مهدیه نادری ۹۸۵۲۲۰۷۶

(g(۱

این مسک ها برای نادیده گرفتن padded tokens هنگام محاسبه امتیاز attention استفاده می‌شوند. به طور خاص، امتیازات attention برای توکن‌های padded روی منفی بینهایت تنظیم می‌شود که معادل داشتن امتیاز 0 پس از اعمال softmax است. بنابراین، توکن‌های padded به جمع وزنی حالت‌های پنهان رمزگذار کمکی نمی‌کنند. این امر ضروری است زیرا توکن های padded حاوی هیچ اطلاعات معنی‌داری نیستند و نباید برای محاسبه امتیازات attention استفاده شوند، که تعیین می‌کند کدام حالت‌های پنهان encoder باید وزن بیشتری در مجموع وزنی داشته باشند.

در محاسبه attention ، مقادیر et را برای سلول هایی که mask درست است، روی −INFINITY قرار می دهیم. با انجام این کار، در محاسبه آلفا، زمانی که از تابع softmax روی مقادیر بسیار منفی استفاده می کنیم، احتمال توجه به آن سلول حدود 0 است. و از آنجایی که آنها سلول های padding هستند، لازم است در خروجی دیکودر، احتمال 0 وجود داشته باشد.

استفاده از مسک ها در این راه برای جلوگیری از توجه مدل به توکن های padded در حالت های پنهان encoder ضروری است. اگر مدل به توکن‌های padded توجه کند، وزن نامناسبی به آن‌ها می‌دهد و احتمالاً باعث می‌شود مدل پیش‌بینی‌های نادرستی انجام دهد. مسک‌ها تضمین می‌کنند که مدل فقط به توکن‌های بدون padded که حاوی اطلاعات معنی‌دار هستند توجه می‌کند.

(h(۱

BLEU score: 19.9724

(i(۱

از مضرات روش Additive attention میتوان به این اشاره کرد که که اغلب کندتر و کارآمدتر از multiplicative attention است، اما این مزیت را دارد که در ابعاد بزرگتر بهتر عمل کند.

نقطه ضعف ضربه داخلی دو بردار (dot product) این است که نیاز دارد decoder و encoder دارای ابعاد یکسان باشند.

یکی از مزیت های آن این است که نیاز به پارامترهای کمتری نسبت به روش multiplicative attention دارد، اگرچه روش هایی برای کاهش تعداد پارامترها در multiplicative attention مانند کاهش رنک وجود دارد.

(a(۲

convolution پنجره ای است که روی چندین قطعه ورودی می لغزد. در این حالت رابطه ای بین embedingهای کاراکتر های مجاور ایجاد می کند. گنجاندن این وابستگی برای ترجمه چینی بسیار مهم است، زیرا اغلب از چند نویسه چینی برای ایجاد معنای معنایی یک کلمه، با نگاه کردن به خطوط مختلف src.vocab استفاده می‌شود.

(i(b(۲

(1) مدل NMT به اشتباه "culprits" را به جای جمع به مفرد ترجمه می کند.

(2) احتمالا به خاطر این است که شکل های مفرد و جمع کلمات امبدینگ های بسیار مشابهی دارند.

(3) شاید بتوان امبدینگ را طوری تنظیم کرد که اسامی را به صورت مفرد از جمع رمزگذاری کند.

(ii(b(۲

مهدیه نادری ۹۸۵۲۲۰۷۶

(1) "resources have been exhausted" به جای ترجمه مستقیم بند اول مربوط به people، دو بار تکرار می شود.

(2) به این neural text degeneration گفته می شود، که در آن استفاده از احتمال به عنوان یک هدف آموزشی منجر به نتایج تکراری می شود.

(3) یکی از راه های رفع این مشکل ممکن است regularizing مدل در برابر ایجاد کلمات و عبارات تکراری باشد. اگرچه در مورد تنها دو تکرار، این ممکن است دشوار باشد.

(iii(b(۲

(1) مدل NMT به اشتباه "a national mourning today" را به "today's day" ترجمه می کند که هیچ معنایی ندارد.

(2) شاید "a national mourning" به خودی خود یک عبارت انگلیسی مبهم باشد. مثلا ".a national day of mourning" راحت تر و قابل درک تر است.

(3) این ممکن است با داده های آموزشی بیشتر از نمونه های مبهم عباراتی که به راحتی از چینی به انگلیسی وصل نمی شوند، و به نوعی اصلاح آن زبان هستند، حل شود.

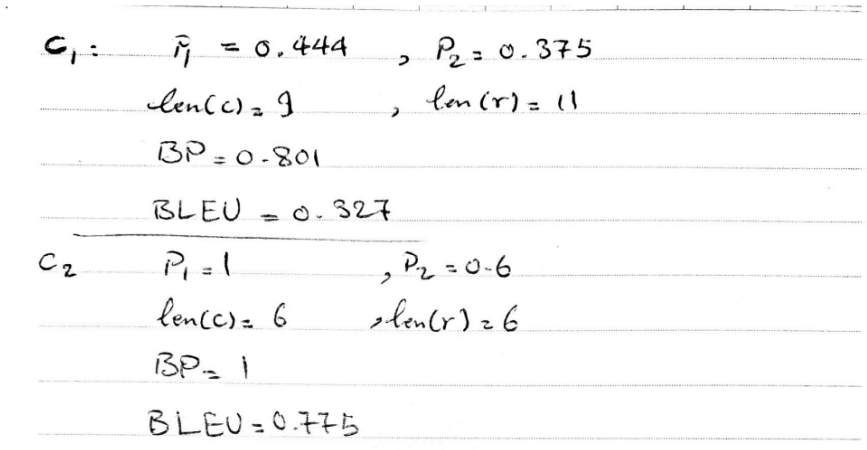
(iv(b(۲

(1) " "act not, err not اصطلاحی است که مدل NMT قادر به تفسیر آن نیست. به اشتباه معنای این عبارت را به "it's not wrong," ترجمه می کند.

(2) با توجه به معیارهای زبان امروزی و همچنین بافت بقیه جمله، این شکل کاملاً غیرعادی است. اگر نقل قول‌ها در میان باشد، احتمالاً شاهد ناپیوستگی در سبک گفتار هستیم

(3) اگر نقل قول‌ها درگیر هستند، شاید بهتر باشد وابستگی مدل به کانتکست مجاور را کاهش دهید.

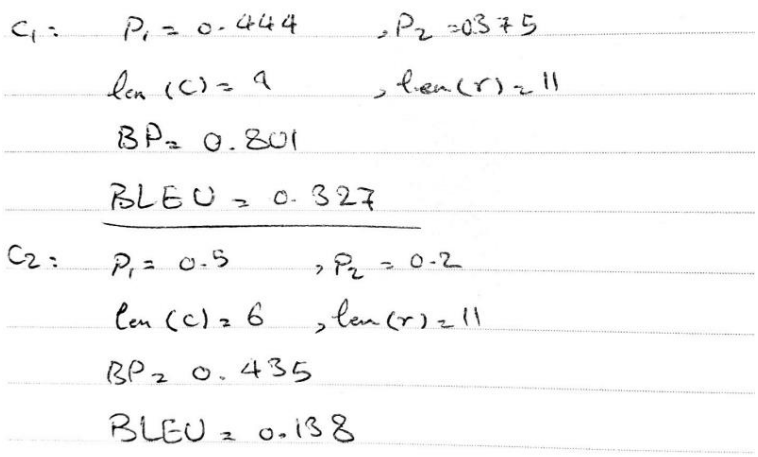
(i(c(۲



طبق BLEU، c2 بهتر است. هرچند موافق نیستم این بهتر باشد. جمله مناسبی نیست؛ این فقط یک رشته از کلمات با همپوشانی بالای n گرم است.

مهدیه نادری ۹۸۵۲۲۰۷۶

(ii(c(۲



طبق BLEU، c1 بهتر است. من با این موافقم، اگرچه این یک نمونه جالب است که در آن ترجمه های مرجع کمتر در واقع منجر به خروجی معقول تر می شود.

(iii(c(۲

اگر یک ترجمه مرجع ضعیف باشد، BLEU مقایسه بدی دارد. حتی اگر ترجمه تک مرجع با کیفیت بالا باشد، BLEU ممکن است واریانس کاملاً منطقی را در ساختار جمله خروجی، ترتیب کلمات و غیره دریافت نکند. خروجی "min" غیر صفر باشد. این به ما امتیاز آماری معنی داری می دهد.

(iv(c(۲

مزایای BLEU این است که (1) ارزان و (2) قابل تکرار است، یعنی به واریانس ارزیابی انسانی بستگی ندارد. معایب آن این است که (1) معمولاً به یک مرجع تکیه می کند، در حالی که انسان ها می توانند از بسیاری از مراجع احتمالی استفاده کنند و (2) نمی توانند کلمات مختلفی را که از نظر معنایی معادل هستند، مانند "can not" و "cannot" که یک انسان به راحتی تشخیص می دهد ، پیدا کنند.