Qualcomm

Variational On-the-Fly Personalization

Jangho Kim*,[1], Jun-Tae Lee*,[2], Simyung Chang[3], Nojun Kwak[1]

[1] Seoul National University
[2] Engineer, Staff, Qualcomm AI Research
[3] Engineer, Sr Staff/Mgr, Qualcomm AI Research
kjh91@snu.ac.kr, juntlee@qti.qualcomm.com, simychan@qti.qualcomm.com, nojunk@snu.ac.kr

* Equal contribution
Jangho Kim completed the research in part during an internship at Qualcomm Technologies, Inc.
Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc

Jangho Kim
Former Interim
Engineering Intern
Qualcomm Korea YH

Index

- Introduction
- Method
- Experiment results
- Conclusion

Introduction

- Problem
  - In edge devices, such as mobile phones and IoT sensors, deep models are required to process (learn or infer) a personal domain where data are generated in a specific environment, which is called *personalization*
  - Despite the importance of personalization, there has been little progress due to practical constraints of edge devices
    - Source-free
    - Few-shot
    - Unsupervised
    - Training-free
- Goal
  - we propose a novel personalization method, Variational On-the-Fly Personalization (VoP) satisfying the constraints (source-free, few-shot ,unsupervised ,training-free)

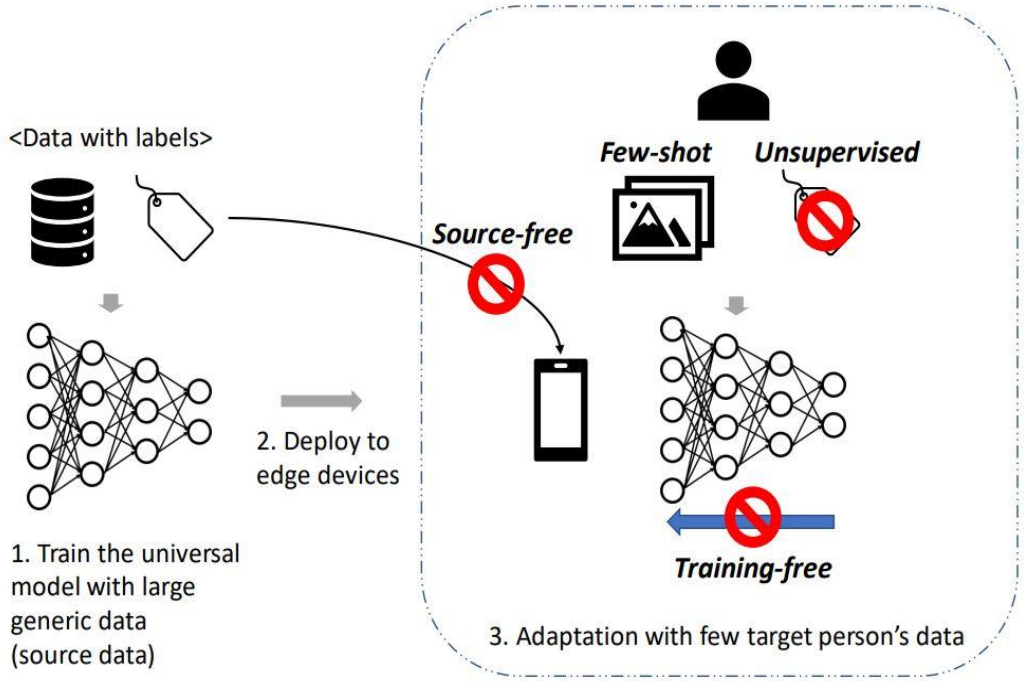

Figure 1: Our personalization scenario on edge devices with practically crucial constraints

Method

- Variational On-the-Fly Personalization
  - we compute the weights of layers specialized to its personality on-the-fly via forwarding only a few personal data
  - The key of our method lies in a small detachable module, the variational hyper-personalizer which is trained to produce an approximated posterior distribution of weights of a layer based on the personality
  - We assume that the data in a personal domain share the same personality

Variational distribution True k-th personality posterior distribution

$$KL(q_\theta(\omega|x_i^k)||p(\omega|x_i^k, y_i^k)) = \int q_\theta(\omega|x_i^k) \ln \frac{q_\theta(\omega|x_i^k)}{p(\omega|x_i^k, y_i^k)} d\omega$$

$$= \int q_\theta(\omega|x_i^k) \ln \left(\frac{q_\theta(\omega|x_i^k)}{p(\omega|x_i^k)} \frac{p(y_i^k|x_i^k)}{p(y_i^k|x_i^k, \omega)} \right) d\omega$$

$$= KL(q_\theta(\omega|x_i^k)||p(\omega|x_i^k))$$

$$- \mathbb{E}_{\omega \sim q_\theta(\omega|x_i^k)} [\ln p(y_i^k|x_i^k, \omega)] + \ln p(y_i^k|x_i^k).$$

Approximating the posterior distribution

Method

- Variational On-the-Fly Personalization
  - Colors in input samples represent personalities for each input sample
  - In the training phase, VoP trains the encoding module and hyper-personalizer to estimate sample-specific weights via black dashed and blue bold arrows
  - At testing phase, for each personality, VoP generates personal weights by forwarding a few enrollment samples via black and red dashed arrows, once

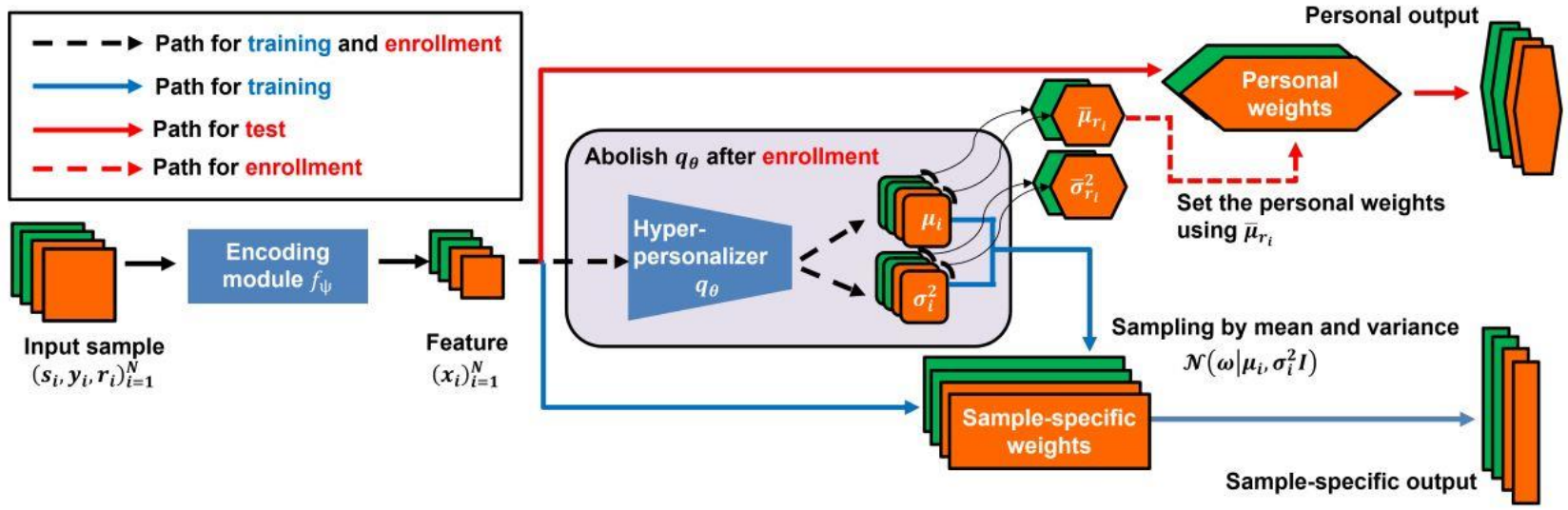

Figure 2: Overall Process of VoP

Experiment results

- To verify the proposed VoP, we apply it to three tasks: keyword spotting, speaker verification and few-shot classification (Details of other experiments are in the paper)

Method	Closed-set	Open-set
Baseline	87.46 ± 1.68	74.45 ± 0.77
Baseline w/ Dropout	81.77 ± 1.75	77.35 ± 1.90
Baseline w/ samovar (2fc)	17.25 ± 1.42	24.35 ± 1.84
Baseline w/ samovar (1fc)	87.47 ± 1.27	81.43 ± 1.02
VoP	92.80 ± 1.40	83.60 ± 0.84

Table 1: Keyword spotting accuracy on Qualcomm keyword speech dataset

Conclusion

- We proposed Variational On-the-Fly Personalization (VoP), a novel personalization method that can produce a personalized network via forwarding a small amount of personal data on-the-fly
- The proposed VoP can effectively estimate the weight distribution suitable for an individual without additional training using a large amount of personal data

Thank you

Qualcomm

Follow us on:

For more information, visit us at:

qualcomm.com & qualcomm.com/blog