BERT - Bidirectional Encoder Representations from Transformers

Bidirectional Encoder Representations from Transformers (BERT) is a pre-trained transformer-based machine learning technique for natural language processing (NLP) developed by Google.

BERT was pre-trained on two tasks: language modelling and next sentence prediction.

* transformer-based model - a deep learning model that adopts the mechanism of self-attention, differentially weighting the significance of each part of the input data
* BERT encodes context bidirectionally - the model learns information from left to right and from right to left
  + existing models were unidirectional and just read the input from one side
* unsupervised learning - a type of machine learning in which models are trained using unlabeled dataset and are allowed to act on that data without any supervision.

Unsupervised Data Augmentation for Consistency Training Summary

The Unsupervised Data Augmentation (UDA) uses a semi-supervised learning (SSL) method to help address a fundamental weakness in deep learning which is that it typically requires a lot of labeled data to work. Works in SSL based on consistency training have shown to work well on many benchmarks and in order to improve consistency training, high quality data augmentation methods were applied.

Architectural Design

Diagram

Description automatically generated

Main Idea:

The predictions of the unlabeled data should align with the predictions of the same unlabeled data after going through data augmentations

UDA was evaluated on a wide variety of language and task vision and on six classification tasks the model achieves significant improvements over state-of-the-art models. On IMDb, UDA with 20 labeled examples outperforms the state-of-the-art model trained on 1250x more labeled data.

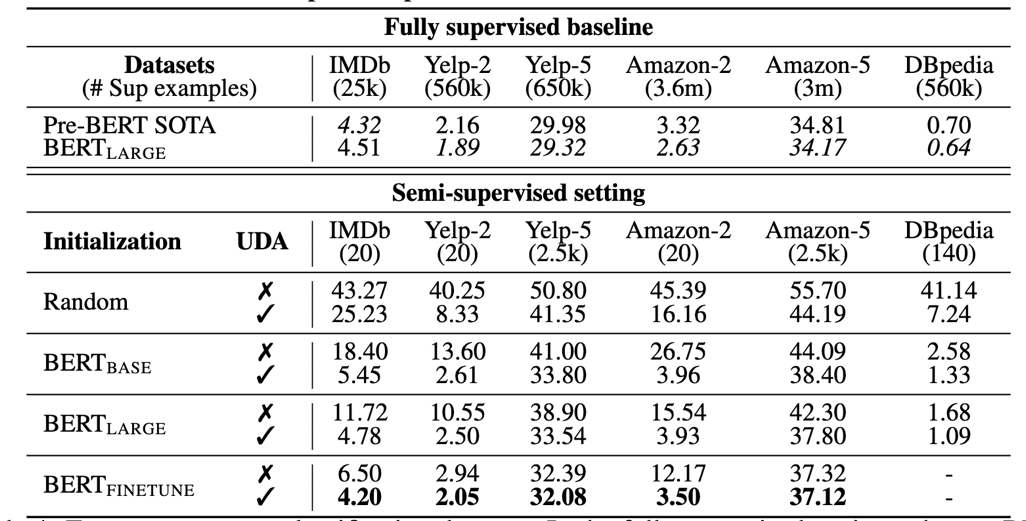
One augmentation method used for text classification was back translation which refers to the procedure of translating an existing example x in language A into another language B and then translating it back into A to obtain an augmented example xˆ. The paraphrases below generated by back-translation sentence are diverse and have similar semantic meanings.

The sentence was translated from English to French and back to English.

Diagram

Description automatically generated

For the evaluation on text classification datasets, UDA was combined with BERT, considering four unsipervised shemas: random Transformer, BERTBASE, BERTLARGE, and BERTFINETUNE. Performances were compared under each of these schemas with and without UDA.



Observations:

* On binary sentiment analysis tasks, with only 20 supervised examples, UDA outperforms the previous SOTA trained with full supervised data on IMDb
* When initialized with BERT and further finetuned on in-domain data, UDA can still significantly reduce the error rate from 6.50 to 4.20 on IMDb.

Fine-tuning BERT on in-domain unsupervised data.

The BERT model was fine-tuned on in-domain unsupervised data using the code released by BERT.

* learning rate of 2e-5, 5e-5 and 1e-4
* batch size of 32, 64 and 128
* number of training steps of 30k, 100k and 300k.

Accuracy on IMDb with different number of labeled examples. In the large-data regime, with the full training set of IMDb, UDA also provides robust gains.

**Accuracy with 20 labeled examples: 95.8%**

Chart, line chart

Description automatically generated

Key points of UDA model

* Uses semi-supervised learning method
* Uses SSL based on consistency training
* Substitute traditional noise injection methods with data augmentation methods
* Evaluated on a variety of language and vision tasks
* [Unsupervised Data Augmentation for Consistency Training](https://arxiv.org/pdf/1904.12848.pdf)
* [UDA GitHub](https://github.com/google-research/uda)
* [Video presentation: Unsupervised Data Augmentation for Consistency Training](https://papertalk.org/papertalks/8414)
* [YouTube video: Unsupervised Data Augmentation](https://www.youtube.com/watch?v=-u8Mi57BDIY)
* [Open Sourcing BERT: State-of-the-Art Pre-training for Natural Language Processing](https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html)