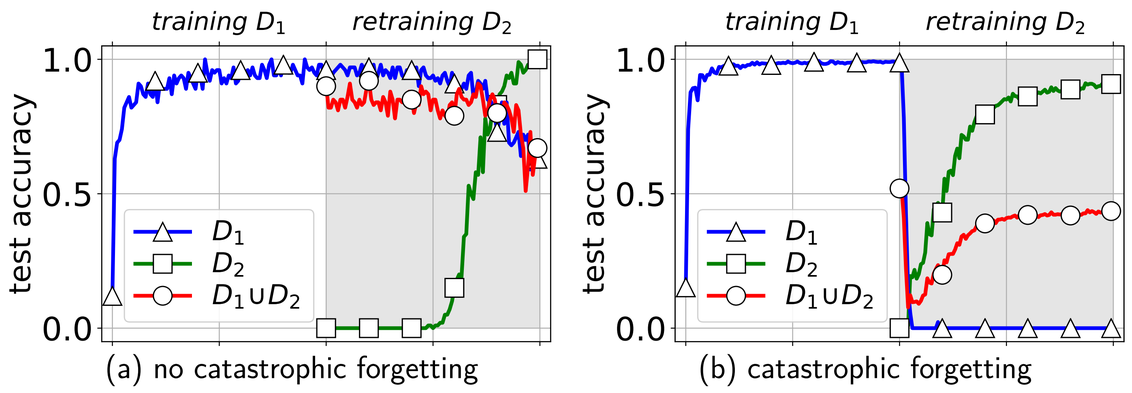
A Comprehensive, Application-Oriented Study of Catastrophic Forgetting in DNNs

Benedikt Pfülb and Alexander Gepperth

{benedikt.pfuelb,alexander.gepperth}@cs.hs-fulda.de, Department of Applied Computer Science, University of Applied Sciences Fulda

1. Problem Definition

DNNs are affected by **catastrophic forgetting** (CF), which leads to an abrupt loss of knowledge after a few training iterations. Sequential Learning Tasks (SLTs) consisting of multiple (here: two) sub-tasks are particularly affected by CF.



(a) no catastrophic forgetting (b) catastrophic forgetting

2. Investigated Approaches

This study evaluates 7 models, some of which are purposely designed to avoid CF.

- fully-connected (FC) DNN (optional with Dropout (D-FC))
- convolutional (CONV) DNN (optional with Dropout (D-CONV))
- the Elastic Weight Consolidation (**EWC**) model
- FC model with Local Winner Takes All (**LWTA**) transfer function
- the Incremental Moment Matching (**IMM**) model

3. Investigated Datasets

To exclude the feasability of CF due to the difficulty of the problem all models are evaluated on 9 different image datasets (10 classes).

- MNIST
- EMNIST
- Fruits 360
- Devanagari
- FashionMNIST
- SVHN
- CIFAR10
- NotMNIST
- MADBase

4. Experimental Setup

1. train model on sub-task D_1, e.g., class 0-4
2. fix hyper-parameters based on the accuracy on D_1
3. retrain model on sub-task D_2, e.g., class 5-9
4. concurrently: evaluate model on D_1, D_2 and $D_1 \cup D_2$

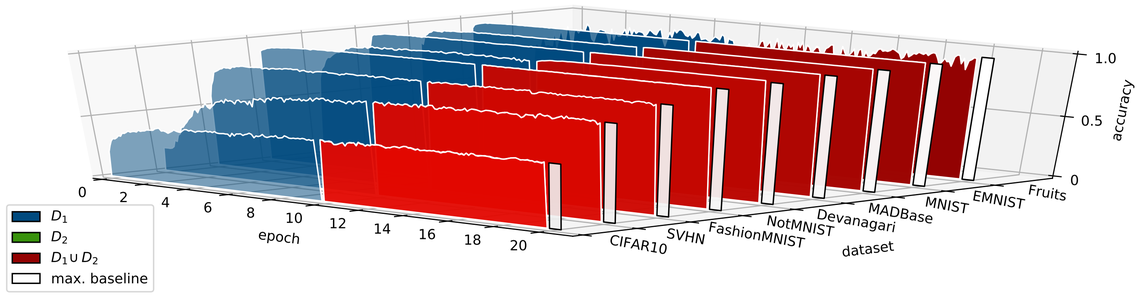
5. Contributions

- Evaluation that respects real-world application constraints:
  - only data of current sub-task available (no look-ahead)
  - model selection only on 1st sub-task
  - origin of the data (w.r.t. sub-task) unknown
  - constant complexity of learning w.r.t. sub-tasks

⚠ Constraints often neglected in previous studies!

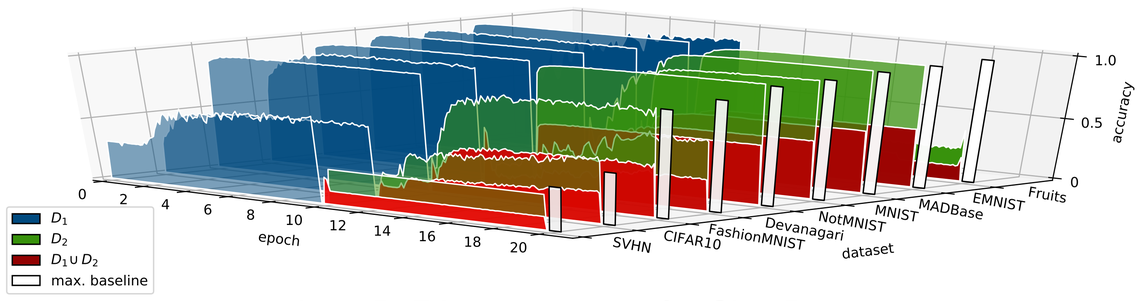
6. Findings

- Permutation-based SLTs shouldn't be used to investigate CF



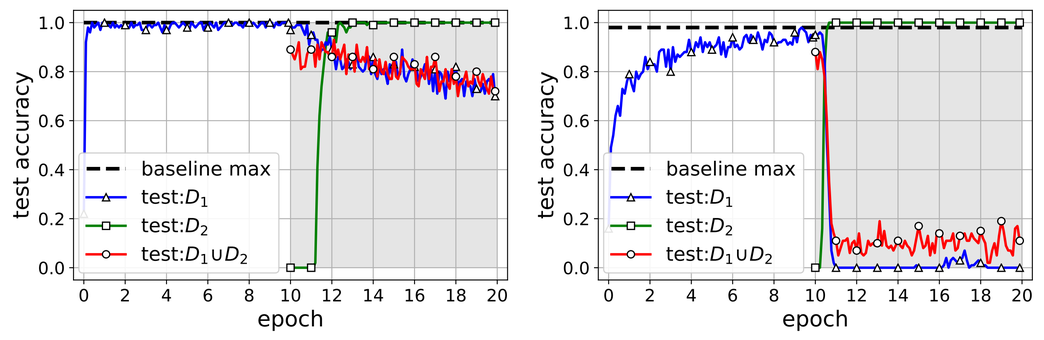
best FC experiments for a permuted SLT

- All examined models exhibit CF (expected for FC models)



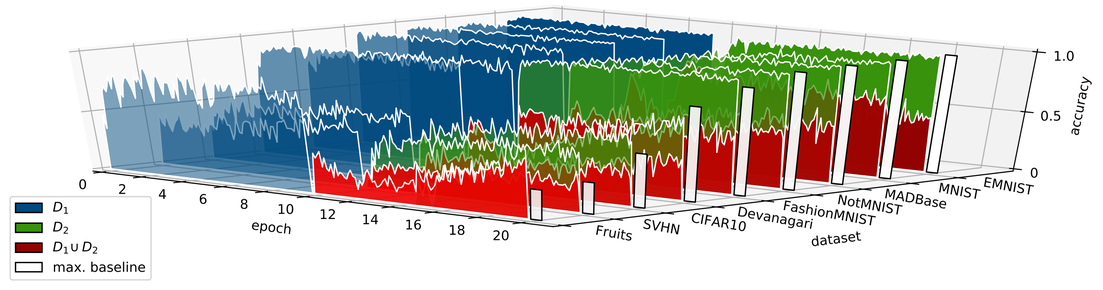
best D-FC experiments for SLT D5-5

- EWC is mildly effective against CF for simple SLTs, but future knowledge is required to tune the λ parameter of EWC
  - for D9-1 tasks it depends on the difficulty of the dataset



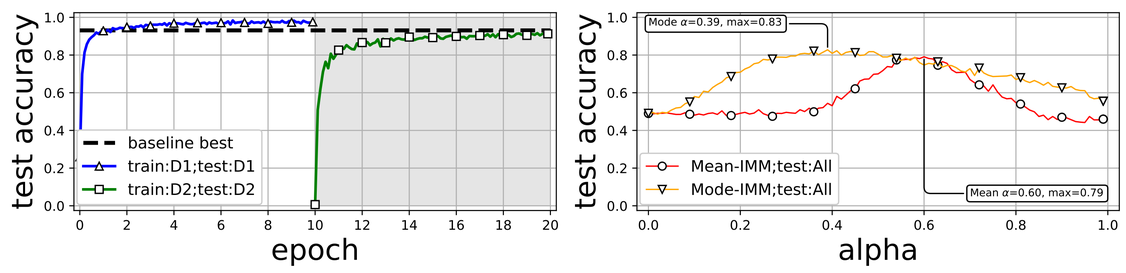
(a) flat linear forgetting (b) catastrophic forgetting

  - for D5-5 tasks, EWC is not effective



accuracy of D5-5 SLT for best EWC experiments

- IMM is effective against CF but requires knowledge of the past to find a good balancing parameter



best IMM experiment for SLT D5-5