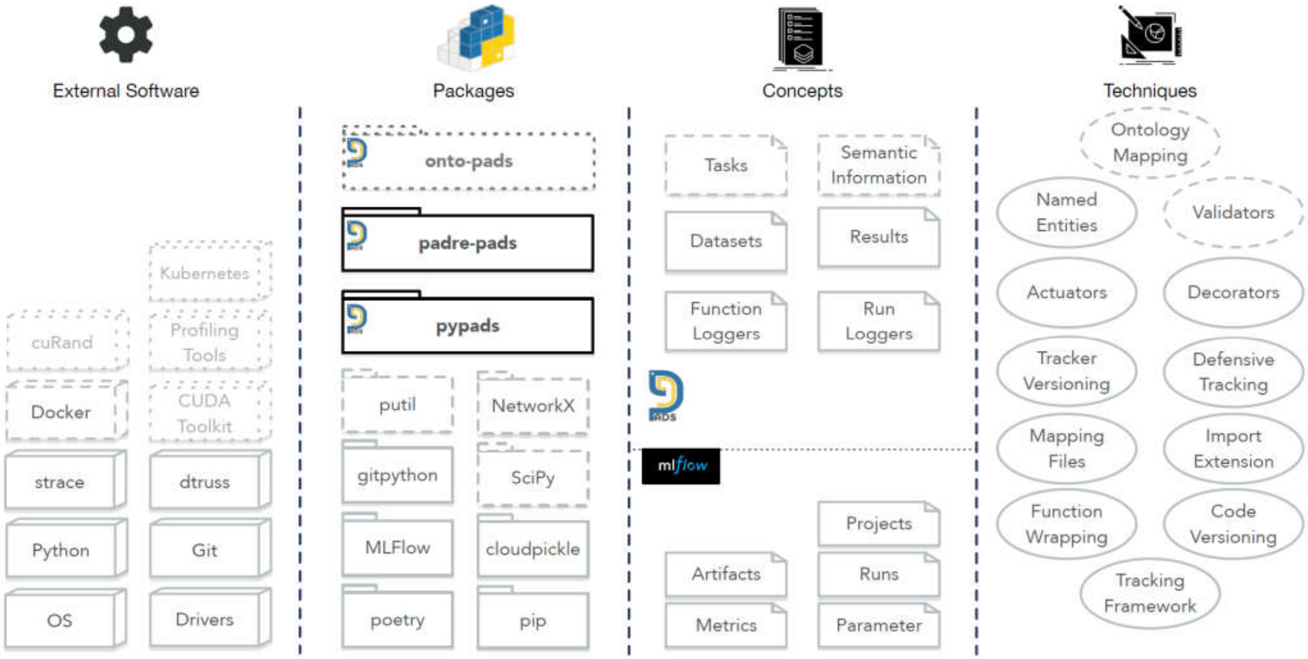
Comparable to general Software Engineering?

Our Contribution - PyPads and Padre

https://github.com/padre-lab-eu/

Python Library plus Wikidata Infrastructure for non-intrusive logging and semantic analysis.

- Monkey-patching of existing frameworks for automatic (detailed) logging
- Post-hoc semantic analysis of logs for inter-library comparison and archiving
- Descriptive metadata format for community driven mappings




```python
from pypads.base import PyPads
tracker = PyPads(...)

from sklearn.decomposition import PCA
from sklearn.linear_model import LogisticRegression
""" Additional experiment code ... """


{"algorithms": [{
                "name": "Object name",
                "implementation":
                    {"package":"package.moduleX.ClassA"},
                "hooks": {"Name of the Hook":
                          ["Hook expressions"]}}],
"metadata": {"author": "John Smith",
             "library": "package",
             "library_version": "x.xx.xx",
             "mapping_version": "x.x"}}


tracker = PyPadrePads(...)

@tracker.decorators.dataset()
def load_data(*args,**kwargs):
    """ Load your dataset and modify it ... """
    return data
```

Conclusion

- ML Engineering is highly undervalued
  „We present a new Algortihm“ vs. „We properly engineered …“
- Comparable to Software Engineering, we need a better ML Engineering
- Tools for automating „as much as possible“
- PyPads and Padre with contributions to automated logging and semantic harmonization

Questions?

Research has been partially funded
by the INTERREG project DataKMU