
Transformers and User Intent

Brandon Szeto

Jacobs School of Engineering
University of California, San Diego
San Diego, CA 92122
bszeto@ucsd.edu

Darren Yu

Jacobs School of Engineering
University of California, San Diego
San Diego, CA 92122
dmyu@ucsd.edu

Nathaniel Thomas

Jacobs School of Engineering
University of California, San Diego
San Diego, CA 92122
nathomas@ucsd.edu

Abstract

Introduction

Related Works

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

In Devlin et al. [2019] the authors introduce a groundbreaking model in natural language processing that utilize the concept of bidirectional context representation learning. By pretraining a transformer-based neural network on large corpora with masked language modeling and next sentence prediction tasks, BERT is able to capture deep contextual information from both left and right contexts of a word. This bidirectional understanding of text enables BERT to achieve state-of-the-art performance on a wide range of NLP tasks, including but not limited to question answering, sentiment analysis, named entity recognition, and machine translation.

Supervised Contrastive Learning

In Khosla et al. [2021] the authors introduce supervised contrastive learning, a method for learning representations of data by maximizing agreement between similar pairs and minimizing agreement between dissimilar pairs. Supervised contrastive learning leverages labeled data to guide the learning process, making it particularly suitable for tasks where labeled examples are abundant. The method has shown promising results in various applications, including image classification, natural language processing, and speech recognition.

SimCSE: Simple Contrastive Learning of Sentence Embeddings

In Gao et al. [2022], the authors propose a simple yet effective approach for learning sentence embeddings through contrastive learning. By formulating the contrastive objective in a straightforward manner, SimCSE achieves competitive performance on various downstream tasks without requiring complex architectures or extensive hyperparameter tuning. The method has been shown to learn semantically meaningful representations of sentences, making it valuable for tasks such as semantic textual similarity, text classification, and paraphrase detection.

Methods

(a) It has at least 3 subsections for baseline/custom/contrastive learning. Each of the subsection should start with the final experimental settings you decide that we do not specify, including parameters of the library functions (except for defaults), hyperparameters you choose etc. These experimental setting should correspond to your best results.

Baseline

Custom

Contrastive

Results

Summarize and document the results for all experiments (Baseline Model, Custom Finetuning Strategies, SupContrast and SimCLR).

(a) Include the table(s) (for the 6 experiments as indicated above with “exp idx”)

(b) Provide your training and validation accuracy in each plot for the following 5 models

- i. Baseline model
- ii. first fine-tuning technique

- iii. second fine-tuning technique iv. SupContrast
- v. SimCLR

Discussion

(a) Answer the 11 questions. Answer in the following format (11 pts): Q1: If we do not fine-tune the model, what is your expected test accuracy? Why? A1 : <Fill in Answer> (b) Feel free to include any other points that need to be discussed.

1. If we do not fine-tune the model, what is your expected test accuracy? Explain Why.
2. Do results match your expectation(1 sentence)? Why or why not ?
3. What could you do to further improve the performance ?
4. Which techniques did you choose and why?
5. What do you expect for the results of the individual technique vs. the two techniques combined?
6. Do results match your expectation(1 sentence)? Why or why not?
7. What could you do to further improve the performance?
8. Compare the SimCLR with SupContrast. What are the similarities and differences?
9. How does SimCSE apply dropout to achieve data augmentation for NLP tasks?
10. Do the results match your expectation? Why or why not?
11. What could you do to further improve the performance ?

Contributions

Brandon Szeto: Dataloader, tokenizer, baseline model

Darren Yu:

Nathaniel Thomas:

References

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.

Tianyu Gao, Xingcheng Yao, and Danqi Chen. Simcse: Simple contrastive learning of sentence embeddings, 2022.

Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning, 2021.