**Summary:**

This paper is about the forgetting problem in continual learning. Their observations suggest that the knowledge outside of local distribution is prone to forgetting, so a simple and novel algorithm is proposed to this problem. Their experiments compare with famous federated learning methods.

**Strengths And Weaknesses:**

Even though the proposed method, seems simple yet sweet. It looks like it works well. The idea could easily be used. A “forgetting measure” is defined to illustrate the behavior and to demonstrate it’s effectiveness.

FL performance is decided by balancing the learning of the new knowledge on p(D) and preserving the old knowledge on p~(D).

In the experimental setup, various different setting and senarios are taken in consideration.

**Limitations:**

More clarity could be provided on what the “p(D) and p~(D) accuracy” means.

I didn’t really understand the definition of Federated Non-True Distillation in 3, a bit easier explanation of the the same could have been provided.