به نام خدا

گزارش تمرین سوم درس مدل‌های زبانی بزرگ

حمیدرضا امیرزاده

۴۰۱۲۰۶۹۹۹

نوتبوک اول

در ابتدا یک نمونه از خروجی سیستم حاصل شده با دادن ورودی تصویر زیر را مشاهده می کنید:

کپشن های تولید شده و سپس ترجمه شده به زبان فارسی برای این تصویر مطابق زیر هستند:

کپشن تولیدی به کمک روش beam search: *چند نفر روی نیمکت کنار یک اسب سفید نشسته اند.*

کپشن تولیدی بدون کمک از روش beam search: *یک سگ سفید و چند نفر روی یک نیمکت نشسته اند.*

همانطور که مشاهده می شود، روش دوم خروجی بهتری دارد.

ارزیابی

در جدول زیر نتایج ارزیابی با توجه به معیارهای ارزیابی BLEU و METEOR برای هر یک از دو روش تولید ذکر شده آورده شده است. توجه شود که این ارزیابی ها روی یک مجموعه داده به اندازه ۱۰۰ که از دیتاست COCO جدا شده اند و کپشن واقعی آن ها به کمک یک مدل ترجمه به زبان فارسی برگردانده شده است، حاصل شده اند.

روش	BLEU Score	METEOR Score	ROUGE-1
به کمک beam search	0.238	0.351	0.441
بدون beam search	**0.250**	**0.357**	**0.453**

روش بدون beam search امتیازهای بهتری گرفته است. از دلایل احتمالی این موضوع می توان به عدم تنظیم مناسب عرضbeam و یا مشکلات مربوط به مدل تولید خروجی GPT2 اشاره کرد.

نوتبوک دوم

پاسخ هر دو روش تک مودالیتی و دو مودالیتی به سوال مذکور به صورت زیر بود:

No, DALL-E2 does not use a CLIP model inside.

اما این پاسخ غلطی است و جواب درست سوال داده شده بله بوده است. این نتیجه باتوجه به اینکه حقیقت های مرتبط متنی موجود در مقاله خود DALL-E2 نیز به ورودی مدل داده شده است کمی عجیب است. از دلایل این شکست می توان به قدرت نسبی پایین مدل Koala-7B در بازیابی اطلاعات مرتبط را اشاره کرد.

نکته دیگر این است که در قسمت دو مودالیتی، از کپشن هایی برای استفاده مدل در پاسخگویی استفاده شده است که عملا ارتباطی با سوال داده شده ندارند. به طور مثال اکثر تصاویر موجود در مقاله درباره نحوه کارکرد خود مدل DALL-E2 اطلاعاتی ندارند، بنابراین استفاده از کپشن مربوط به این نوع تصاویر به عنوان حقیقت برای کمک به مدل زبانی در استخراج جواب برای سوال مذکور، عملا بهره ای ندارد و کمکی نمی کند.