Yale

Improving Zero and Few-Shot Abstractive Summarization with Intermediate Fine-tuning and Data Augmentation

Alexander R. Fabbri[1],Simeng Han[3], Haoyuan Li[4], Haoran Li[2],

Marjan Ghazvininejad[2], Shafiq Joty[3], Dragomir Radev[1], Yashar Mehdad[2]

Yale University[1] Facebook AI[2] Nanyang Technological University[2] Renmin University of China[1]

LILY Lab

Introduction

Models trained with self-supervised pretraining objectives have achieved state-of-the-art results on text summarization. However, typically such models are fine-tuned on large supervised datasets. Collecting such datasets for new domains is unfeasible, so we analyze the performance of transferring pertained models in zero and few-shot settings. We propose a method, **WikiTransfer**, to improve zero-shot performance which takes advantage of characteristics of the target dataset which are known a priori. We create dataset-specific data for intermediate fine-tuning, which generally improves transfer to that domain. WikiTransfer models achieve state-of-the-art, zero-shot abstractive summarization performance on the CNN-DailyMail dataset and demonstrate the effectiveness of our approach on three additional diverse datasets. These models are more robust to noisy data and also achieve better or comparable few-shot performance using 10 and 100 training examples when compared to few-shot transfer from other summarization datasets. We further study the effect of the components of our unsupervised fine-tuning data and analyze few-shot performance along with data augmentation techniques using both automatic and human evaluation.

Methods

Intermediate Fine-tuning: Assume that we want a summary of M sentences from source documents of N sentences on average

We then iterate the following procedure on all Wikipedia articles available in a Wikipedia dump:

- Remove the first M sentences from the Wikipedia article for use as a summary.
- Select the M sentences in the remaining N article sentences with the highest individual ROUGE scores against the pseudo summary
- Filter examples based on how extractive the pseudo-summaries are

Round-trip Data Augmentation: We translate the input and summaries to and from a non-English language to create additional training data.

Consistency Regularization: We add a consistency regularization term to ensure that the model is not affected by perturbations in the round-trip augmentation data.

Target Dataset	WikiTransfer	Other Transfer
CNNDM	**39.11/17.25/35.73**	36.81/14.18/32.62 (Reddit)
XSum	**31.85/10.44/23.75**	24.04/6.43/18.99 (Reddit)
Reddit	**21.47**/4.10/17.62	21.37/**4.14/17.76** (CNNDM)
BigPatent	**35.58/10.91/31.53**	33.57/9.34/25.76 (CNNDM)

Table 1. Zero-shot ROUGE-1/2/L transfer performance from our WikiTransfer data vs transferring from other summarization datasets

Model	ROUGE-1/2/L
WikiTransfer	**39.11/17.25/35.73**
TED (Yang et al., 2020)	38.73/16.84/35.40

Table 2. A comparison of SOTA zero-shot abstractive performance on CNNDM

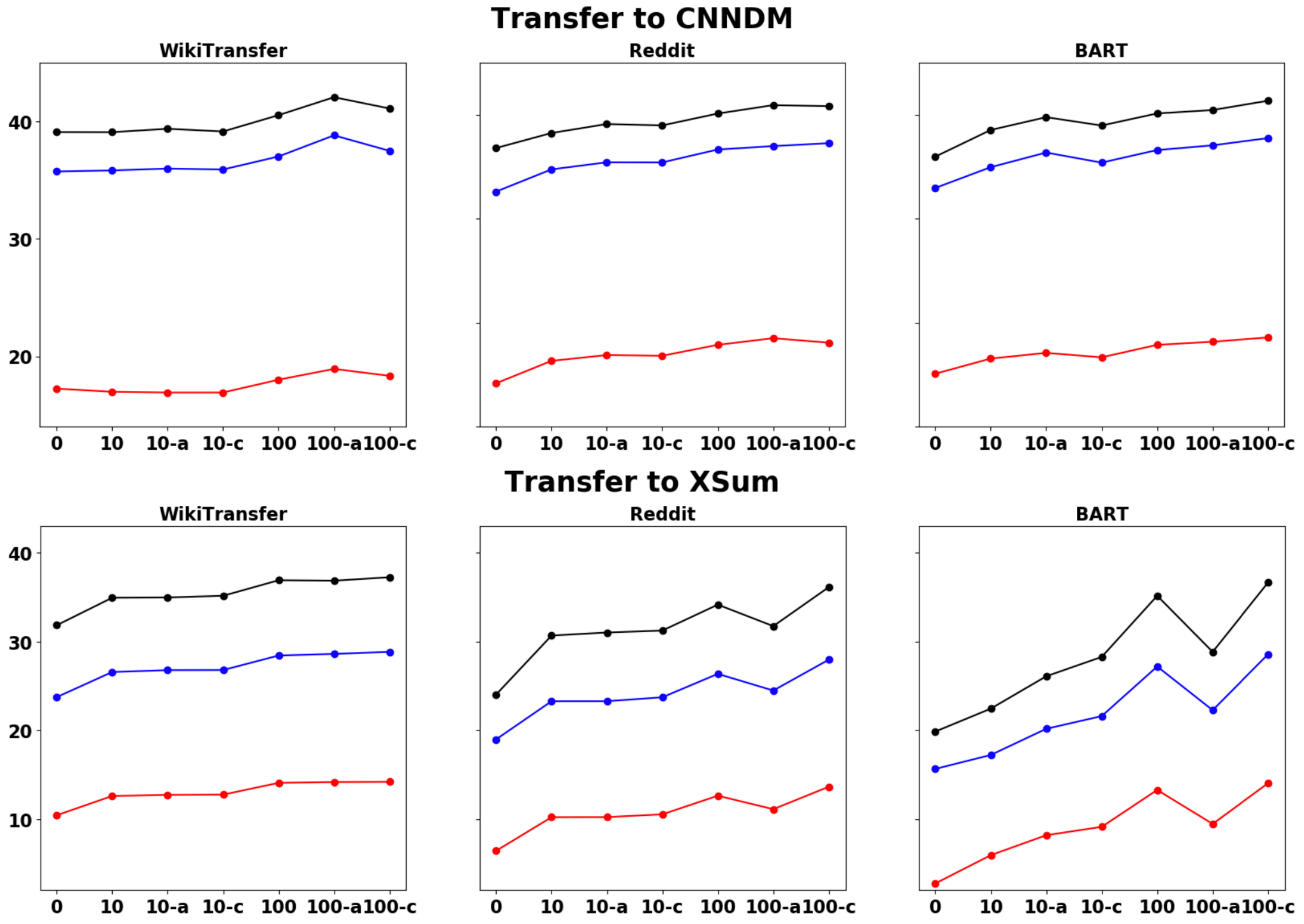

Figure 1. A comparison of few-shot transfer across dataset size and augmentation techniques

Results

Zero-shot results

Zero-shot transfer results from WikiTransfer are shown in Table 1. WikiTransfer models outperform zero-shot transfer from other domains in all domains except for Reddit, for which transfer from CNNDM outperforms in ROUGE-2 and ROUGE-L. We show a comparison of WikiTransfer zero-shot performance to the zero-shot performance of TED on CNNDM in Table 2. We outperform the TED model, which was designed specifically for the news domain, showing the generalizability of our approach.

Few-shot results

We show the performance of transferring WikiTransfer models on CNNDM and XSum in Figure 1. We show a comparison varying the amount of training data, from zero-shot to 100-shot, as well as with round-trip data augmentation (*-a) and consistency regularization. We see that transfer from WikiTransfer outperforms transfer from Reddit or transferring a vanilla BART model. Data augmentation helps WikiTransfer, especially on CNNDM, while consistency regularization gives greater improvements on XSum, making the models more robust to noise in abstractive data augmentation

Conclusion

We introduced WikiTransfer, a novel and generalizable method for fine-tuning pretrained models on dataset-specific unsupervised data obtained from generic Wikipedia data. WikiTransfer models achieve state-of-the-art zero-shot abstractive summarization performance on the CNN-DailyMail dataset and generalize across three additional datasets. In few-shot settings, WikiTransfer models are robust to noise introduced through data augmentation and benefit from consistency loss on more abstractive datasets. Human assessments of the summaries do not show significant differences between the WikiTransfer few-shot summaries and fully-supervised summaries, demonstrating the efficiency of our approach.

Acknowledgement

Thanks to Faiaz Rahman for assisting with the human evaluation