

# به نام خدا

تمرین سری هفتم

درس مبانی پردازش زبان و گفتار

سینا علی نژاد

شماره دانشجویی: ۹۹۵۲۱۴۶۹

نتایج مدل اول (PersianMind):

Zero-shot = [1,1,1,1,1]

One-shot = [1,1,1,1,1]

Few-shot = [1,1,1,1,1]

Real labels = [1,0,0,1,0]

برای این مدل با اینکه در one-shot و five-shot نمونه هایی دادم اما همچنان خروجی ۱ میداد. ۱ به معنای اینکه دو سوال را مشابه میدانست. دقت آن بسیار پایین بود. البته این که توانستم یک خروجی مشخص بگیرم، بخاطر پرامپت ورودی بود. در ابتدا مدل جواب مشخصی به همان شکل که انتظار داشتم نمیداد. اما با مشخص کردن عبارت PersianMind: در تمپلیت پیام، مدل جواب خود را (صفر یا یک) در آن قسمت قرار داد.

نتایج مدل دوم (PersianLlama):

Zero-shot = [1,1,0,0,0]

One-shot = [0,1,0,0,1]

Few-shot = [N,1,N,N,1]

Real labels = [1,0,0,1,0]

در این مدل خروجی صفر هم گرفتم و در حالت zero shot ۳ مورد را درست جواب داد اما در حالت one shot اوضاع بدتر شد و در حالت five shot حتی خروجی مناسب که بتوان پاسخ را فهمید نمیداد و صرفاً متن سوال را در جواب کپی میکرد. به نظرم این اتفاق بدین دلیل است که وقتی در حالت few shot میرویم حجم متن ورودی طولانی میشود و مدل منحرف میشود. با پرامپت های کوتاه تر و ساده تر به جواب های بهتری میرسیدم.

نتایج مدل سوم (MetaLlama):

Zero-shot = [1,1,0,1,0]

این مدل در حالت zero shot بهترین عملکرد را به نسبت بقیه مدلها داشت و فقط یک غلط داشت. در اینجا هم با یک تغییر کوچک در پرامپت و با قرار دادن Response: توانستم خروجی دلخواه را از مدل دریافت کنم. هرچند ۵ ورودی برای تصمیم گیری اصلا کافی نیست اما به دلیل زمان بر بودن لود شدن مدلها و همچنین inference آنها، مجبور بودم تعداد کمی ورودی را امتحان کنم. برای حالت one-shot و five-shot این مدل متأسفانه نتوانستم به خروجی برسم زیرا انگار مدل وقتی ورودی طولانی (حالت one shot و five shot) می‌گرفت، در تولید خروجی وارد یک حلقه بینهایت میشد و هرگز متوقف نمیشد. کد مربوطه را در نوتبوک قرار دادم.

برای پرامپت در همه مدلها از یک تمپلیت ثابت استفاده کردم:

```
TEMPLATE = "Context: {context}\nPrompt: {prompt}\nResponse: "
```

و بخش پرامپت را به صورت زیر دادم:

```
PROMPT = " دو تا متن سوال به تو می دهم و تو باید بگویی که این دو سوال،  
معادل یکدیگر هستند یا خیر. اگر معادل هستند، عدد ۱ را خروجی بده و در غیر  
"سوال دوم: {سوال ۲}\سوال اول: {سوال ۱}. اینصورت عدد ۰ را خروجی بده
```

در بخش context نیز در حالت one shot و five shot صرفاً نمونه هایی از ورودی و خروجی قرار دادم.

برای لود کردن مدلها از روش quantization استفاده کردم با استفاده از پارامتر load\_in\_8bit هنگام لود مدل و همچنین کتابخانه bitsandbytes.

همچنین فایل های json مربوط به خروجی های دو مدل اول را نیز در فایل زیپ فرستادم.