



دانشکده مهندسی کامپیوتر

درس پردازش زبان طبیعی

تمرین سری چهارم

مهلت تحویل:

1404 / 2 / 26

استاد :

دکتر مرضیه داوودآبادی

طراح تمرین :

فرناز میرنظامی

موضوع تمرین: تولید داده مصنوعی، تنظیم و ارزیابی مدل زبانی در حوزه پزشکی

حال که با مدل‌های زبانی بزرگ (LLMs)، روش‌های تنظیم (fine-tuning) و ارزیابی عملکرد آن‌ها آشنا شده‌اید، در این تمرین قصد داریم به صورت عملی، داده مصنوعی تولید کرده، مدل را روی داده مصنوعی تنظیم کرده و عملکرد مدل را ارزیابی کنیم.

گام اول: تولید داده مصنوعی در حوزه پزشکی

با استفاده از تکنیک‌های پرامپت‌سازی (prompt engineering) و با استفاده از ChatGPT، یک مجموعه داده مصنوعی تولید کنید. برای تولید این دیتاست از منبع داده شده استفاده کنید. (لطفاً باتوجه به [جدول اسامی](#)، فایل مربوط به خود را از [گوگل درایو](#) دانلود کنید.)

○ فرمت داده‌ها باید به صورت زیر باشد:

- ورودی: علائم بیمار (symptoms)
- خروجی: تشخیص بیماری (diagnosis)

نمونه :

```
[  
  {  
    "symptom": "مردی ۵۰ ساله با کاهش ناگهانی بینایی در چشم راست و بدون درد، با نوار چشم‌سنجی حاکی از "نبود جریان خون شبکه‌ای مراجعه کرده است",  
    "diagnosis": "انسداد شریان مرکزی شبکیه"  
  }  
]
```

- تعداد نمونه‌های تولیدی: 100 جفت پرسش و پاسخ
- دیتاست به زبان فارسی و فرمت آن json باشد.

گام دوم: تنظیم مدل روی دیتاست ساخته شده

از مدل [Llama-3.2-1B](#) به عنوان Base Model استفاده کرده و آن را روی داده تولیدشده در مرحله قبل، به کمک تکنیک [LoRA](#) تنظیم (fine-tune) نمایید.

مدل نهایی را به عنوان یک مدل جدید ذخیره و مستندات مربوط به آموزش آن (پارامترهای آموزش، تنظیمات LoRA ، کتابخانه ها) را در گزارش ارائه دهید.

گام سوم: ارزیابی عملکرد مدل جدید

برای ارزیابی مدل جدید، با استفاده از پرامپت مناسب و ChatGPT ، یک دیتاست مصنوعی به فرم سوالات چهارگزینه ای همراه با پاسخ درست (MCQs) تولید کنید:

نمونه:

```
[
{
  "question": "کدام یک از موارد زیر از عوارض شایع آبله مرغان است؟",
  "options": [
    "پانکراتیت",
    "پنومونی",
    "هپاتیت ویروسی",
    "فلج اطفال"
  ],
  "answer": "پنومونی"
},
]
```

برای این دیتاست نیز 100 سوال و جواب تولید شود و به فرمت json باشد.

معیار ارزیابی دقت (Accuracy) می باشد .

با استفاده از پرامپت مناسب ، خروجی مدل را روی دیتاست ارزیابی ، تولید کنید و سپس با جواب درست در دیتاست مقایسه کنید و دقت را محاسبه کنید.

گزارش تمرین باید شامل موارد زیر باشد:

- پرامپت های استفاده شده
- مشخصات پیکره بندی مدل جدید
- نمونه دیتاست های ساخته شده برای fine-tuning و ارزیابی
- نتیجه ارزیابی
- نمونه خروجی مدل روی دیتاست ارزیابی
- جدول مقایسه ای عملکرد مدل پایه و مدل fine-tuned

انجام هر یک از موارد زیر نمره امتیازی دارد .

- استفاده از مدل های دیگر برای fine-tuning علاوه بر مدل گفته شده
- مقایسه عملکرد چند مدل مختلف در ارزیابی
- پیاده سازی مرحله ی Continuous Pretraining پیش از fine-tuning

نکات تکمیلی:

- پاسخ کامل را در یک فایل نوتبوک (.ipynb) قرار دهید و خروجی ها را نیز نمایش دهید.
- فایل جیسون دیتاست آموزش و فایل جیسون دیتاست ارزیابی به همراه کد و گزارش ارسال شود.
- فایل ها را با نام StudentID_FirstName_LastName_HW4.zip فشرده و در سامانه بارگذاری کنید.
- کدهای بدون گزارش یا تحلیل، امتیازی نخواهند داشت.
- تأخیر در تحویل باعث کسر ۵۰٪ نمره به ازای هر روز خواهد شد.
- ذکر منابع مورد استفاده الزامی است.