Reviewer #258421:

We are grateful for the reviewer's constructive comments. The followings are our rebuttal for your reviews.

1. **"… Technical depth …"**

   Our paper is inspired by [Santurkar et al., 2018, ICML18] which is an in-depth study of the mode collapsing problem in GANs. We propose an elegant solution to address data distortion, missing modes, and boundary distortion mentioned in that paper. We obtain convincing experimental results. Finally, our application is novel and is used in a new important domain wherein machine learning, deep learning and domain adaptation can be applied. For these reasons, we strongly believe in the quality of our work from both a technical and application perspective.

   Among three papers listed, the first paper [1] is directly relevant to our problem of domain adaptation for sequential data. It is an oversight on our part and we thank the referee for pointing this out. We will include this discussion in our final version of the paper. However, the work in [1] is based on the Variational RNN and the adversarial training (i.e., GAN as in Ganin and Lempitsky, 2015]) is applied on the latent representation z^i. It is likely that the mode collapsing problem could still happen due to the nature of GAN. In terms of using adversarial training to close the gap of source and target data, our study goes much further than this by proposing additional components and principles to circumvent the mode collapsing problem based on a recent study in [Santurkar et al., 2018].

2. **"… Only align the autoregressive context vector and no explanation …"**

   We employed two generators and discriminators and elegantly designed their operating principles to reduce the negative impact of data distortion, missing modes, and boundary distortion. In our paper, we described these phenomena and provided high-level details of how these components together can solve these problems (See Fig. 3 and Sub-section the rationale for our Dual-discriminator Code Domain Adaptation approach). The underlying principle here is that we ensure D_S and D_T obtain high values using both G_S(x^S) and G_T(x^T). We would be grateful if the reviewer could revisit these relevant parts in our paper.

3. **"… Double the capacity of the discriminator for DCDA …"**

   We do not think this helps because the main task here is of the generator to produce appropriate representations. The two discriminators in our model need to cooperate and are not simply equivalent to a double discriminator.

4. **"… Also, I'm not fully convinced about the existence of the mode collapse in the proposed setting. Do you have any evidence of the existence of the phenomenon in your cases? ..."**

   Mode collapse is an inherent problem of GANs. This is because we do not solve the min-max problem properly. Instead, we alternatively update the generator and discriminator. The discriminator tries to discriminate between the source and target data, while the generator attempts to make them indistinguishable. There is no principle to prevent the generator maps the target data in a small region to a region inside the source data and vice versa (See Fig. 3).

It is quite challenging to visualize and quantify the magnitude of mode collapse since the data are very high-dimensional. To this end, we reuse the experimental setting in Sec 5.2 in [Santurkar et al., 2018] and the experimental results obtained strongly support our claim.

5. "… In my understanding, the dual-discriminator architecture is only meaningful for asymmetric loss function. Why don't you use just another type of divergence well-used in GAN scenarios, such as Wasserstein distance? ..."

According the experiments in [Santurkar et al., 2018, ICML18], WGAN with gradient penalty also suffers from the mode collapsing problem as other GANs. In fact, in the original WGAN with gradient clipping and later WGAN with gradient penalty, the authors only claimed that their WGANs are stable with the loss functions gradually decreasing during training process. Still, the quality of samples was improved along with the reduction of the loss. The authors did not claim WGAN can address the mode collapsing problem better than other GANs.

6. "**The authors says "Notably, the work of [Ganin and Lempitsky, 2015; Tzeng et al., 2015; Shu et al., 2018] employed generative adversarial networks (GAN) [Goodfellow et al., 2014]", but this sentence is wired for my since the none of the methods model the data distribution**"

Most of GAN-based domain adaptation papers uses the GAN principle to ensure the source and target distributions in the joint space are identical. In other words, we need to train the generator in such a way that the induced source and target distributions in the joint space are as close as possible. If so, the domain classifier trained on the source data in the joint space can be nicely transferred to the target domain. Therefore, our sentence "Notably, the work of [Ganin and Lempitsky, 2015; Tzeng et al., 2015, Shu et al., 2018] employed generative adversarial networks (GAN) [Goodfellow et al., 2014] to close the discrepancy gap between source and target data in the joint space" absolutely makes sense.