Poster: Towards Condition-Independent Deep Mobile Sensing

Taesik Gong, Yeonsu Kim, Jinwoo Shin, and Sung-Ju Lee
{taesik.gong,yeonsu.kim,jinwoos,profsj}@kaist.ac.kr
KAIST

ABSTRACT

Deep mobile sensing applications are suffering from various individual conditions in the wild. We propose a meta-learned adaptation technique to adapt to a target condition with a few labeled data. We evaluate our system on a public dataset and it outperforms baselines.

CCS CONCEPTS

• **Human-centered computing → Ubiquitous and mobile computing systems and tools**; *Mobile computing*.

ACM Reference Format:
Taesik Gong, Yeonsu Kim, Jinwoo Shin, and Sung-Ju Lee. 2019. Poster: Towards Condition-Independent Deep Mobile Sensing. In *The 17th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '19), June 17–21, 2019, Seoul, Republic of Korea.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3307334.3328622

1 INTRODUCTION

While recent studies have opened the possibility of novel sensing applications with smart devices, they face performance challenges when deployed in the wild. Performance degradation is mainly caused by the existence of a vast number of *individual conditions*; we define an individual condition as a combination of multiple dependencies (e.g., the user's behavior and device) that occur when deployed to an individual, which is usually different from the condition where the sensing model was trained. Taking human activity recognition (HAR) with smartphones as an example; users would have their own patterns of activities (e.g., walking speed and stride) which in turn generate dissimilar sensor values across users. Moreover, users hold smartphones in various ways, e.g., in a pocket or on a hand that affects the orientation and position of the sensors. In addition, individuals have different smartphones that have heterogeneous sensor readings due to the diversity of both software and hardware specifications. Each of these dependencies, and the countless combination of these dependencies, is known to significantly degrade the performance of mobile sensing systems when deployed in practice [3, 4]. These differences in individual conditions hinder mobile sensing systems from operating well for *unseen* users and thus overcoming this issue has become an important research question.

We present a system that adapts deep sensing models for a target user with a few *shots* from the target (one shot is one labeled sample per class). Figure 1 is an overview of our meta-learned adaption. We trains a base model by giving multiple synthetic conditions via our task generation strategies. With each task the base model experiences various individual conditions and make its parameters adaptive to a new/unseen condition. When the base model is deployed to a user, it adapts to the target user's condition with only a few shots (e.g., one or two). Since our approach works with only a few shots, it significantly diminishes the laborious data collection process for each user while still achieving desirable performance. It is applicable to any deep learning models (i.e., model-agnostic) and does not limit its coverage to specific type of sensors or applications (i.e., condition-agnostic). Our solution also entails significantly less adaptation time compared to the traditional deep neural networks training, which is undoubtedly beneficial to the resource-constrained mobile devices.

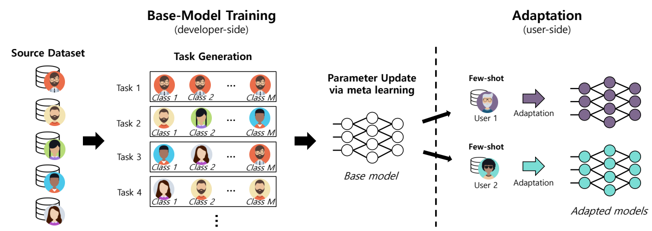


Figure 1: Meta-learned adaptation overview.

2 WHY CONDITIONS MATTER

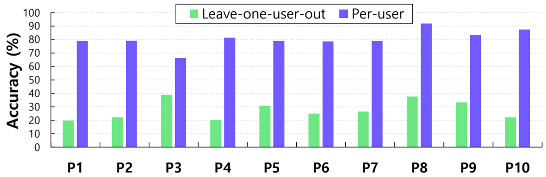
While recent studies have shown the potential of a variety of mobile sensing applications powered by deep learning, they must overcome the challenge of diverse *individual conditions* for wider adoption. Mobile sensing applications get input from the sensors in smart devices for their services, e.g., Inertial Measurement Unit (IMU) for motions and microphone for audio. The sensed values, however, are highly dependent on various conditions which in turn deteriorates the performance when faced with an untrained condition. We summarize two major categories where the individual conditions come from:

User dependency: Humans have different physical conditions and behaviors that make them unique between each other. In human activity recognition (HAR) for instance, users have dissimilar patterns of "walking" in term of the speed and stride, which could be confused with someone's "running". In addition, some people prefer to put their phone in their pocket, while others hold in hand, and each smarthphone position makes different sensor readings even with the same device. Since users' behaviors are unbounded and cannot be easily characterized in advance, user dependency is one of the major obstacles for mobile sensing to overcome.

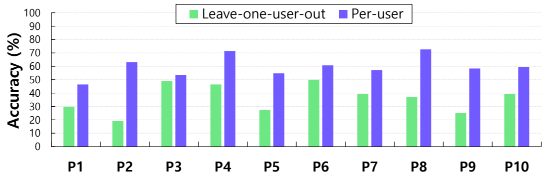
Device dependency: Users have their own devices that have a different shape, weight, sensor specification, and so on, which make the model get different sensor values. Especially for IMU sensors,

MobiSys '19, June 17–21, 2019, Seoul, Republic of Korea

ACM ISBN 978-1-4503-6661-8/19/06.
https://doi.org/10.1145/3307334.3328622



(a) Activity recognition with motion sensors.



(b) Speech recognition with audio inputs.

Figure 2: Per-user accuracy of activity recognition and speech recognition.

different devices have different sensor biases, errors, and sampling rates. In addition, software heterogeneity (e.g., different versions of OS) makes sensor readings different. With the recent spread of wearable devices, some users might run the sensing application in their wearable devices instead of smartphones. As the number of unique Android devices has already exceeded over 24,000 in 2015 [1], it seems infeasible to collect data from all possible devices in advance to train and make the model work effectively for every device.

3 CASE STUDY

Figure 2 shows the accuracy for activity recognition and speech recognition of our models trained with leave-one-user-out and per-user manner, respectively. For our activity recognition dataset (Figure 2a), even though leave-one-condition-out models utilized nine times more training data, they show significantly worse performance than the per-user models. This suggests the combination of multiple dependencies makes the model difficult to generalize its performance to unseen conditions. Figure 2b shows that a similar trend exists for our speech recognition data. Both results show adapting the sensing models to unseen conditions is important.

4 META-LEARNED ADAPTATION

We consider a practical scenario where a model developer has a *source dataset* collected under several *conditions*.Under the scenario, our goal is to adapt to a new/unseen user’s condition when only a few target user’s data samples are available. We denote a labeled data instance for each class as a *shot*. Note that we assume a few shots (e.g., one or two) are given from the target user and the original source dataset does not contain any data samples from the target user. To handle adaptation with only a few shots, we design a meta learning framework, also known as learning to learn, to train the model. Meta learning generally aims to learn a new task or environment rapidly, by learning how to learn. Hence, the meta-objective is learning effective parameters that has an ability to adapt to an unseen condition.

Figure 1 shows an overview of our meta-learned adaptation. Specifically, we makes the base model learn how to adapt to a new condition with only a few shots.The base model is trained on a set of *tasks*, where each task is generated from the source dataset. Individual task mimics a situation where the model performs under a new untrained condition. After training, the base model has the knowledge of how to adapt to a new condition with a few shots. In the adaptation step, a target user provides a few shots to the model and the model adapts its parameters. After the adaptation process, the model is ready for the target user’s conditions.

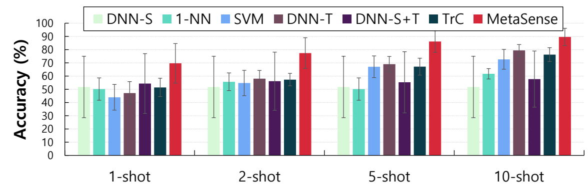


Figure 3: Accuracy for HHAR dataset.

5 RESULT

We use a public dataset, Heterogeneity Human Activity Recognition (HHAR) [3] to evaluate our system’s performance. HHAR dataset was collected with nine users for six human activities. Each user was equipped with eight smartphones around the waist and four smartwatches in the arms and logged accelerometer and gyroscope values for each activity. This dataset has user and device-model dependency but does not include various device positions as each mobile device is located at specific positions. We use the 256-length window with 50% overlapping between two consecutive windows [3]. After eliminating duplicate device models and conditions with less than 10 shots, we have six users and four different devices that result in a total of 24 conditions. We evaluate each 24 conditions with 15 (5×3) conditioned datasets, ensuring no overlap in either the target device or the user. We report the average accuracy of the 24 conditions.

Figures 3 shows the accuracy of the baselines and the proposed approach with the HHAR dataset. We specify the accuracy of 1, 2, 5, and 10-shot cases. DNN-S is using only the source dataset. 1-NN and SVM are shallow classifiers. DNN-T is using only the target dataset. TrC refers to the state-of-the-art transfer learning for activity recognition with a few shots [2]. The result indicates that the effectiveness of our system over other baselines.

ACKNOWLEDGEMENTS

This work was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No.2016R1A2B4014068).

REFERENCES

[1] L. Mirani. There are now more than 24,000 different android devices, Aug 2015.

[2] S. A. Rokni, M. Nourollahi, and H. Ghasemzadeh. Personalized human activity recognition using convolutional neural networks. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[3] A. Stisen, H. Blunck, S. Bhattacharya, T. S. Prentow, M. B. Kjærgaard, A. Dey, T. Sonne, and M. M. Jensen. Smart devices are different: Assessing and mitigating-mobile sensing heterogeneities for activity recognition. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*, pages 127–140. ACM, 2015.

[4] G. M. Weiss and J. W. Lockhart. The impact of personalization on smartphone-based activity recognition. In *AAAI Workshop on Activity Context Representation: Techniques and Languages*, pages 98–104, 2012.