

Final Project - Question Answering

شرح پروژه

الف) روی مجموعه داده PQuAD مدل ParsBERT را Fine-Tune کنید. نتایج معیارهای F1-Score و Exact Match (EM) گزارش دهید.

ب) از طریق روش های Zero-Shot و Few-Shot مدل Llama-3.1-8B-4bit را برای پرسش و پاسخ استفاده کنید. نتایج را بر اساس معیارهای F1-Score و Exact Match (EM) تحلیل و مقایسه کنید.

ج) روی مجموعه داده PQuAD مدل Llama-3.1-8B-4bit را با روش هایی نظیر LoRA و QLoRA Fine-Tune کنید. در خروجی پروژه یک مدل آموزش دیده و بهینه سازی شده که قادر به پاسخدهی به سوالات فارسی باشد مورد انتظار است. و در معیار هایی نظیر F1-Score و Exact Match (EM) دقت بالایی داشته باشد.

بخش امتیازی (حداکثر ۱.۵ نمره)

ایجاد مدلی که از روش RAG برای بازیابی اطلاعات و پاسخدهی به سوالات استفاده کند. برای مدل بازیابی از TF-IDF یا BM25 و مدل زبانی از Llama-3.1-8B-4bit استفاده نمایید. برای مجموعه داده از فایل PDF پیوست شده امکان استفاده دارید. نتایج معیارهای F1-Score و Exact Match (EM) برای ارزیابی کیفیت مدل گزارش دهید.

طراحی واسط کاربری ساده (Web Interface) برای نمایش عملکرد مدل های Question Answering.

ساختار تحویل پروژه

- کدها در قالب یک فایل زیپ یا لینک به یک مخزن گیت‌هاب باشند.
- مستندات پروژه در قالب PDF یا Markdown نوشته شوند.
- همه فایل‌های مربوط به پروژه تا پایان مهلت ارسال در کوئرا آپلود شده باشد.
- پیش‌پردازش داده‌ها، معماری مدل، ارزیابی عملکرد، و چالش‌های مواجهه شده در مسیر انجام پروژه باید شرح داده شود.

نکات مهم

- پیاده‌سازی‌ها باید قابلیت تست گرفتن توسط مصحح را داشته باشند.
- اگر دسترسی به GPU و سخت‌افزار محدود دارید استفاده از Google Colab و یا Kaggle مجاز است.
- می‌توانید پروژه را به صورت انفرادی یا در گروه‌های حداکثر سه نفره انجام دهید.
- هر دانشجو موظف است در هنگام تحویل حضوری، مستقلاً به تمام جزئیات مسلط باشد.
- با موارد تقلب و کپی‌کردن، طبق تشخیص دوستان حل تمرین و استاد درس، برخورد جدی خواهد شد.
- زمان بندی تحویل و ارائه حضوری پروژه، متعاقباً اعلام خواهد شد.

منابع مورد نیاز

- لینک دیتاست‌ها :

<https://huggingface.co/datasets/Gholamreza/pquad>

در صورت محدودیت سخت‌افزاری امکان استفاده از دیتاست PersianQA هم دارید

https://huggingface.co/datasets/SajjadAyoubi/persian_qa

- لینک مدل‌ها :

<https://huggingface.co/unsloth/llama-3-8b-bnb-4bit>

<https://huggingface.co/HooshvareLab/bert-base-parsbert-uncased>

<https://huggingface.co/Qdrant/bm25>

همیشه در پناه خدا موفق باشید

امین رضائزاد – محمد نورسته