Variational On-the-Fly Personalization

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

[1] Qualcomm AI Research
[2] Seoul National University
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.
[1] Qualcomm AI Research is an initiative of Qualcomm Technologies, Inc

Jangho Kim
Interim Engineering Intern
Qualcomm Korea YH

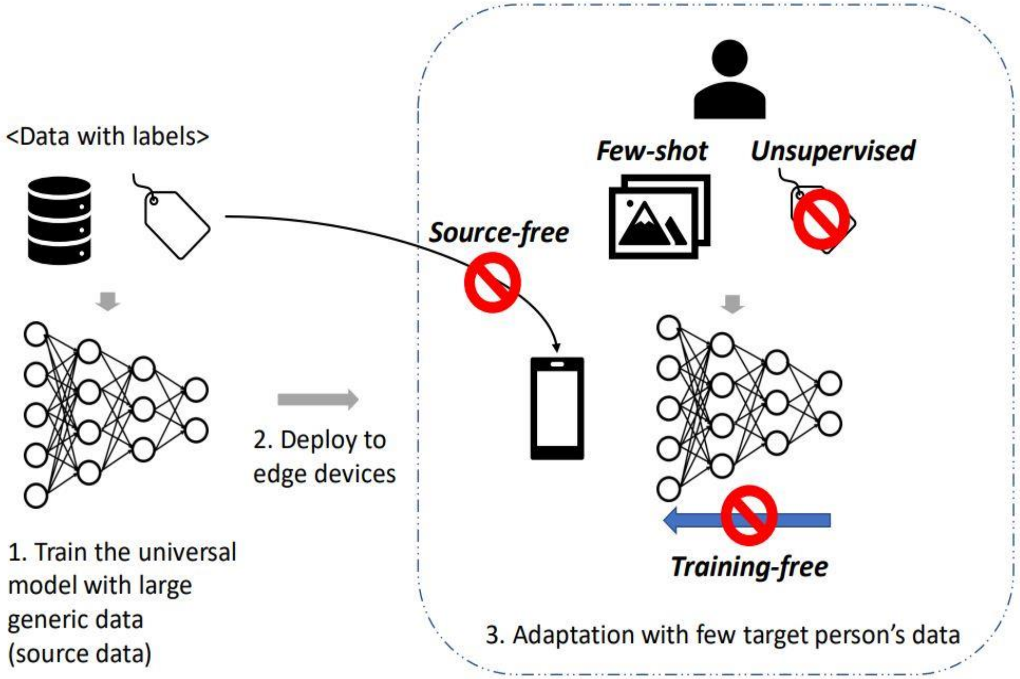
Index

- Introduction
- Method
- Experiment results
- Conclusion

Source sample text

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)



<Data with labels>
Few-shot
Unsupervised
Source-free
2. Deploy to edge devices
1. Train the universal model with large generic data (source data)
Training-free
3. Adaptation with few target person's data

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

$$KL(\overset{\text{Variational distribution}}{\underline{q_\theta(\omega|x_i^k)}}||\overset{\text{True k-th personality posterior distribution}}{\underline{p(\omega|x_i^k, y_i^k)}}) = \int q_\theta(\omega|x_i^k) \ln \frac{q_\theta(\omega|x_i^k)}{p(\omega|x_i^k, y_i^k)} d\omega$$

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

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

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

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

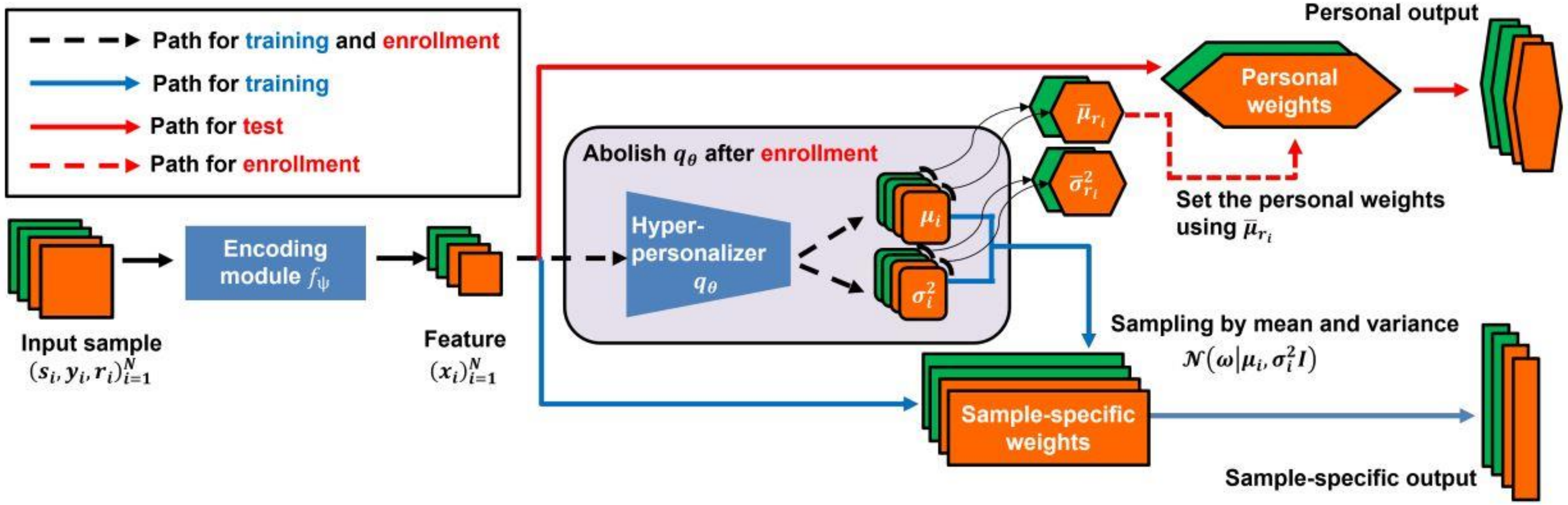
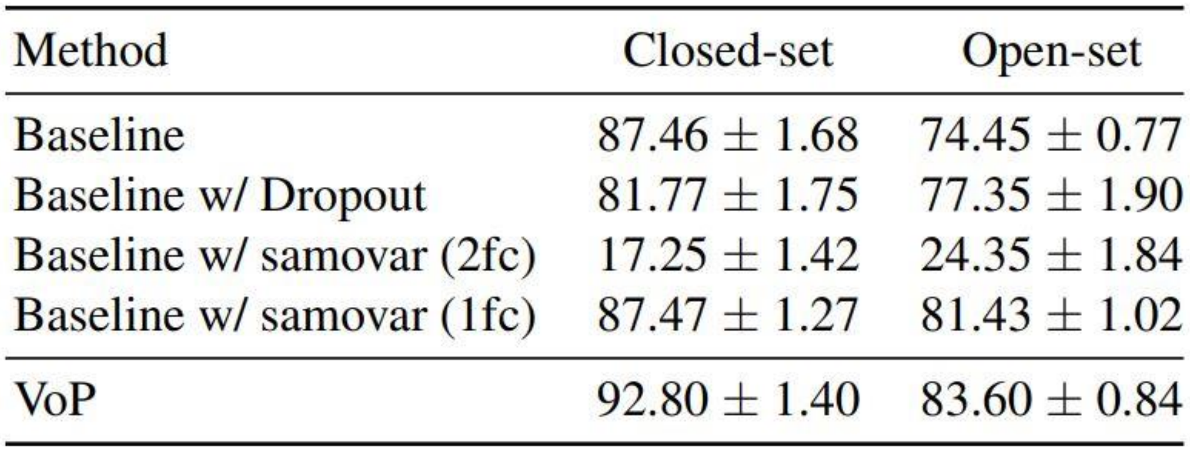


Figure 2: Overall Process of VoP

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
- we showed that VoP successfully generates an accurately personalized model without increasing the computational cost

Source sample text

Thank you

Qualcomm

Follow us on:

For more information, visit us at:

qualcomm.com & qualcomm.com/blog

Nothing in these materials is an offer to sell any of the components or devices referenced herein.

References in this presentation to “Qualcomm” may mean Qualcomm Incorporated, Qualcomm Technologies, Inc., and/or other subsidiaries or business units within the Qualcomm corporate structure, as applicable. Qualcomm Incorporated includes our licensing business, QTL, and the vast majority of our patent portfolio. Qualcomm Technologies, Inc., a subsidiary of Qualcomm Incorporated, operates, along with its subsidiaries, substantially all of our engineering, research and development functions, and substantially all of our products and services businesses, including our QCT semiconductor business.