Optimization of Knowledge Distillation in Heterogeneous Federated Problems

Wenhao Li, Christos Anagnostopoulos, Kevin Bryson

School of Computing Science

University of Glasgow, UK
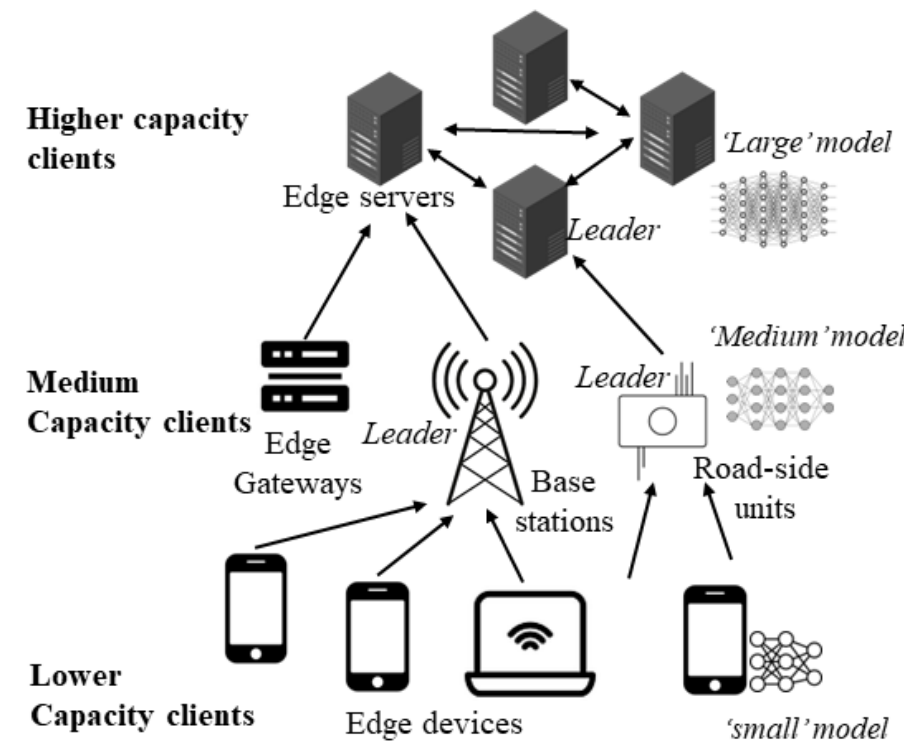
Motivation

Heterogeneous Federated Learning
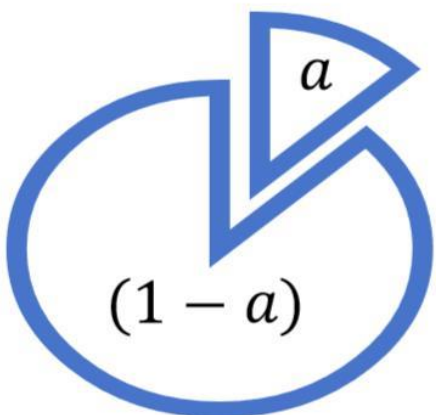
Multi-Level Client

Knowledge Distillation

Knowledge Distillation in FL

➢ Compress knowledge to align teacher and student performance

Commonly adopted methodology :
✓ **Soft Label – based KD**

For student

$$\mathcal{L}^S := \underbrace{(1 - \alpha)\,\mathcal{L}_{\mathrm{CE}}(\boldsymbol{p}^S, \boldsymbol{y})}_{①} + \underbrace{\alpha\,\mathcal{D}_{\mathrm{KL}}(\boldsymbol{p}^S \parallel \boldsymbol{p}^T)}_{②}$$

① Cross-entropy loss -- Ground Truth (Hard) Knowledge

② KL divergence loss– Soft label (Soft) Knowledge

For teacher

$$\mathcal{L}^T := \underbrace{(1 - \beta)\,\mathcal{L}_{\mathrm{CE}}(\boldsymbol{p}^T, \boldsymbol{y})}_{①} + \underbrace{\beta\,\mathcal{D}_{\mathrm{KL}}(\boldsymbol{p}^T \parallel \boldsymbol{p}^S)}_{②}$$

a

$(1 - a)$

Research questions

RQ1: How to match teachers and students in a multi-level structure?

Knowledge distillation weight design

RQ2: How to dynamically trade off hard knowledge and soft knowledge

RQ3: How to design a more personalized distillation learning process

Proposed Research

Label distribution-based assessment considers similarity, trend toward uniformity, and complementarity (RQ1)

Dynamic adjustment design(RQ2):

Early stages, focus on personalized hard knowledge (**accuracy**);	During training, focus on consistency soft knowledge (**robustness**)
$\alpha \downarrow$	$\alpha \uparrow$

Personalized group distillation(RQ3):Mutual learning between teachers and students, and learning with peers within the group. (loss-based revenue)

Research challenges

Selection of teacher? Personality or Consistency

Multi-level distillation process

Path constraints for distillation

Thank you for your attention!

? Questions?

✉ w.li.7@research.gla.ac.uk