Inter-University Research Institute Corporation
Research Organization of Information and Systems
National Institute of Informatics

TOWARDS DATA DRIFT MONITORING FOR SPEECH DEEPFAKE DETECTION IN THE CONTEXT OF MLOPS

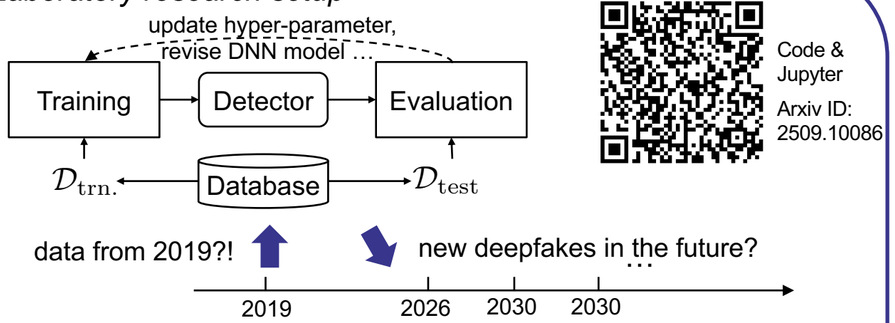
Xin Wang, Wanying Ge, Junichi Yamagishi

SLP P28.3

Three-minute summary

- Motivations
  - Laboratory research setup assumes a fixed database(s)
  - In applications, new & evolving data may **drift away** from the fixed training & test data
- Research questions
  - Can we monitor the data drift?
  - Can we reduce the data drift by fine-tuning on new data?
- Take-home messages
  - Distance in embedding space helps to monitor the drift
  - Fine-tuning helps to reduce the drift if data is representative

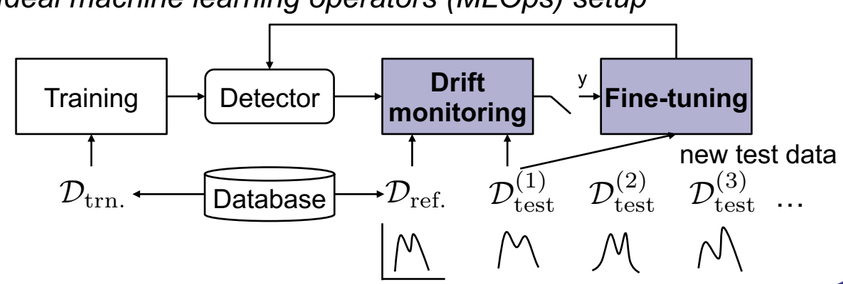
Laboratory research setup



Code & Jupyter

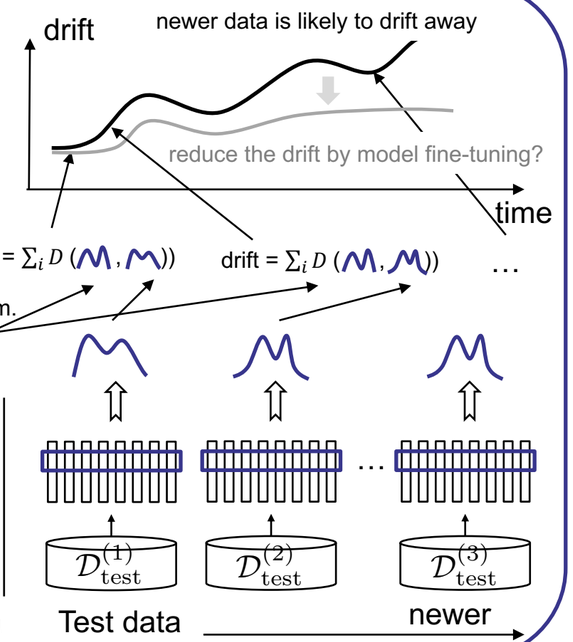
Arxiv ID: 2509.10086

Ideal machine learning operators (MLOps) setup



Core method: drift detection

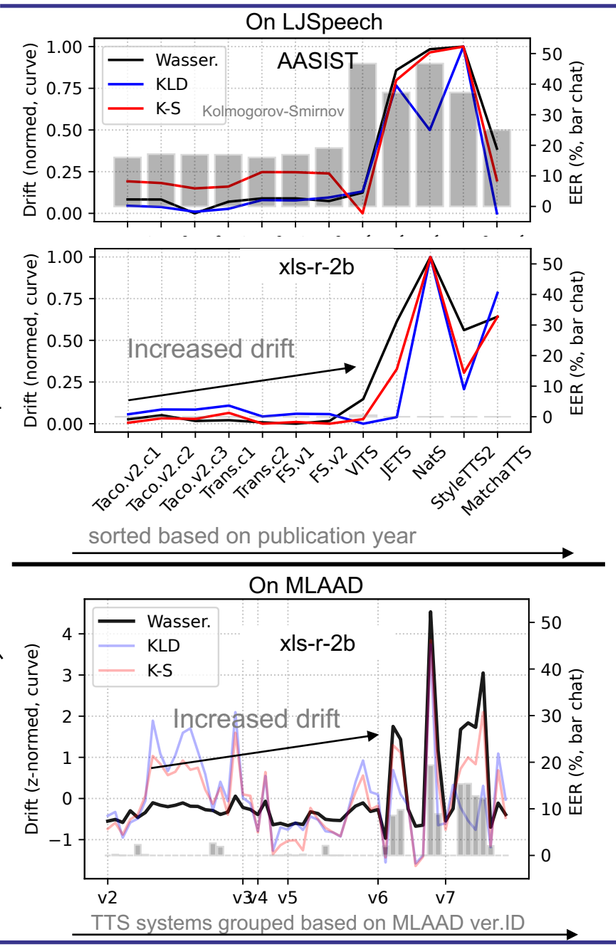
- Drift = distance between dist. of embedding features
  - embedding after global average pooling
  - independence across dim.
    sum over dimensions
- Distance metric *D*
  - Wasserstein-1
  - KLD …



Experiment 1: can we monitor the drift?

- We only monitor fake data in this work
- Models & their training data self-supervised learning
  - **`AASIST`**: trained on ASVspoof 2019 LA, non-SSL
  - **`xls-r-2b:`** AntiDeepfake checkpoints, SSL front-end+pooling+linear layer
- Reference data: ASVspoof 2019 LA dev.
- Test data text-to-speech
  - `LJSpeech`: 12 TTS systems in LJspeech voice
    autoregressive (tacotron), diffusion (MatchaTTS) …
  - `MLAAD eng.`: 54 TTS systems that mimic the target speakers, grouped as v2, v3, …, v7
    v2: Tacotron, fast-pitch … ; v7: Llasa-3B, Chatterbox …
    – 5 hours randomly sampled from each MLAAD version
- Observations
  - More drift on more advanced TTS
  - Similar trend on another SSL model (see paper)
  - EER correlates with drift to varied degrees

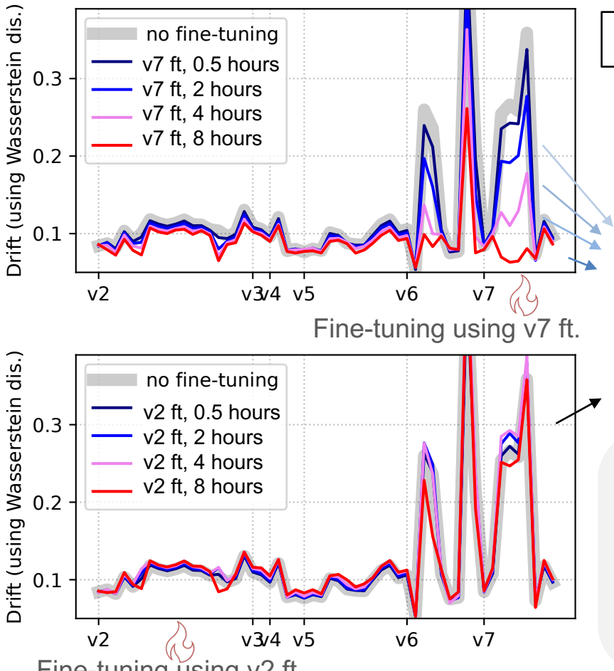
  Bona fide data is fixed.
  No drift is measured on bona fide data

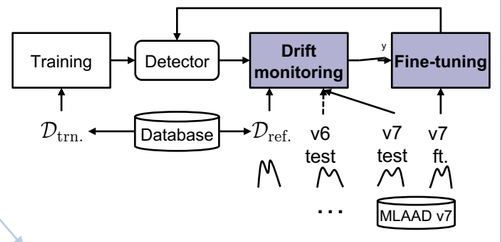


Experiment 2: can fine-tuning reduce the drift?

- For MLOps, suppose that we can annotate newly incoming data and use it to fine-tune the model; we hope it works on similar data

  fine-tuning (ft.) data: MLAAD eng. ← disjoint utterances → test data: MLAAD eng.
- Model: `xls-r-2b`; Wasser. distance; ref. & test data: as Exp.1





More fine-tuning data, less drift

Fine-tuning on v7 also reduces drift caused by "older" data (v6)

Fine-tuning on older data DOES NOT reduce the drift on "newer" data

✓RQ1: drift can be monitored using embedding features

✓RQ2: drift can be reduced if the data is representative

* Fine-tuning reduced EERs (see paper)