

<div><p>دانشگاه علم و صنعت ایران</p></div> <div><p>دانشکده</p><p>مهندسی کامپیوتر</p><p>مدرس:</p><p>دکتر مرضیه داوود آبادی</p></div>	<p>به نام او</p>	<p>مبانی پردازش زبان طبیعی</p>
<div><div><ul style="list-style-type: none">تاریخ تحویل: ۱۴۰۳/۰۴/۰۱اگر برای جواب دادن به سوالی نیاز به پیش فرضی دارید، فرض خود را صریحا نوشته و با توجه به فرض خود به ارائه جواب بپردازید.به هیچ وجه تمرینی را از دیگران کپی نکنید. در صورت مشاهده تقلب و کپی در تمرینات، نمره هردو طرف <u>صفر</u> در نظر گرفته می شود.طراح: محمدمامین عباسی</div></div>		
نمره		
<div><div><div><h3>استفاده از LLM های فارسی برای Query Paraphrasing</h3><p>در این تمرین، شما باید از چندین مدل زبان بزرگ (LLM) برای حل مسئله Query Paraphrasing یا تشخیص جمله های مترادف استفاده کنید. دیتاست Query Paraphrasing شامل زوج های سؤال است که باید تعیین کنید آیا این دو سؤال مترادف یکدیگر هستند یا خیر. هدف این تمرین، آشنایی با مدل های زبان بزرگ و ارزیابی عملکرد آنها در تشخیص جمله های مترادف است. شما با استفاده از سه مدل مختلف و در سه سناریوی مختلف (Zero-shot، One-shot و Five-shot) عملکرد این مدل ها را مقایسه خواهید کرد. این تمرین به شما کمک می کند تا با Prompt Engineering و In Context Learning آشنا شوید. در این تمرین نیاز به آموزش مدل نیست و فقط از تعامل با مدل ها تست های مرتبط را انجام دهید.</p><div><div>1</div><div><h4>اهمیت تشخیص جمله های مترادف</h4><p>تشخیص جمله های مترادف یکی از مسائل مهم در پردازش زبان طبیعی است. این مسئله در کاربردهای مختلفی مانند جستجوی اطلاعات، پاسخ به سوالات، و سیستم های توصیه گر اهمیت دارد. به عنوان مثال، در یک سیستم جستجوی اطلاعات، تشخیص اینکه دو سوال مختلف به یک موضوع اشاره دارند، می تواند به بهبود نتایج جستجو کمک کند.</p><div><h4>مدل های مورد استفاده</h4><p>شما باید از مدل های زیر استفاده کنید:</p><ul style="list-style-type: none">meta-llama/Meta-Llama-3-8B-InstructViraIntelligentDataMining/PersianLLaMA-13B-Instructuniversitytehran/PersianMind-v1.0</div></div></div></div></div></div>		

سناریوهای آزمایش

برای هر مدل، شما باید سه سناریوی زیر را آزمایش کنید:

- **Zero-shot Learning**: در این سناریو، مدل بدون هیچ داده آموزشی، باید تشخیص دهد که آیا دو سؤال مترادف هستند یا خیر. این سناریو نشان می‌دهد که مدل تا چه حد می‌تواند بدون آموزش قبلی، وظیفه را انجام دهد.
- **One-shot Learning**: در این سناریو، یک مثال آموزشی به مدل داده می‌شود و سپس مدل باید تشخیص دهد که آیا دو سؤال مترادف هستند یا خیر. این سناریو نشان می‌دهد که مدل تا چه حد می‌تواند از یک مثال آموزشی یاد بگیرد.
- **Five-shot Learning**: در این سناریو، پنج مثال آموزشی به مدل داده می‌شود و سپس مدل باید تشخیص دهد که آیا دو سؤال مترادف هستند یا خیر. این سناریو نشان می‌دهد که مدل تا چه حد می‌تواند از چند مثال آموزشی یاد بگیرد.

دیتاست

دیتاست Query Paraphrasing شامل زوج‌های سؤال به همراه برچسب (0 یا 1) است که نشان می‌دهد آیا دو سؤال مترادف هستند یا خیر .

<https://github.com/persiannlp/parsinlu/tree/master/data/qqp>

نکات مهم

- سعی کنید بهترین Prompt را برای هر مدل پیدا کنید.
- با استفاده از مثال‌های آموزشی، مدل‌ها را بهبود دهید و عملکرد آنها را ارزیابی کنید.
- روند کار، پارامترهای استفاده شده برای تعامل با مدل و نتایج به دست آمده را به دقت تحلیل کنید و تفاوت عملکرد مدل‌ها را در سناریوهای مختلف بررسی و گزارش کنید.

موفق باشید.