

Evaluating Lottery Tickets Under Distributional Shifts

Shrey Desai¹ Hongyuan Zhan² Ahmed Aly²

¹The University of Texas at Austin ²Facebook Assistant



@shreydesai

Lottery Tickets

Introduction

> Compression algorithm that finds **small, sparse subnetworks** within **large, overparameterized networks**

> Works using four simple steps:

1 **Initialize** a neural network with parameters drawn from a random distribution $f(x; \theta_0)$
 $\theta_0 \sim \mathcal{N}(0, 1)$

2 **Train** the network to convergence (e.g., several epochs or iterations) $\text{train } f(x; \theta_0) \rightarrow \theta_n$

3 **Prune** the network using a heuristic to *permanently* remove the least important weights (e.g., least magnitude) $m \in \{0, 1\}^{|\theta|}$

4 **Retrain** the pruned network using the *original* parameters before training $f(x; \theta_0 \odot m)$
 $\text{train } f(x; \theta_0 \odot m)$

Related Work

Frankle and Carbin (2019)

Lottery tickets exist! Resetting network weights to their initial values is crucial.

Actually, initial values are questionable. Randomly initialized tickets still perform well if the learning rate is tuned.

Liu et al. (2019)

Morcos et al. (2019)

Lottery tickets obtained on ImageNet transferred to other datasets. Initial values seem to be important?

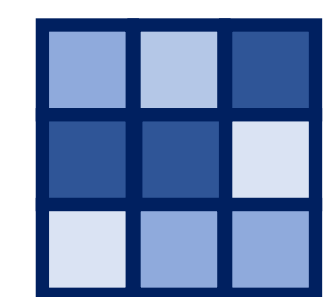
Our Work

> How well do lottery tickets generalize?



Figure 1: If a lottery ticket was obtained using New York Times samples, then how would it perform when trained with Twitter samples?

> If so, what is *required* for generalization?



initial params



mask

Figure 2: How important is it to reset models to their initial values? Can randomly initialized tickets generalize as well?

Methodology

Cross-Domain Setup

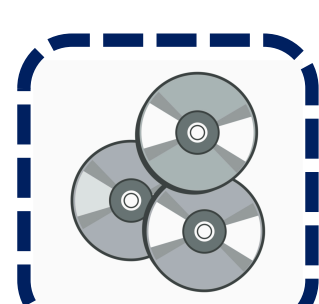
> In order to test generalizability, we must introduce **domain shifts**; we use multiple domains from the **Amazon Reviews Dataset** to set up a binary sentiment analysis task



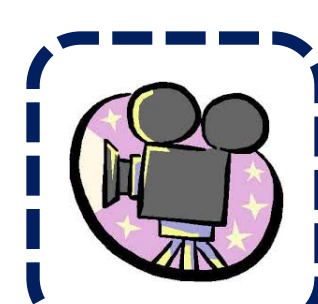
Books



Electronics



CDs



Movies

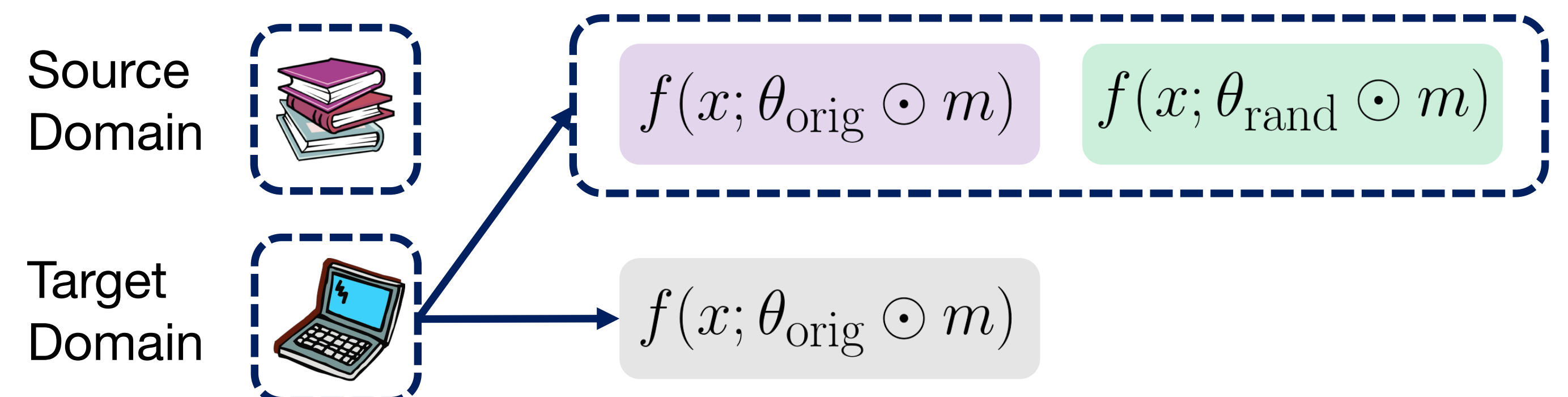


Home

20K Train
10K Validation
10K Test

Testing Generalizability

> Obtain a **source domain** lottery ticket and **target domain** lottery ticket; then, train variants of the **source domain** lottery ticket using **target domain** samples



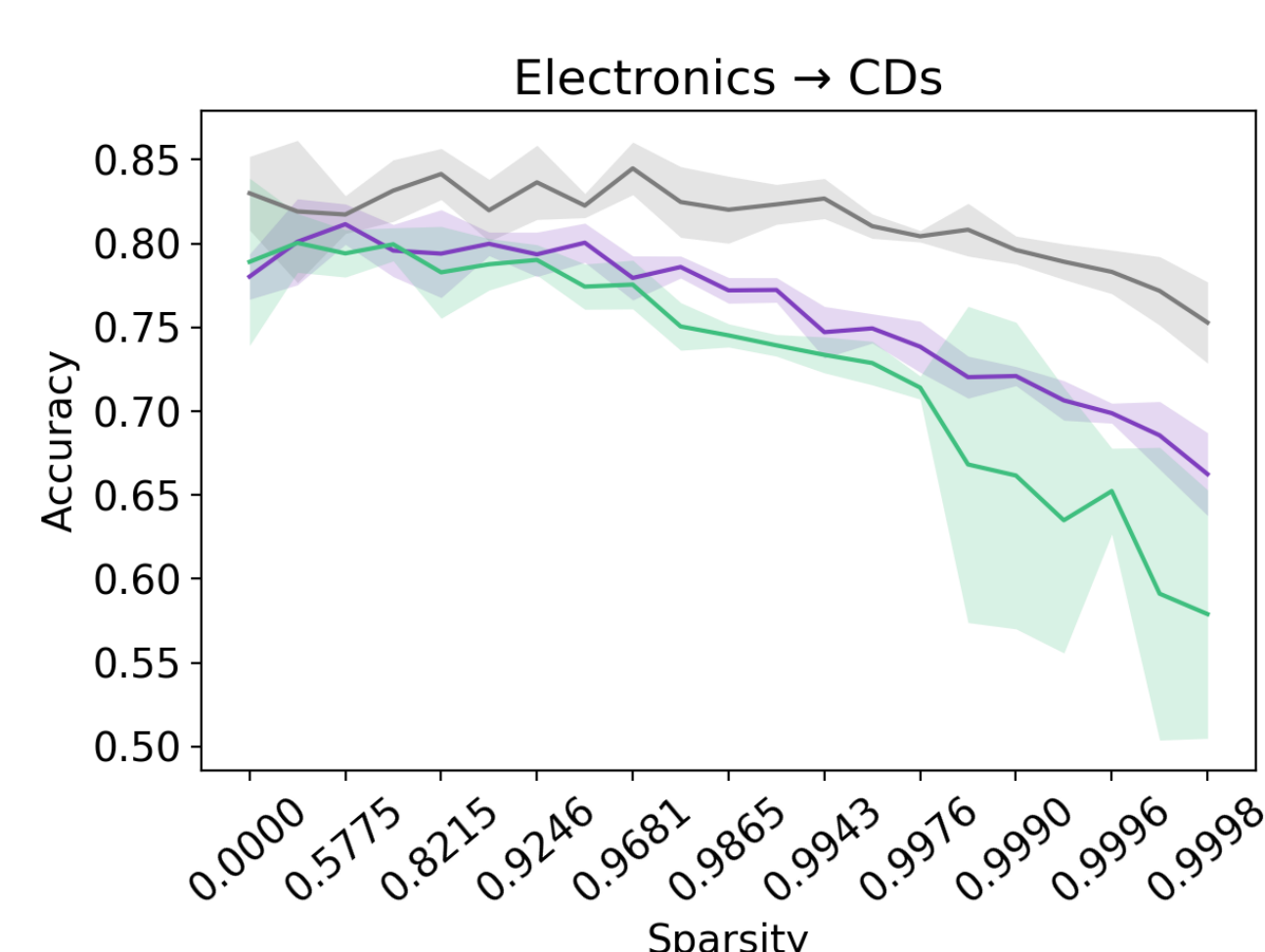
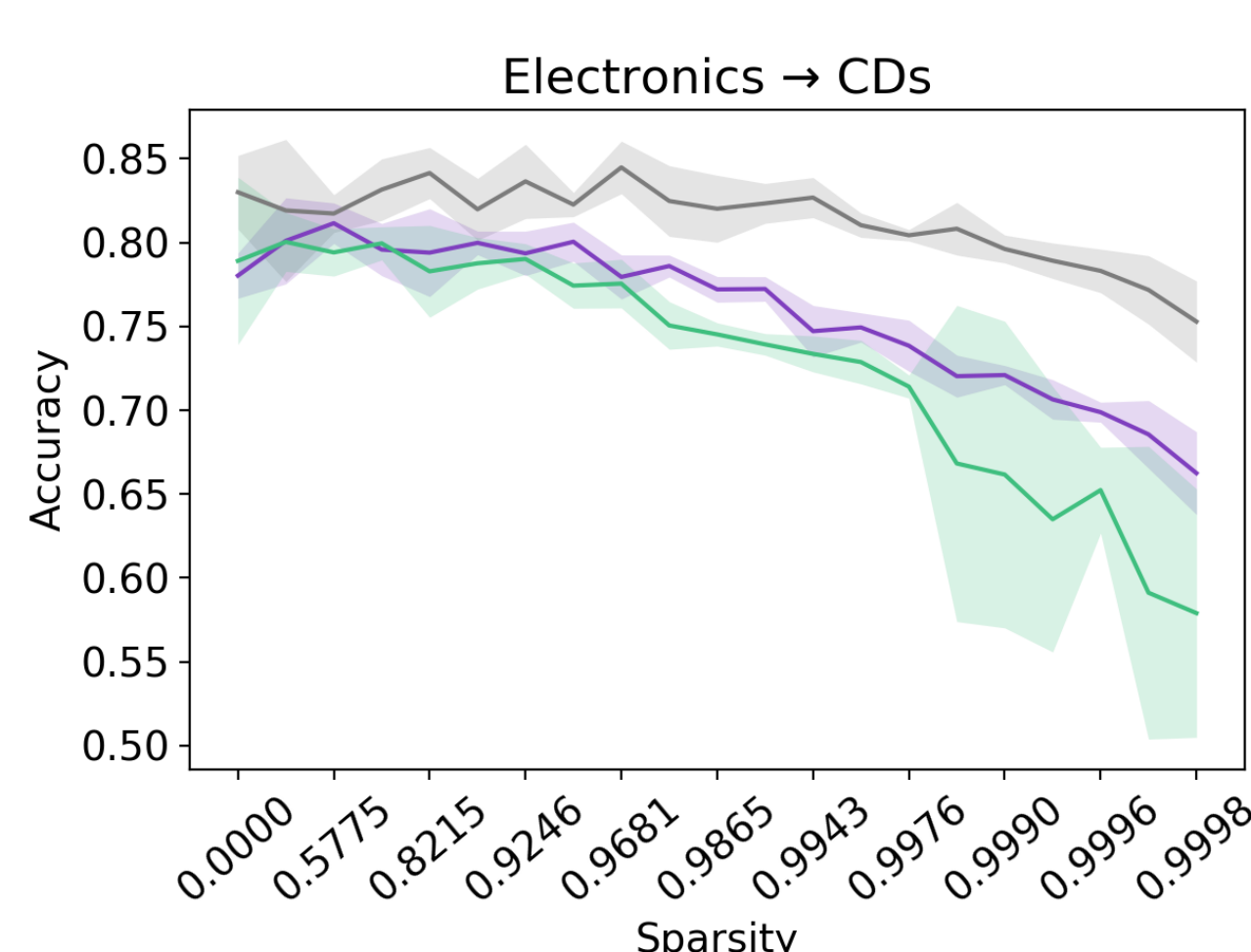
Experiments & Results

Source LT (original init)

Source LT (random init)

Target LT (original init)

> We find **compelling evidence** of source lottery ticket generalization ability. In addition, there is a **phase transition point** (~99.5%) where initialization matters.



> Sometimes, source lottery tickets perform **on-par or even better** than the target domain tickets. Future work can explore what makes source tickets **more effective** in a target domain.

