DISTRIBUTED ONE-CLASS LEARNING

Ali Shahin Shamsabadi[1], Hamed Haddadi[2], Andrea Cavallaro[1]

[1]Queen Mary University of London, UK, [2]Imperial College London

1. Introduction

Objective: Training a filter in the service provider cloud on users data with users collaboration to preserve privacy against

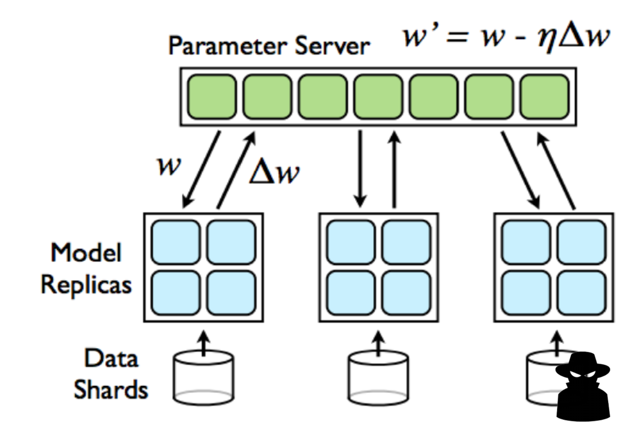
- Malicious users
- Malicious service provider

Challenges

- Sensitive information in users data
- Parameters of filter memorise training data

2. Related work

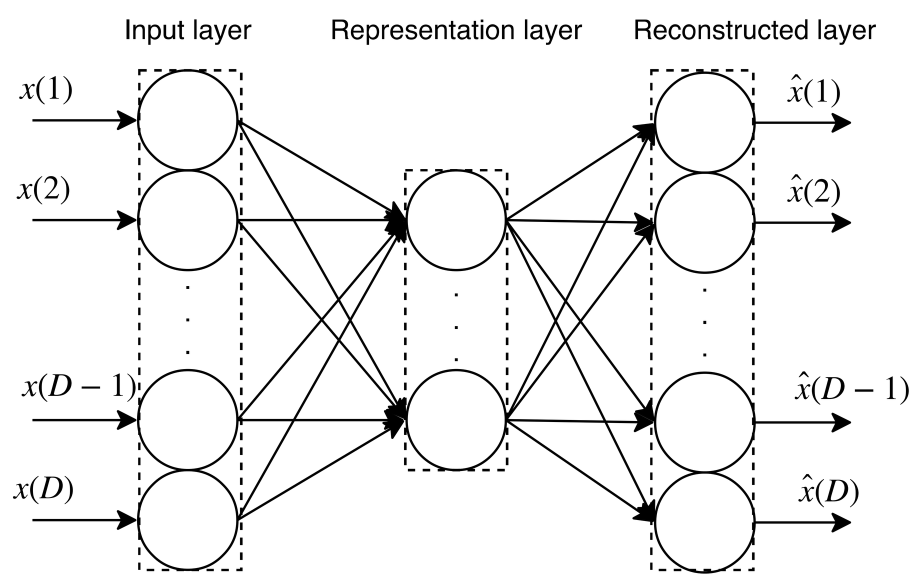
- Centralised learning
  - Users upload their data to the cloud
  - Service provider train the filter
  - **Limitation**
    - Service provider has access to the user data
- Distributed learning [1,2]
  - Users train a local copy of the filter on their devices
  - Users upload only the parameters of their filter Δw
  - Service provider fine-tunes the filter $w' = w - \eta\Delta w$
  - **Limitations**
    - Parameters shared among the service provider and users
    - Each user should have access to data of several classes



3. Proposed approach: Distributed One-Class Learning (DOCL)

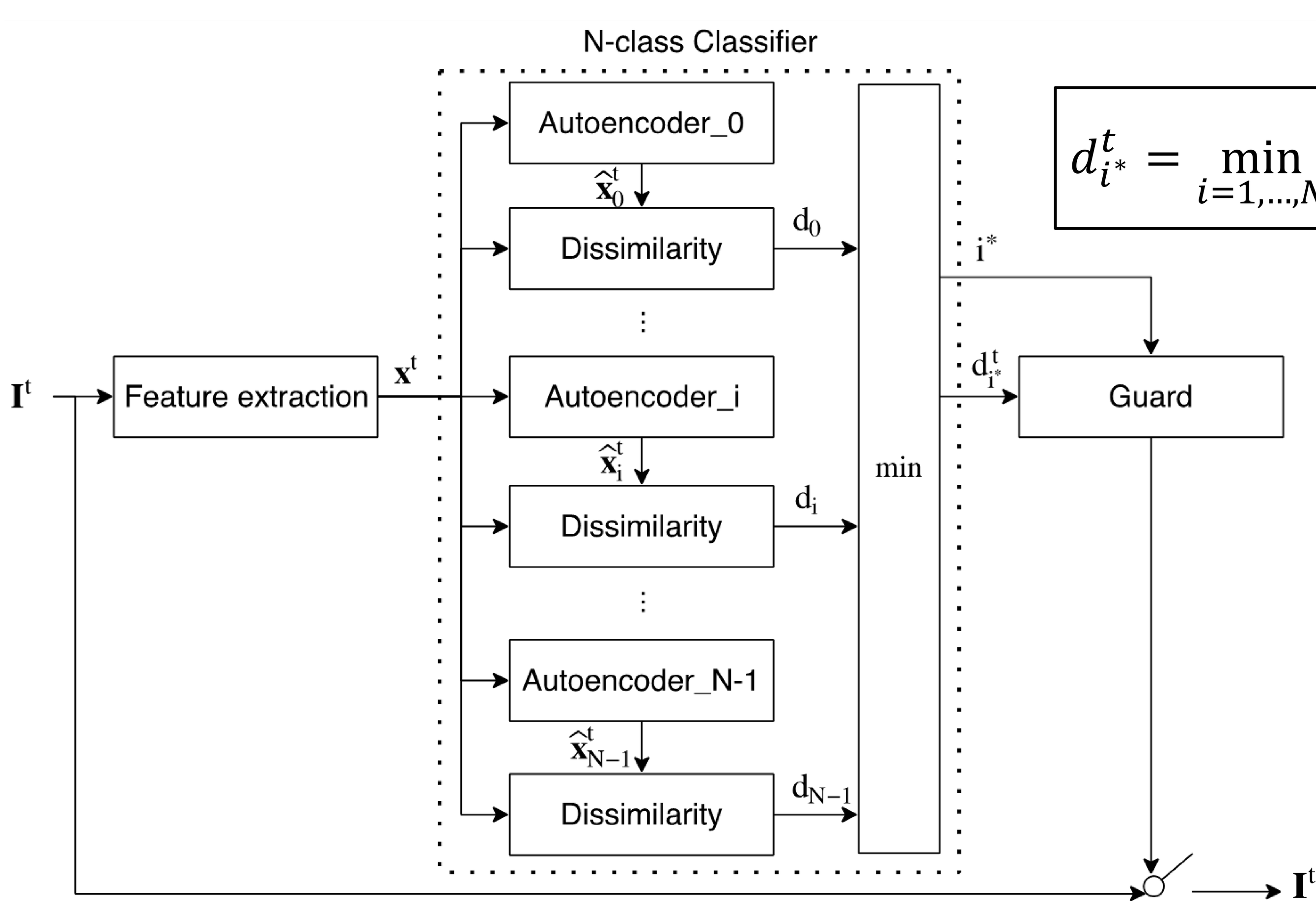
- Assume N users and one service provider
- Training a centralised N-class classifier (filter) without sending the training data and addressing malicious users
- Decompose the global filter to N one-class classifiers
- Distribute N one-class classifiers among N users
- Each user locally trains a one-class autoencoder on their private data independent of other users

input image $x \in \mathbb{R}^D$



reconstructed image $\hat{x} \in \mathbb{R}^D$

- Users upload the parameters of their one-class classifiers to the service provider
- Service provider
  - Aggregates all the N one-class classifiers to discriminate between classes
  - Checks the legitimacy of each uploaded image I^t by user u_{i^*} prior to sharing that image

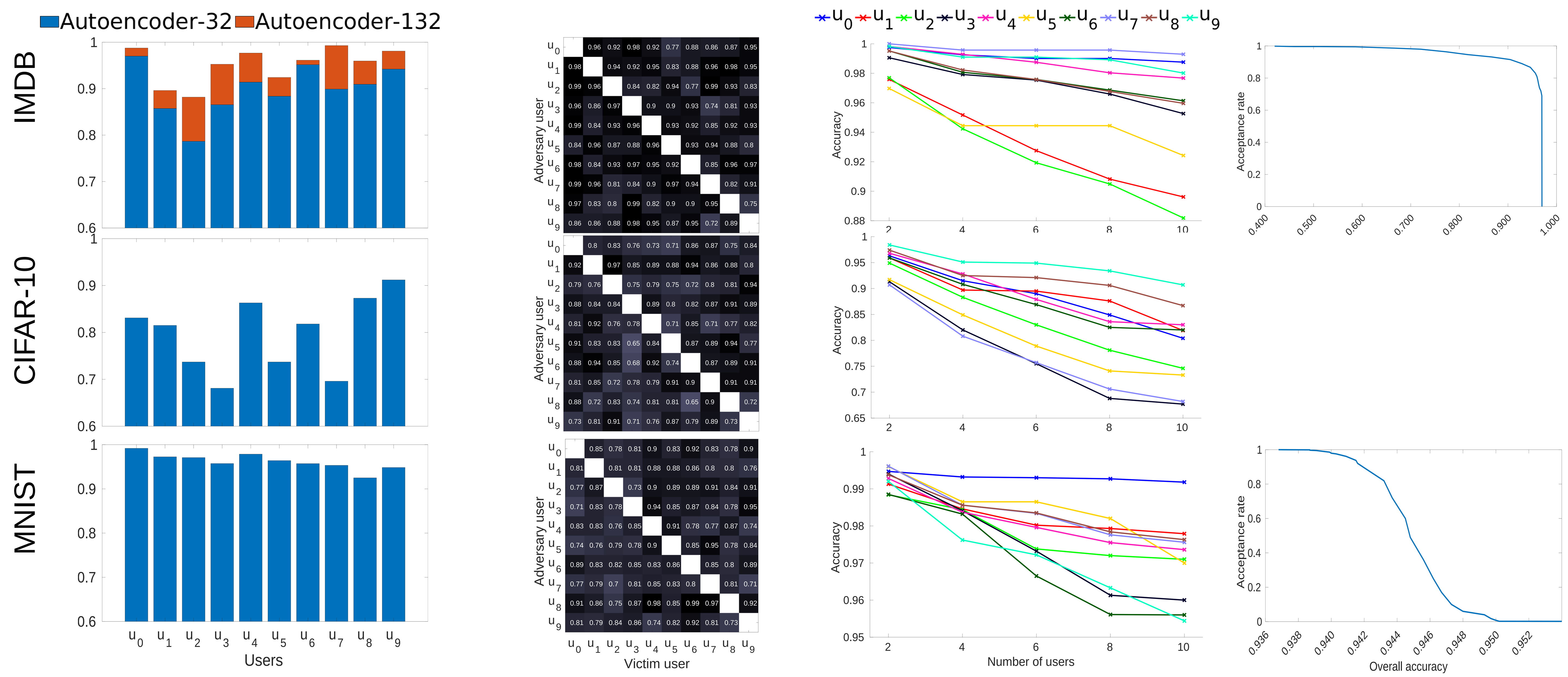


$$d^t_{i^*} = \min_{i=1,\dots,N} \sum_{l=1}^{D} \left\| x^t(l) - \hat{x}^t_i(l) \right\|_2^2$$

4. Experimental results

Setting

- Extracts features of IMDB and CIFAR-10 with ResNet
- 10 users $u_i, i = 0, \dots, 9$



Number of correctly predicted labels

Per user accuracy $A = \frac{n_c}{n_t}$

Total number of images of each user

Preserve privacy against malicious user who has full knowledge of training data of victim user

Negligible effect on per user accuracy by increasing the number of users

Two scenarios for non-private images:

- Substantially different from private images
- Similar to private images

5. Conclusions

- The proposed filter outperforms on MNIST and IMDB than on CIFAR-10, which has a high inter-class variability
- The more similar the images in one class, the smaller the decrease in per-class and overall accuracy when the number of classes increases
- A new user can join at any time by training a new one-class classifier

References

[1] R. Shokri and V. Shmatikov, "Privacy-preserving deep learning", Computer and Communications Security, 2015.

[2] H.B.McMahan, et al., "Communication-efficient learning of deep networks from decentralized data", Artificial Intelligence and Statistics, 2016.