

Unsupervised Domain Adaptation for Robust Sensory Systems

Akhil Mathur

Nokia Bell Labs and UCL
Cambridge, United Kingdom
akhil.mathur@nokia-bell-labs.com

Anton Isopoussu

Nokia Bell Labs
Cambridge, United Kingdom
anton.isopoussu@nokia-bell-labs.com

Nadia Berthouze

University College London
London, United Kingdom
n.berthouze@ucl.ac.uk

Nicholas D. Lane

University of Oxford
Oxford, United Kingdom
nicholas.lane@cs.ox.ac.uk

Fahim Kawsar

Nokia Bell Labs and TU Delft
Cambridge, United Kingdom
fahim.kawsar@nokia-bell-labs.com

ABSTRACT

Despite significant advances in the performance of sensory inference models, their poor robustness to changing environmental conditions and hardware remains a major hurdle for widespread adoption. In this paper, we introduce the concept of unsupervised domain adaptation which is a technique to adapt sensory inference models to new domains only using unlabeled data from the target domain. We present two case-studies to motivate the problem and highlight some of our recent work in this space. Finally, we discuss the core challenges in this space that can trigger further ubicomp research on this topic.

KEYWORDS

domain adaptation; neural networks; audio sensing; activity recognition; unsupervised learning

ACM Reference Format:

Akhil Mathur, Anton Isopoussu, Nadia Berthouze, Nicholas D. Lane, and Fahim Kawsar. 2019. Unsupervised Domain Adaptation for Robust Sensory Systems. In *Proceedings of ACM Conference (Conference'17)*. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 INTRODUCTION

Driven by the advancements in computational models that infer user context from sensor data, a number of consumer

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Conference'17, July 2017, Washington, DC, USA

© 2019 Association for Computing Machinery.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...\$15.00

<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

devices and mobile sensing applications have been developed for various sensing modalities. For example, commercial smart devices have been launched for monitoring a users sleep [6], physical activity [2], dietary actions [4], stress [5], daily activities [2], ambient environments [1, 3], and emotional well-being [5]. In addition to these dedicated consumer devices, a range of mobile sensing applications have been proposed to detect context and activities such as sleep [9], exercise [11], facial expressions [10], transportation mode [8] and emotions [15].

The training of sensory inference models follows a supervised learning approach wherein a labeled training dataset is used to model the relationship between target classes and the sensor data. A fundamental assumption that drives the performance of supervised learning algorithms is that the data distribution remains the same during training and testing stages. If this assumption is violated, the accuracy of supervised classifiers is likely to degrade. Indeed, prior research[7, 8, 16] has shown that differences in sensor devices and operating conditions can introduce unexpected variability in the sensor data. These real-world diversities lead to a discrepancy (or shift) between the training and test data distributions, a phenomenon known as *domain shift* in machine learning. Therefore, if a classifier is trained on data collected from one domain (e.g., device A) and tested on a different unrelated domain (device B), it is likely to perform poorly. An easy solution to this problem is to train a model separately for each domain (e.g., a model per device), however this requires collecting large amounts of labeled training data for each domain, which is both expensive and time-consuming. Ideally, we desire a solution which can generalize a sensor classifier to a new domain using zero or minimal amount of labeled data.

To this end, *unsupervised domain adaptation* (UDA) has emerged as a promising technique to adapt inference models across domains using only unlabeled data from the target

domain. At a high-level, the idea behind UDA is as follows: given a pre-trained classifier for a source domain and an unlabeled dataset from a target domain, how can we adapt the weights of the source model such that it shows better performance in the target domain.

In this paper, we briefly present two real-world sensing scenarios where domain shift caused by operating conditions is prevalent. We then present our recent work on designing and applying unsupervised domain adaptation algorithms to improve the inference accuracy in these scenarios. Finally, we identify the core challenges and open research questions in this space which could pave the way for future research.

2 SCENARIOS

We proceed to discuss two examples of domain shift in mobile sensing. In the first, we look at how audio sensing models cope with perturbations in the audio signal caused by the change of microphone hardware. In the second, we look at the variation caused by different body placements of inertial sensors and how they impact the accuracy of human activity recognition (HAR) models.

Microphone Variability in Speech Models

Speech inference models, once trained, are deployed on numerous mobile and wearable devices, many of which are not known while the models are trained and could come from different hardware manufacturers. This is a challenging scenario because different manufacturers use different hardware components (i.e., microphones) and may also have different software pipelines which process the raw audio signal before exposing them to user applications. Therefore, inference models need to be robust against these forms of microphone heterogeneity expected in the wild.

In our previous work, we quantified the accuracy loss caused by microphone heterogeneity in mobile and embedded microphones [12, 13] in the context of a keyword detection model. We recorded a large spoken keyword dataset simultaneously on three embedded microphones (namely Matrix Voice, ReSpeaker, PlugUSB) and trained microphone specific CNN models for performing keyword detection. Once the models were trained, they were deployed on test microphones different from the training microphones. As shown in Figure 1, we observe that there is a significant drop (red bars) in inference accuracy when there is a mismatch between training and test microphones. For example, when the model trained on Matrix Voice is deployed on ReSpeaker and PlugUSB microphones, there is an absolute accuracy drop of 12.4% and 6.7% respectively when compared with the accuracy upper bound (78.8%).

We propose a CycleGAN-based domain adaptation technique, named Mic2Mic [12], which uses unlabeled and unpaired data from the two microphones to learn a mapping

or domain translation function between them. Once trained, this translation function is then used as a component in the audio inference pipeline to reduce the domain shift between microphones. Our results show that Mic2Mic is able to recover up to 87% of the lost accuracy in the best-case scenario and outperforms a number of calibration and speech enhancement (SE) baselines.

Body Position Variability of Inertial Sensors

Due to the various form-factors and individual wearing preferences, mobile and wearable devices can be worn or carried by users in diverse ways, for example, a smartphone can be carried in pocket or held in hand. This wearing diversity poses a major challenge for human activity recognition (HAR) models in that the models need to be robust against the significant variabilities in motion data obtained from the diverse body positions. However as shown in Figure 2a, the data distribution across body positions show significant divergence, thereby causing the accuracy of a classifier trained on the source body position to degrade in the target domain. Figure 2b shows that this accuracy loss can be as high as 80.7% when the classifier is trained on data collected from thigh and tested on motion data from forearm.

In our ongoing work, we are designing domain adaptation techniques which align the feature spaces of different body positions by learning features which minimise a distance metric between the distributions. The distance are measured using integral probability metrics, such as the maximum mean discrepancy, or via coupling metrics such as the Wasserstein distance. Once a suitable alignment of feature spaces is done, using only unlabeled data, we can expect that a classifier trained on one body position will generalize more effectively to a new body position.

3 OUTLOOK

Our past and ongoing work shows that unsupervised domain adaptation carries significant promise for improving the robustness of sensory inference models. In this section, we provide a discussion on the research challenges involved in scaling, evaluating and generalizability of the domain adaptation techniques, and identify potential research directions in this space.

Extending UDA to Multimodal Sensing

Our current experiments are focused on applying UDA to inference models built using a single modality (e.g., audio or motion). However as multimodal sensing becomes more prominent, it is imperative that we study approaches of extending UDA to them. Indeed, a naive approach is to perform UDA on each modality separately and then fuse their features into the model, however there is also scope for performing UDA on the joint distribution of multiple modalities.

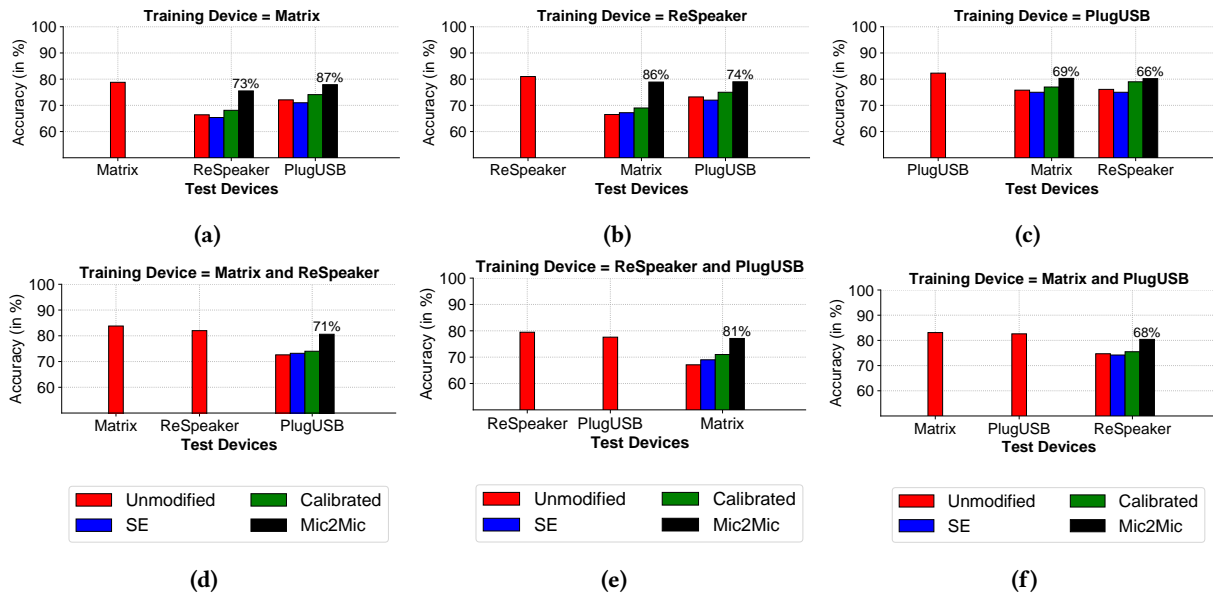
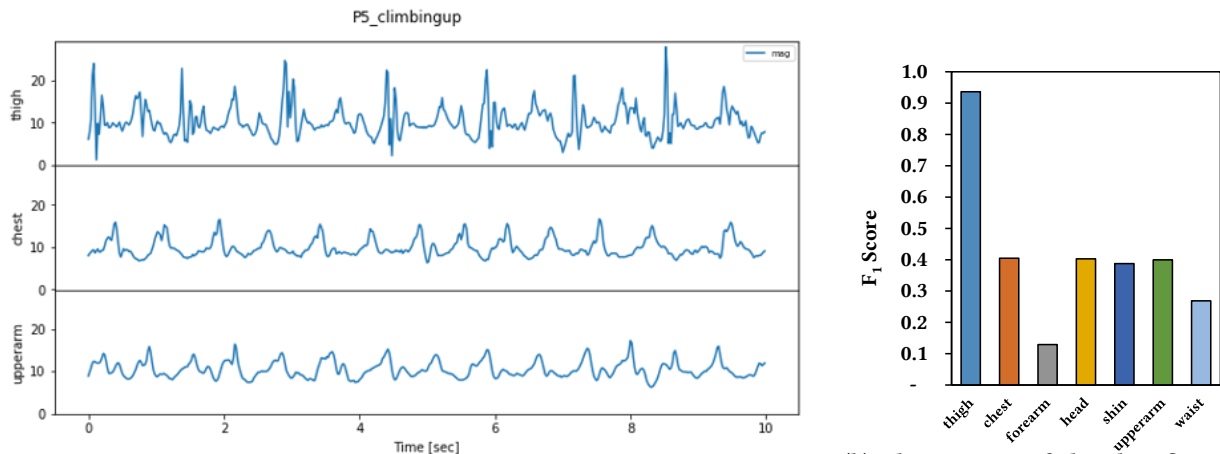


Figure 1: Accuracy of the Keyword Detection model under different scenarios of microphone variability. The numbers on the bars denote the percentage of lost accuracy recovered using Mic2Mic.



(a) Accelerometer traces collected from a trouser pocket, a chest pocket and an arm band for the same physical activity. Variations across body positions are clear and significant.

(b) The F_1 score of the classification model trained on *thigh* when tested on other body positions.

Figure 2: The wearing variability is typical in many wearable devices. Such variability result in huge variance in collected accelerometer signals (Figure 2a). Human activity recognition models that do not consider wearing variability may suffer from significant accuracy drops (Figure 2b).

Evaluation Metrics of Domain Adaptation

Depending on the application, the following three desiderata are often required and used as evaluation criteria.

- *Adaptability* refers to the performance associated with adapting models from one domain to another. This is an important metric when models need to be adapted between static domains – for instance, if a source model

was trained on wrist-worn sensors and subsequently need to be deployed for chest-worn sensors, we desire a high performance in wrist→chest adaptation. More specifically, *Adaptability* can be computed based on the test accuracy of the adapted model in the target domain. The higher the test accuracy in the target domain, the higher the adaptability of the domain adaptation algorithm.

- *Persistence* measures the performance of an inference model on the original source domain after it undergoes the domain adaptation process. It is particularly important because we would like the model to retain its performance in the source domain along with becoming better at doing inferences in the target domain.
- *Generalizability* measures how well a model performs on data domains for which it was neither trained nor adapted. This property is critical because in real-world scenarios, domains such as wearing positions do not remain static – i.e., users can wear their device in unexpected positions and orientations. Therefore, a technique which produces more generalizable models is desirable. Moreover, generalizability may obviate the need for performing domain adaptation on each new domain when scaling to a very large number of domain.

An ideal domain adaptation method satisfying all three properties would be flexible and depending on application, may allow the simplest system design.

Optimizing Training and Inference Pipeline for UDA

A particularly interesting area of research is to make the training and inference pipelines for UDA better aware of user preferences and resource requirements on mobile devices. For example, in our current implementation, the domain adaptation is performed on the cloud and requires end-users to upload their private sensor data to the cloud. Future research should look at privacy-preserving approaches which can perform domain adaptation using no or minimal raw sensor data from the users. Similarly, the adapted models need to be optimized for inference on mobile and wearable devices so that there is minimal latency overhead caused by applying domain adaptation in the inference pipeline.

Label Space Mismatch

Existing domain adaptation techniques assume that label/class spaces across the source and target domains are identical, even though their underlying data distributions might be different. In the presence of domain shift it is challenging to determine a source domain which has identical labels or classes to the intended target domain. As such, it is likely that there are outlier classes either in the source or target domain. The presence of such outlier classes in the adaptation task can lead to negative transfer [14], a phenomenon wherein the outliers even degrade the transfer of knowledge between shared classes. Therefore, a major challenge for making domain adaptation practical in ubiquitous systems is to design algorithms which can work under label space mismatch.

When to Apply Domain Adaptation

Even though domain adaptation has significant potential in improving the performance of sensory models, it is not a silver bullet that can work in all scenarios. For instance, if the domain shift between the distributions is extremely high, it is likely that domain adaptation cannot recover the lost accuracy. As a topic of future research, we plan to investigate whether feasibility of domain adaptation could be quantified in an empirically robust manner.

4 CONCLUSION

In this paper, we presented two short case studies to motivate the need for sensory model adaptation. We found that variabilities in sensor hardware as well as user behavior can drastically degrade the performance of state of the art methods. Our work argues that unsupervised domain adaptation is a promising technique to address these challenges and we highlighted several research avenues on this topic that could be of benefit for the workshop attendees as well as the ubicomp community in general.

REFERENCES

- [1] 2016. Google Glass. <https://developers.google.com/glass/distribute/glass-at-work>. Accessed: October 26, 2019.
- [2] 2017. FitBit. <https://www.fitbit.com>.
- [3] 2017. Narrative Clip. <http://getnarrative.com/narrative-clip-1>. Accessed: September 1, 2017.
- [4] 2017. New technology tracks food intake by monitoring wrist movements. <http://gadgetsandwearables.com/2017/03/29/food-tracking/>. Accessed: October 26, 2019.
- [5] 2018. Empatica Wristband. <https://www.empatica.com/en-gb/research/e4/>. Accessed: October 26, 2019.
- [6] 2018. Nokia Sleep Tracker. <https://health.nokia.com/uk/en/sleep/>.
- [7] Yohan Chon et al. 2013. Understanding the coverage and scalability of place-centric crowdsensing. In *UbiComp*. ACM, 3–12.
- [8] Blunck et al. 2013. On heterogeneity in mobile sensing applications aiming at representative data collection. In *Proceedings of the 2013 ACM UbiComp*. ACM, 1087–1098.
- [9] Tian Hao, Guoliang Xing, and Gang Zhou. 2013. iSleep: Unobtrusive Sleep Quality Monitoring Using Smartphones (*SenSys '13*). ACM, 14. <https://doi.org/10.1145/2517351.2517359>
- [10] Seungchul Lee, Chulhong Min, Alessandro Montanari, Akhil Mathur, Youngjae Chang, Junehwa Song, and Fahim Kawsar. 2019. Automatic Smile and Frown Recognition with Kinetic Earables. In *Proceedings of the 10th Augmented Human International Conference 2019 (AH2019)*. ACM, New York, NY, USA, Article 25, 4 pages. <https://doi.org/10.1145/3311823.3311869>
- [11] Hong et al. Lu. 2010. The Jigsaw continuous sensing engine for mobile phone applications. In *Proceedings of Sensys '10*. ACM, 71–84.
- [12] Akhil Mathur, Anton Isopoussu, Fahim Kawsar, Nadia Berthouze, and Nicholas D. Lane. 2019. Mic2Mic: Using Cycle-consistent Generative Adversarial Networks to Overcome Microphone Variability in Speech Systems. In *Proceedings of the 18th International Conference on Information Processing in Sensor Networks (IPSN '19)*. ACM, New York, NY, USA, 169–180. <https://doi.org/10.1145/3302506.3310398>

- [13] Akhil Mathur, Tianlin Zhang, Sourav Bhattacharya, Petar Veličković, Leonid Joffe, Nicholas D Lane, Fahim Kawsar, and Pietro Lió. 2018. Using deep data augmentation training to address software and hardware heterogeneities in wearable and smartphone sensing devices. In *Proceedings of the 17th ACM/IEEE International Conference on Information Processing in Sensor Networks*. IEEE Press, 200–211.
- [14] Sinno Jialin Pan and Qiang Yang. 2010. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* 22, 10 (2010), 1345–1359.
- [15] Kiran Rachuri et al. 2010. EmotionSense: a mobile phones based adaptive platform for experimental social psychology research. In *Proceedings of Ubicomp '10*. ACM, 281–290.
- [16] Siqi Yang, Arnold Wiliem, and Brian C Lovell. 2016. To face or not to face: Towards reducing false positive of face detection. In *Image and Vision Computing New Zealand (IVCNZ), 2016 International Conference on*. IEEE, 1–6.