Ask Project Nexodus Docs/Project Aspen

Intern Final Presentation

Christina Xu
Ask Project
Nexodus Docs: Leveraging LLMs for Documentation
Q&A
Main strategies for QA tasks

**Extractive**
“Extracts” answer directly from docs based on its similarity to the question.

**Abstractive**
Generates answer based on question and context - context being extracted directly from docs.

**Generative**
Generates answer not explicitly found in docs eg) ChatGPT.
What is Fine-tuning

- LLMs are pre-trained on specific domains and tasks such as text generation, question answering, etc. We might want to train the LLM to adapt to our data and task.

**Adapter Tuning**

Add more layers to the pre-trained model and train weights only in those additional layers.
What is Fine-tuning?

Traditional approaches are not practical

- LLMs are trained on specific domains and tasks such as text generation, question answering, etc. We might want to train the LLM to adapt to our data and task.

**Drawbacks**

- Inference latency - the more layers, the longer it takes for the model to generate an answer.
Fine-tuning with LoRA (Low Rank Approximation)

Update pretrained weights in the model
During training...

Decompose $\Delta W$ into $A$ and $B$

Consider a 100 x 100 matrix $\Delta W$. That would mean we would have to train 10,000 parameters. If we decompose it into matrices $A$ and $B$, which are 100 x 1 and 1 x 100, respectively, we only have 100 parameters to train in each or 200 in total.
After training...

Merge $W$ with $\Delta W$
LLM Strategy Evaluation

Key Idea: human language is difficult to quantitatively evaluate

https://huggingface.co/spaces/evaluate-metric/wer
https://huggingface.co/spaces/evaluate-metric/rouge
https://huggingface.co/spaces/evaluate-metric/bertscore
Project Aspen: Bus Factor
What is Project Aspen?

Analyzes data from open source projects to empower contributors and participants to make data driven decisions about open source communities and projects.
Bus Factor

How high the risk is to a project should the most active people leave?

- Quantifies the amount of contributors a project can afford to lose before it stalls by hypothetically having these people get run over by a bus

- Typically, it is the smallest number of people that make up 50% of contributions
How do we define “contributions”? 

We can analyze bus factors from different perspectives

Commits $\rightarrow$ Issues $\rightarrow$ Pull Requests
There appears to be a trend in the top 10 contributors across all perspectives eg) 01012f1b, 01000c4d, 01000cc2

The proportion between the top 10 and ‘other’ contributors for each perspective matches our intuition
Bus factor as a function of time
Thank you

Special thanks to Sanjay Arora, James Kunstle, Heidi Dempsey, Jen Stacy, and my fellow research interns

For questions or concerns regarding my projects, feel free to reach out to me via: chrxi@redhat.com

Let’s connect!

https://github.com/oss-aspen/Rappel

https://github.com/christinaexyou/ask_project_nexodus_docs_(WIP)

https://www.linkedin.com/in/christinaexyou/

https://medium.com/@christinaexyou