University of Glasgow

School *of* Computing Science
Knowledge & Data Engineering Systems

Knowledge Re-usability at the Edge

Dr Christos (Chris) Anagnostopoulos
Senior Lecturer in Data Engineering & Distributed Computing
Knowledge & Data Engineering Systems Group
School of Computing Science

WORLD CHANGING GLASGOW

Scottish Autonomous Networked Systems (SANS)
12-13th Dec2022
James Watt School of Engineering, University of Glasgow

University of Glasgow

School of Computing Science
Knowledge & Data Engineering Systems

Knowledge & Data Engineering Systems

The **Knowledge & Data Engineering Systems** (KDES) research group is part of the Information, Data and Analysis (IDA) Section.

Vision:

- KDES brings together the fundamental research areas of **Distributed Data Systems**, **Data Engineering**, and **Data Science**.
- KDES's activities range *from* large-scale Distributed Computing, *to* Edge Computing and Distributed Machine Learning / AI.

Research Topics:

- Distributed Data Management at the Edge
- **Resilient & Distributed Machine Learning**
- Large-scale Data Analytics
- Information Processing Systems

Knowledge & Model Re-usability

Context: Unprecedented growth of data *surpasses* current predictive models and processing capabilities.

i.e., we generate more data (big data) *than* we (want to) process (relevant data)

Fact: Redundancy because of *similar* analytics tasks, *similar* predictive models, even *similar* data!

Rhetorical?: Do we *need* all the data/models? Do we need to *analyse* all the data/*train* all these models? Do we need all these redundant models?

Challenge: Can we extract *similar performance patterns* from models and data so that ***existing*** predictive models can be ***reused*** or made ***reusable***?

Benefit: Avoid *building* and *maintaining reduplicative* predictive models since reusable models can be ‘reused’ by other nodes’ predictive tasks (e.g., classification, image recognition, regression).

Reuse existing models
Or, make models *reusable*

Principles for Model Re-usability

Multi-task Learning (MtL): case of **Personalized Federated Learning (PFL)**, which addresses challenges raised by *heterogeneous* data and *heterogeneous* tasks via training *personalized* models.

Principle: MtL paradigm learns *multiple tasks* simultaneously by optimizing *multiple* loss functions *at once*.

i.e., leverages statistical information contained in *each task individually* to promote overall model's performance on *all* tasks.

Tasks: *e.g.*, image classification of identified objects in AVs, object recognition, regression.

Our Target: Models *useful* in multiple contexts/tasks/data, therefore, being *reusable*. Thus, nodes can *reuse* those models for a family of tasks (e.g., object classification) *without the need of* training new ones.

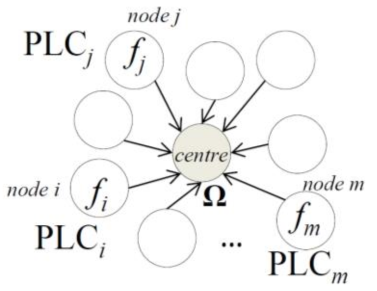
Our Idea: Instead of training independent (local) models on nodes with *less* capacity to be reused, we contribute with MtL models that learn from *all* of nodes' tasks at once.

Fact: MtL excels when tasks/data have some level of correlation/similarity, *which is the reality in our case*.

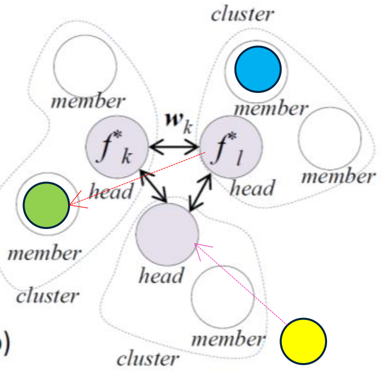
Contribution: A Two-phase Distributed Multi-task Learning (DMtL) Framework making *reusable* models (being *re-used* by nodes).

Narrative:

- Nodes initially train their *local models* & produce their *performances* on *local* tasks.
  - Partial Learning Curves (PLC): universal indicators of model performances used in many contexts (e.g., hyper-parameter selection of deep learning models).
  - *Why?* PLCs are (i) cheap to store and compute, thus, suitable for resource-constrained nodes and (ii) produce insights for model performance in meta-learning context.
- Identify correlations among models' performances and tasks via PLCs, thus, *nodes are grouped together*; Cluster-heads are then selected.
- Distributed MtL runs across *only* cluster-heads *capturing* the per-group diversity of tasks & data
- Cluster-heads generate models, which can be reused by *any* member of *any* group.



(a)



(b)

University of Glasgow

School *of* Computing Science

Knowledge & Data Engineering Systems

Thank you!

References:

- Q. Long, C. Anagnostopoulos, F. Deligianni, K. Kolomvatsos, *Model Reuse in Distributed Computing: A Multitask Learning Approach Based on Partial Learning Curves*. Available at SSRN: *https://ssrn.com/abstract=4253481*
- Q. Long, K. Kolomvatsos, C. Anagnostopoulos, *Knowledge reuse in edge computing environments*, Journal of Network and Computer Applications, Elsevier, 206, 103466, 2022 (*https://doi.org/10.1016/j.jnca.2022.103466*)
- Skotti, X., Kolomvatsos, K. Anagnostopoulos, C. (2022) *On the Reusability of Machine Learning Models in Edge Computing: A Statistical Learning Approach*. In: Arai, K. (ed.) Future Technologies Conference (FTC) 2022, Volume 3.

https://www.gla.ac.uk/schools/computing/research/researchsections/ida-section/

https://www.gla.ac.uk/schools/computing/research/researchsections/ida-section/knowledgeanddataengineeringsystems/