Knowledge and Information Reuse in the Edge Computing Environments

Mr Qianyu Long

Supervisors: Dr Christos Anagnostopoulos, Dr Fani Deligianni

School *of* Computing Science
Essence: Data Science & Distributed Computing Lab

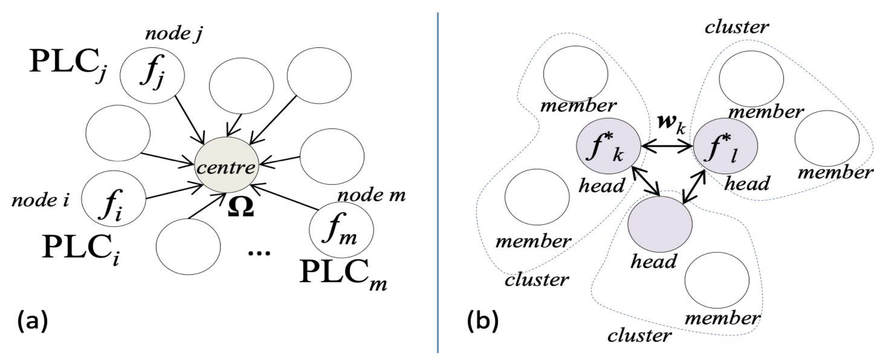
Introduction

• **Edge Computing (EC)** is the idea of moving Cloud services near to the users, which distributed the computation and sources to edge servers and devices.
• Several approaches have been proposed to deal with challenges in EC, such as task offloading [1], service migration [2] and collaborative computing [3]. **Knowledge reuse** attracts attention recently [4], leveraging existing knowledge and models to deal with learning tasks or objectives.
• Application scenarios can be in **Edge Intelligence (EI),** e.g., Smart City, Intelligent Grids and Intelligent Healthcare.
• We notice the potential of knowledge reuse in edge computing and get inspired by recent developments in research and it is in its infancy stage.

Research Questions

What are the advantages of reusing existing knowledge at edge networks?
• In the context of meta-learning, which meta-features should be used to decide the feasibility of reuse?
• How to address challenges existed in EC environments, like non-Identical-Independent-Distributed (non-iid) cases?
• In reuse, what are the *common* and *different* aspects for statistical machine learning (SML) & deep learning (DL).

Figure 1. Multitask relationship Learning based on information reuse



Methods & Results

Reuse in SML & DL

The feasibility of reuse depends on the similarity between tasks and data. Intuitively, if <u>similar</u> tasks and data is reused, then redundant storage and computation for knowledge (including models and data) can be avoided.

• In ML, researchers leverage meta-features to measure the aforementioned similarity. Meta-features can be learned from task/data properties, model evaluation and prior models (parameters).
• The concept of **knowledge/information reuse in DL** also relies on the meta-learning. Specifically, it is meta-representation learning, which ranges over many areas such as parameter initialization, embedding functions and instance weight.

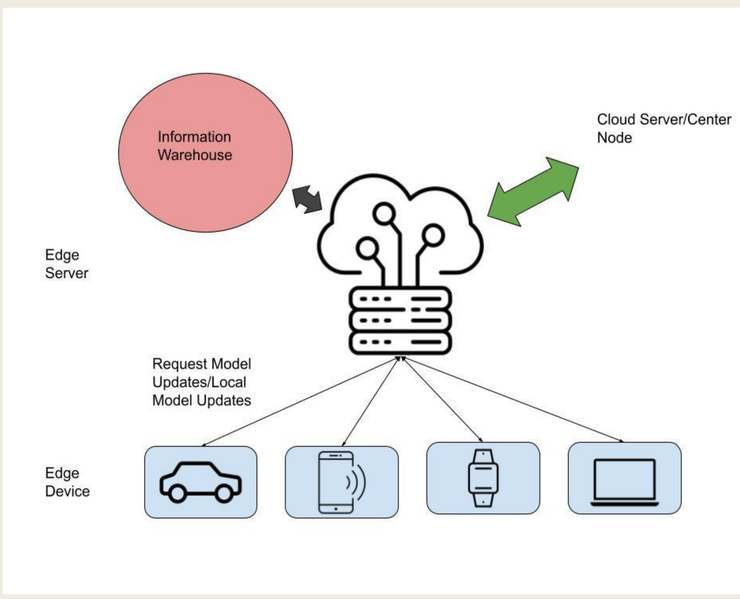


Figure 2
Reuse in Cloud-Edge Computing with Federated Learning

All edge devices upload their updates model information to the neighborhood edge server. An information warehouse is connected with the edge server, which store the historical information. The edge server communicates with the cloud server and edge devices and it decides which information is informative to upload and download based on information warehouse.

Achieved Approaches

We used the **Maximum Mean Discrepancy (MMD)** and Cosine Dissimilarity (CD) as <u>synopses</u> to decide whether model reuse is proper between two edge across networks.

A monitor mechanism with Holt-Winter models based on Sum Square of Errors (SSE) is proposed to ensure the model reusability.

• We leveraged **Learning Curves (LC)** to cluster tasks in multitask learning. Not only the partial LC provides information for relationship among tasks but also it help tuning the models.
• All experiments are conducted based on the synthetic and real world datasets such as CIFAR10, Sentiment, Air Pollution and GNFUV (Autonomous Vehicles).

Conclusion & Future work

Reusing executed models or knowledge has attracted recent research attention due to the large amount of data and redundancy in tasks. According to the advantages brought by knowledge reuse, showcased by our experiments, we believe it will play an important role in efficient information management in edge networks.

In this work, we have summarized the past works, which show good potential of reuse information for statistical machine learning models. Current work can be extended with other suitable meta-features such as KL-Divergence and statistics for the two-sample tests. Reuse information is also necessary in the deep learning. We schedule to investigate how to reuse knowledge to accelerate training and inference of Deep Neural Networks (DNN) with techniques like weight reuse, layer reuse and entire module reuse in the edge computing environments.

References

[1] T. X. Tran and D. Pompili, "Joint task offloading and resource allocation for multi-server mobile-edge computing networks," IEEE Transactions on Vehicular Technology, vol. 68, no. 1, pp. 856–868, 2018.
[2] T. Ouyang, Z. Zhou, and X. Chen, "Follow me at the edge: Mobility-aware dynamic service placement for mobile edge computing," IEEE Journal on Selected Areas in Communications, vol. 36, no. 10, pp. 2333–2345, 2018.
[3] Y. Sahni, J. Cao, and L. Yang, "Data-aware task allocation for achieving low latency in collaborative edge computing," IEEE Internet of Things Journal, vol. 6, no. 2, pp. 3512–3524, 2018.
[4] J. Lee, A. Mtibaa, and S. Mastorakis, "A case for compute reuse in future edge systems: An empirical study," in 2019 IEEE Globecom Workshops (GC Wkshps). IEEE, 2019, pp. 1–6.