Team Name

Small NLP

Project Title

Lightweight Sentiment Analysis: Comparing Small LLMs, Large Models, and Traditional ML Approaches

Project summary

With the recent rise of extremely powerful LLMs which perform sentiment analysis at record high scores there is also a greater need for considerable amounts of compute to run such models. The deep learning models which excel at this NLP task usually require cloud compute or a local GPU cluster which users may have to outsource to. As such this project seeks to evaluate and compare the performance of a fine-tuned very small deep learning model which will be able to run on a lot more hardware then the leading LLMs such as BERT or GPT.

The smaller model will be fine-tuned to IMDb's movie review dataset to assess its performance in sentiment analysis. We will also compare it to the computationally intensive models. As well as evaluating the small deep learning models performance versus industry grade models it will also be compared to traditional ML techniques, and standard non-transformer deep learning networks. Some of the other models for evaluation include: CNNs, LSTM, LSTM-CNN, Logistic Regression, Naive Bayes, and Random Forest learners. These benchmarks of performance will be attained from the existing literature for this well tested dataset.

With the research into how a smaller sized model can hopefully outperform the traditional ML techniques while achieving reasonable scores when compared to the larger models we hope to show how a locally runnable model can still perform sentiment analysis well. This can show ways to reduce the deployment and computation costs of the larger models with sufficient tuning of the smaller deep learning model. We can also perhaps show that there are diminishing returns when using more and more complex NLP models for tasks that aren't too hard (like most sentiment analysis)

What you will do

We will first setup SmolLM, a publicy available 135M-parameter LLM on Georgia Tech's PACE ICE GPUs. Then we will setup code to establish the baseline performance of this model used for sentiment analysis. We will then try to use few-shot learning and fine-tuning to improve the model's performance on sentiment analysis, using a dataset from IMDB movie reviews. We will perform an in-depth analysis explaining what model parameters are important, which can be tuned and why, and how the fine-tuning process can improve over our baseline performance. We will also assess what performance can be reached with the same evaluation procedure, when replacing the LLM to another model such as the ones mentioned in the project summary.

Resources / Related Work & Papers

For state-of-the-art sentiment analysis, transformer-based models like BERT, RoBERTa, and GPT-3 have demonstrated remarkable performance due to their ability to capture contextual information and semantic nuances. These models leverage pre-trained embeddings fine-tuned on vast datasets, resulting in a deep understanding of sentiment at both the word and sentence levels. Other transformer-based approaches, such as T5 and DeBERTa, utilize enhanced architecture to improve the capture of syntactic and semantic structures, further boosting performance. Notably, SOTA models now integrate task-specific tuning or domain adaptation for optimal results in specific applications, like social media sentiment analysis. These models frequently surpass traditional benchmarks and achieve top scores on popular sentiment datasets, including SST-2 and IMDB. The advantage of complex models is that for the harder examples (such as people using sarcasm), a real in-depth understanding of humans and their use of language can be required.

For non-transformer-based methods, RNNs (e.g., LSTMs and GRUs) and CNNs were previously popular due to their effectiveness in sequential text processing, capturing dependencies in text through memory and convolutional filters. Though less context-aware than transformers, these architectures still perform well when combined with pre-trained word embeddings (like GloVe or Word2Vec), offering competitive sentiment analysis performance at lower computational costs.

Relevant papers

[1] Links to some benchmarks for models on the IMDB dataset we are using. [2] talks about traditional methods and their performance on the dataset. [3] is another paper on deep NNs run on the same dataset (CNN, LSTM, and LSTM-CNN). [4] is very close to [3]. [5] focuses on traditional ML methods for our dataset.

- [1] Papers With Code Websitelink (https://paperswithcode.com/sota/sentiment-analysis-on-imdb)
- [2] Tripathi, Sandesh, et al. "Analyzing sentiment using IMDb dataset." link (../documentation/Analyzing_Sentiment_using_IMDb_Dataset.pdf)
- [3] Haque, Md Rakibul, Salma Akter Lima, and Sadia Zaman Mishu. "Performance analysis of different neural networks for sentiment analysis on IMDb movie reviews." link (../documentation/Performance_Analysis_of_Different_Neural_Networks_for_Sentiment_Analysis_on_IMDb_Movie_Reviews.pdf)
- [4] Ali, Nehal Mohamed, Marwa Mostafa Abd El Hamid, and Aliaa Youssif. "Sentiment analysis for movies reviews dataset using deep learning models." link (../documentation/ssrn-3403985.pdf)
- [5] Ghosh, Ayanabha. "Sentiment analysis of IMDb movie reviews: A comparative study on performance of hyperparameter-tuned classification algorithms." link (../documentation/Sentiment_Analysis_of_IMDb_Movie_Reviews__A_comparative_study_on_Performance_of_Hyperparameter-tuned_Classification_Algorithms.pdf)"

Datasets

IMDb Movie Reviews for Sentiment Analysis:

- https://ai.stanford.edu/~amaas/data/sentiment/ (https://ai.stanford.edu/%7Eamaas/data/sentiment/)

Other links to the dataset:

- https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews (https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews)
- https://paperswithcode.com/dataset/imdb-movie-reviews (https://paperswithcode.com/dataset/imdb-movie-reviews)

List your Group members.

- Jaden Zwicker
- Brieuc Popper
- Houshmand Abbaszadeh