

update

May 16, 2023

1 CS 449 Final Project Update

Due: May 23, 2023 at 11:59pm

1.1 1. Names and Net IDs

- Kyle Hwang (ksh6947)
- Michael Lin (qlb968)
- Dylan Wu (dwg0364)

1.2 2. Abstract

Write a new abstract that describes the motivation for your project and proposed methods. This should reflect any changes you've made, but you'll explain those changes later.

Our final project seeks to use `pytorch` to replicate the “Attention is All You Need”, which introduced the Transformer as a way to improve upon existing sequence transduction language models. We will attempt to implement the model’s architecture and train the model on a subset of the WMT 2014 English-French dataset. We will then perform analyses on the training results, and, where appropriate, compare the model performance against results from the paper.

1.3 3. Big Changes

Describe any major changes that have taken your project in a new direction. What made you decide to make these changes? Were some things easier or harder than you expected?

So far, we don’t have any big changes. To focus our implementation to value-added parts of the model, we will use an off-the-shelf tokenizer and BLEU score implementations.

1.4 4a. Describe your dataset(s)

This should be an updated version of your datasets section from your proposal. Write this section as if the reader hasn’t read your original proposal. List the datasets you’re using, where you found them, and what they contain. Be detailed! For each dataset, what does the data look like? What is the data representation? (e.g., what resolution of images? what length of sequences?) How is the data annotated or labeled? Include citations for the datasets. Include at least one citation of previous work that has used this data, or an explanation of why it hasn’t been used before

For example: we use the popular Fashion MNIST dataset, which contains ... grayscale images of clothing. The dataset was created in ...

1.5 4b. Load your dataset(s)

Load at least a few examples and visualize them as best you can

1.6 4c. Small dataset

Describe the small dataset you constructed as requested in the project proposal. How many examples did you use? How large was the dataset in total?

1.7 5. Updated Methods

This should be an updated version of your methods section from your proposal. Write this section as if the reader hasn't read your original proposal. For example:

This is a self-supervised learning task, and we will (as much as we can) implement the Transformer architecture in the paper from scratch using `pytorch`. We will then train the model using KL divergence loss on the WMT 2014 English-French dataset to predict English text from French text. This is the loss function used in the paper, which makes sense because at each token position, there is a predicted probability of the token to place there and an actual token that should be placed there. Given KL divergence loss measures the dissimilarity between the prediction distribution and the true labels, this is a good choice of loss function. We will evaluate the model's performance using BLEU score, which is used by the paper's authors.

1.8 6. Proposed Deliverables

Copy the deliverables from your proposal and write a sentence saying whether you completed this goal, made progress on this goal, or abandoned this goal. Unless you completed the goal, give an explanation of how it went.

1.8.1 6.1 Previous Essential Goals

- For example: We will use Logistic Regression, Support Vector Machine, and Decision Tree models to train and test on our MNIST data.
 - We completed this goal by using the `sklearn` packages.
- (List all your proposed goals here)

1.8.2 6.2 Previous Desired Goals

- For example: We will conduct a thorough hyperparameter search for all our models and provide a detailed and practical guide for how a researcher should search for good hyperparameter values.
 - We made progress on this goal by running a hyperparameter grid search for both `sklearn` and `torch` models, but...
- (List all your proposed goals here)

1.8.3 6.3 Previous Stretch Goals

- For example: We will conduct a manual analysis of the digits that our model gets wrong and use a GAN to create new images that help us learn a more robust classifier.
 - We abandoned this goal because we didn't want to conduct a manual analysis.
- (List all your proposed goals here)

1.9 7. Results So Far

Show the results of your methods trained on your data. You should at least have empirical results for the small dataset described in 4c. Unless you are doing generation or another task where simple models (e.g., linear regression or logistic regression) aren't applicable, we strongly encourage you to compare your initial results against a simple baseline. If you have additional results, you're welcome to include those as well.

For example: our logistic regression baseline achieves 80% accuracy on Fashion-MNIST, and our MLP achieves 85% accuracy, when applied to our small dataset containing 1,000 subsampled images.

1.10 8. New Deliverables

For any proposed deliverables that you did NOT complete or abandon, copy them into these lists below. Then, if you have abandoned or completed your Desired or Stretch goals, add new goals until you have at least two goals per category.

8.1 New Essential Goals

- Copy the old essential goals that you haven't yet completed or abandoned.
- Don't add any new Essential goals.

8.2 New Desired Goals

- Copy the old desired goals that you haven't yet completed or abandoned.
- Make sure you have at least two Desired goals left to complete here, adding new goals as necessary.

8.3 New Stretch Goals

- Copy the old stretch goals that you haven't yet completed or abandoned.
- Make sure you have at least two Stretch goals left to complete here, adding new goals as necessary.

1.11 9. Hopes and Concerns

What's gone well so far? What's been the hardest thing?

For example: we got really bored of staring at pixelated hand-written digits for hours on end.

1.12 10. References

Cite the papers or sources that you used to discover your datasets and/or models, if you didn't include the citation above. For example:

Vaswani, Ashish, et al. “Attention Is All You Need.” ArXiv (2017): /abs/1706.03762.

Kitaev, Nikita, et al. “Reformer: The Efficient Transformer.” ArXiv (2020): /abs/2001.04451.