

به نام خدا



دانشگاه تهران



دانشکده مهندسی برق و کامپیوتر

درس شبکه‌های عصبی و یادگیری عمیق

تمرین امتیازی

| | | |
|-------------|--------------------|--------|
| آرین فیروزی | نام و نام خانوادگی | پرسش ۱ |
| 810100196 | شماره دانشجویی | |
| آرمان مجیدی | نام و نام خانوادگی | پرسش ۲ |
| 810100205 | شماره دانشجویی | |
| ۱۴۰۱.۱۱.۱۴ | مهلت ارسال پاسخ | |

فهرست

| | |
|---|--|
| پرسش ۱ ۱ | تنظیم دقیق مدل های زبانی بزرگ برای گفتگو در زبان فارسی |
| ۱ ۱ | دادگان و انتخاب مدل |
| ۳ ۳ | Soft Prompt |
| ۵ ۵ | LoRA |
| ۷ ۷ | تغییر وزن برخی لایه ها |
| ۱۲ ۱۲ | جمع بندی |
| پرسش ۲ - تولید کپشن برای تصاویر (Image Captioning) ۱۳ | |
| ۱۳ ۱۳ | آماده سازی دیتاست |
| ۱۳ ۱۳ | انتخاب مجموعه داده |
| ۱۳ ۱۳ | پیش پردازش تصاویر |
| ۱۴ ۱۴ | پیش پردازش متن (Captions) |
| ۱۴ ۱۴ | تقسیم داده ها |
| ۱۷ ۱۷ | نمایش داده های پردازش شده |
| ۱۷ ۱۷ | CNN-RNN |
| ۱۷ ۱۷ | طراحی مدل |
| ۱۷ ۱۷ | پیاده سازی بخش رمزگذار (Encoder) |
| ۱۷ ۱۷ | پیاده سازی بخش رمزگشای (Decoder) |
| ۱۷ ۱۷ | اتصال رمزگذار و رمزگشای (Encoder-Decoder) |
| ۱۷ ۱۷ | آموزش مدل |
| ۱۸ ۱۸ | پیاده سازی مدل Attention based CNN-RNN |
| ۱۹ ۱۹ | CNN-Transformer |
| ۲۴ ۲۴ | بخش امتیازی |
| ۶-۲ | 6-2 |

شکل‌ها

پرسش 1

شکل 1.1: دو نمونه از فرمت دادگان در slimOrca

شکل 1.2: نمودار خطاب برای مدل PEFT Prompt Tuning

شکل 1.3: نمودار خطاب برای مدل LoRA

شکل 1.4: نمودار خطاب برای مدل سنتی

پرسش 2

شکل 2.1. 5 تصویر تصادفی به همراه کپشن آن‌ها

شکل 2.2. 5 تصویر تصادفی به همراه یکی از کپشن آن‌ها

شکل 2.3. نمودار پراکندگی (scatter plot) طول کپشن‌ها در دیتابست

شکل 2.4. هیستوگرام کلمات پرکاربرد

شکل 2.5.تابع هزینه برای مدل CNN-RNN

شکل 2.6. 5 تصویر به همراه کپشن تولیدی مدل CNN-RNN

شکل 2.7. تابع هزینه برای مدل Attention based CNN-RNN

شکل 2.8. 5 تصویر به همراه کپشن تولیدی مدل Attention based CNN-RNN

شکل 2.9. تابع هزینه برای مدل CNN-Transformer

شکل 2.10. 5 تصویر به همراه کپشن تولیدی مدل CNN-Transformer

شکل 2.11. تابع هزینه برای مدل جدید CNN-Transformer

شکل 2.12. تصویر به همراه کپشن تولیدی مدل جدید CNN-Transformer

شکل 2.13. تصویر به همراه کپشن تولیدی مدل جدید CNN-Transformer

شکل 2.14. تصویر به همراه کپشن تولیدی مدل جدید CNN-Transformer

شکل 2.15. تصویر به همراه کپشن تولیدی مدل جدید CNN-Transformer

شکل 2.16. تصویر به همراه کپشن تولیدی مدل جدید CNN-Transformer

جدول‌ها

پرسش 1

جدول 1.1: پاسخ مدل‌ها با فاین تیون مختلف به پرسش‌های یکسان

جدول 1.2: مقایسه تعداد پارامتر‌های قابل آموزش هر مدل

پرسش 2

پرسش ۱. تنظیم دقیق مدل های زبانی بزرگ برای گفتگو در زبان

فارسی

۱-۱. دادگان و انتخاب مدل

دادگان مورد استفاده در این تمرین دیتاست slimOrca فارسی با حدود 50000 داده است. این دادگان شامل چندین گفت و گو با GPT-4 است که هر یک از داده ها شامل ۳ بخش است:

- پaramp: یک پارامپت معمولاً کوتاه با role سیستم که در واقع نحوه‌ی پاسخ دادن مدل را تعیین می‌کند. شامل اطلاعاتی برای تنظیم مدل و قرار دادن آن در context سوال است.
- پارامپت کاربر: یک پارامپت معمولاً با طول بیشتر که role کاربر دارد و پارامپت اصلی کاربر است.
- جواب GPT-4: که برای مقایسه مدل و تنظیم دقیق آن به کار برده شده است.

نمونه‌ای از پرسش و پاسخ‌های این مدل در شکل ۱.۱ قابل مشاهده است.

| dataset.head(2) | | | | |
|-----------------|---|---|--|--------------|
| | 0 | 1 | 2 | |
| 0 | {'content': 'فاصله‌ها را اضافه...'} ...'کنید: باری‌های مصنوعی هستید. ب | {'content': 'شما یک دستیار هوش...'} ...'این اصطلاح ه...زیر: "این اصطلاح ه...است، همانطور ک | {'content': 'در اینجا عبارات جدا...'} ...'آیه به معنای نشانه...'} ...'این اصطلاح ه...زیر: "این اصطلاح ه...است، همانطور ک | None None |
| 1 | | | | |

شکل ۱.۱. دو نمونه از فرمت دادگان در slimOrca

برای مدل، ما مدل Gemma2-2b-it را انتخاب کردیم که مدل زبانی ارائه شده توسط گوگل برای گفت و گو است که در چندین سایز و نوع ارائه می‌شود. دلیل انتخاب Gemma به جای llama این بود که دسترسی کاربران ایرانی به مدل دوم محدود است و همچنین متأثر برای دسترسی به llama باید مجوز بدهد.

همانطور که پیشتر گفته شد، مدل Gemma در سایز‌های مختلفی ارائه می‌شود (از 27 میلیارد تا 2) که ما کوچکترین مدل را به علت محدودیت پردازشی انتخاب کردیم. همچنین مدل‌های زبانی معمولاً برای گفت و گو تولید نمی‌شوند، بلکه ابتدا برای تکمیل متن آموخته داده می‌شوند و سپس نسخه‌ی قابل گفت و گو تولید می‌شود که در واقع خودش یک پردازش دقیق از نسخه اصلی است. از جایی که ما تسلیک گفت و گو داشتیم و دادگان طبق آن داده شده بود، نسخه instruct قابل استفاده بود.

با این که کوچکترین سایز انتخاب شده بود، اما اینفرنس مدل همچنان کند بود و حتی در مرحله تست که تنها یک پرامپت میگرفت، مدت زمان زیادی برای جواب دادن صرف میکرد، در نتیجه تصمیم گرفتیم مدل را به صورت quantize شده استفاده کنیم. این پروسه یکی از راه های کم کردن حجم مدل با tradeoff کمی از دقت است و کارکرد کلی آن این است که اعداد وزن ها را تا چند اعشار قطع میکنیم که حجم مدل و پیچیدگی محاسبات آن کاهش پیدا کند. برای کوانتزیشن از کانفیگ bits and bytes استفاده شد و مدل را به صورت 4bit-quantized لود کرده، برای آموزش با این شرایط آماده کردیم.

آماده سازی دادگان:

توکنایزر مورد استفاده ما ابتدا توکنایزر ارائه شده توسط خود مدل بود که تبدیل دادگان به فرمت قابل قبول برای توکنایزر (دادگان 3 بخش داشتند و باید یک string به توکنایزر میدادیم) با مشکلاتی همراه بود. در نهایت طبق راهنمایی دستیاران آموزشی، از توکنایزر chatml استفاده کردیم. با استفاده از این توکنایزر، میتوان داده ها را با استفاده از یک دیکشنری که نقش مدل و پرامپت مربوط به آن را وارد کرد، که این مشابه دادگان ما بود، در نتیجه به راحتی میشد دادگان را به فرمت قابل تزریق به مدل تبدیل کرد. پس از انجام این تغییر، دادگان را از طریق کتابخانه pandas خواندیم و پس از اعمال توکنایزر آن را به dataframe تبدیل کردیم. دادگان به نسبت 2 به 10 برای ولیدیشن و آموزش جدا شدند.

تست مدل:

برای تحلیل مدل، چند پرامپت به زبان فارسی نوشتیم و بعد از تبدیل آن به توکن، با مدل امتحان کردیم. نتیجه این تست در جدول 1.1 وارد شده است.

مدل در حالت اولیه، در عین تولید جواب های قابل قبول، مشکلاتی برای زبان فارسی دارد. به عنوان مثال، مشکل تکرار یک کلمه که از مشکلات رایج هنگام آموزش مدلهاست، گاها تولید جملات نامفهوم و همچنین تولید خروجی انگلیسی برای ورودی فارسی (تا نصف جمله فارسی میگفت بعد انگلیسی).

برای تولید پرامپت های تست، تلاش کردیم سوالاتمان انواع مختلف پرسش و پاسخ ها را شامل شود و جنبه های مختلف مدل مورد آزمون قرار گیرد. مشخصا برای تبدیل پرامپت به توکن نقش کاربر به توکنایزر داده شد.

Soft Prompt .۱-۲

پاسخ به سوالات (از سایت (huggingface

- در روش soft prompt، برخلاف روش hard prompt که متن های آماده‌ی واقعی با ورودی گستره به مدل داده می‌شود و تولید پرامپت‌های خوب بسیار زمانگیر است، از تنسور‌های قابل آموزش استفاده می‌شود که پیوسته بوده و به همراه embedding‌ها به مدل داده می‌شوند. این روش سریعتر است، ولی این virtual token‌ها قابل خواندن توسط انسان نیستند.
- توضیح سه روش به اختصار به شرح زیر است:
 - در این روش پارامتر‌های مدل دست نخورده باقی می‌مانند و فقط توکن‌های مربوط به امبدینگ‌ها تغییر پیدا می‌کنند. به همین خاطر است که از جمله روش‌های بهینه و سریع برای فاین‌تیون مخصوص می‌شود.
 - مشابه روش قبل است، ولی این بار توکن‌های امبدینگ را به تمام لایه‌های مدل وارد می‌کنیم. با این عمل، فاین‌تیون تقریباً به اندازه فاین‌تیون کامل مدل موثر است، در حالی که بیش از هزار برابر کاهش پارامتر داشتیم.
 - در این روش مثل روش‌های قبلی تنسور را به توکن‌ها اضافه می‌کند، اما در این حالت از یک انکودر برای پارامتر‌ها استفاده می‌کنیم که میتواند نوعی از مموری مثل lstm باشد، و همچنین توکن‌های پرامپت در لایه‌ی دلخواه میتواند وارد شود و محدود به استفاده آن در لایه اول نیستیم.
- با توجه به توضیحات بالا، prompt tuning ساده‌ترین و در عین حال efficient ترین راه حل از لحاظ پردازشی است و به دلیل مشکلات سخت افزاری، این روش را انتخاب کردیم تا جواب سریعتری بگیریم.

پیاده سازی:

برای پیاده سازی از کتابخانه PEFT استفاده کردیم و کانفیگ prompt tuning با 20 توکن مجازی (که به نظر حجم مناسبی بود) را به همراه مدل کوانتایز شده به مدل کتابخانه دادیم. پس از این عمل، تنها 46,080 پارامتر از حدود دو میلیارد و شصصد میلیون پارامتر مدل قابل آموزش بودند. برای آموزش از ضریب $5e^{-8}$ استفاده کردیم که ضریب رایجی برای فاین‌تیون است و طبق نتایج loss خوب عمل کرده، مدل را به مدت 10 ایپاک آموزش دادیمو نتایج آن را به استفاده از progress bar ذخیره کردیم. نتیجه‌ی آموزش در تصویر 1.2 قابل مشاهده است که ضریب خطای

ولیدیشن در آن با شبیب قابل قبولی کاهش پیدا کرده، ولی خطای دادگان آموزش افت و خیز زیادی دارد و با اینکه ترنند آن رو به پایین است، آموزش stable نیست.

آموزش ابتدا به صورت دستی انجام شده بود که زیاد استیبل نبود و همچنین اضافه کاری زیادی داشت (کد به صورت کامنت موجود است) در نتیجه از trainer برای انجام آموزش استفاده شد.



شکل 1.2. نمودار خطا برای مدل PEFT Prompt Tuning

یکی از مشکلات آموزش این بود که حجم gpu پر میشد و کرش میکرد و مجبور شدیم که سایز batch را یک قرار بدهیم و در واقع از روش بدون batch برای آموزش استفاده کردیم که مدت زمان بیشتری طول میکشید ولی حجم کمتری از gpu را اشغال میکرد. به علت زمان زیاد، از تمام دادگان برای آموزش استفاده نشد و بخشی از دادگان برای آموزش به کار گرفته شدند. دادگان استفاده شده به عنوان دیتا mini نوشته شده اند.

همچنین مدل بعد از آموزش توکن های ورودی را مثل قبل نمیتوانست دریافت کند و خروجی متنی از آن نمیتوانستیم بگیریم، راه حل این مشکل استفاده نکردن از cache بود که باعث میشد ورودی قبلی با ورودی بعد از توکن های مجازی تداخل پیدا کند و مدل inconsistent باشد و با استفاده از آپشن های generate هندل شد.

مشکل دیگری که تا حدودی رفع شد اما راه حل کامل برای آن پیدا نکردیم، stable نوشن آموزش بود که به علت batch نشدن ورودی بود. با اینکه ram سیستم توانایی تولید batch در همان ابتدا نداشت، ولی تبدیل لیست به batch کمتر فضای نیاز داشت، در نتیجه از یک data collator استفاده کردیم و ورودی را به صورت batch دادیم. این راه حل به خصوص برای LoRA خوب عمل کرد و در این بخش با اینکه تاثیر مثبت گذاشت ولی مشکل را به طور کامل رفع نکرد. احتمالاً با تغییر lr و افزایش سایز بچ بتوان نتایج بهتری گرفت.

یک مشکل دیگر هنگام آموزش، این بود که مدل به جای خروجی یک سری عدد رندم میداد که به علت initialize کردن اشتباه مدل و استفاده از lr اشتباه بود که با کمتر کردن ضریب و استفاده از batch حل شد.

در مورد بررسی پرامپت ها، همانطور که از نمودار خطای مشخص است، نتایج بهبود زیادی پیاده نکرده و مدل تصمیم گرفته به زبان انگلیسی پاسخ دهد. سوال مربوط به ترجمه به طور استثنایی خوب ترجمه شده ولی سایر جواب ها بی ربط هستند. یک مورد نسبتاً درست دیگر داستان است که به زبان انگلیسی گفته شده.

قابل توجه است که هنگام inference، از محدودیت طول پاسخ و همچنین یک پارامتر برای عدم تکرار کلمات استفاده کردیم و بخشی از بهبود های مدل به خاطر همین نوع inference بود.

LoRA ۱-۳

: پاسخ به سوالات (باز هم از huggingface

Low-Rank Adaptation • روشهای کاهش تعداد پارامتر های قابل آموزش در مدل هاست. طبق اظهار مقاله، این روش میتواند تا 10000 برابر تعداد پارامتر ها را کاهش دهد. در این روش تعدادی ماتریس rank decomposition به هر لایه ترانسفورمر اضافه میشوند (که روشهای تقسیم یک ماتریس بزرگ به ضرب چند ماتریس و در نتیجه کاهش تعداد پارامتر ها است) و وزن های مدل دست نخورده باقی میمانند.

این روش به لایه های attention و ffn در ساختار ترانسفورمر اعمال میشوند. •

متدهای مبتنی بر lora : •

DoRA ○ این روش برای بالا بردن performance لورا به کار برده میشود. در این راهکار فرایند به روز رسانی به دو بخش magnitude و direction تبدیل میشود.

آموزش lora داده میشود و بخش magnitude توسط پارامتر قابل آموزش دیگری انجام میشود.

روشی است که در آن از روش ضرب Low-Rank Hadamard Product :LoHa ○ هادامار استفاده میشود و ماتریس ها شبیه lora به ضرب چند ماتریس شکسته میشوند. این روش نسبت به لورا پارامتر های کمتری تولید میکند و دقت آن نیز قابل مقایسه با لورا است.

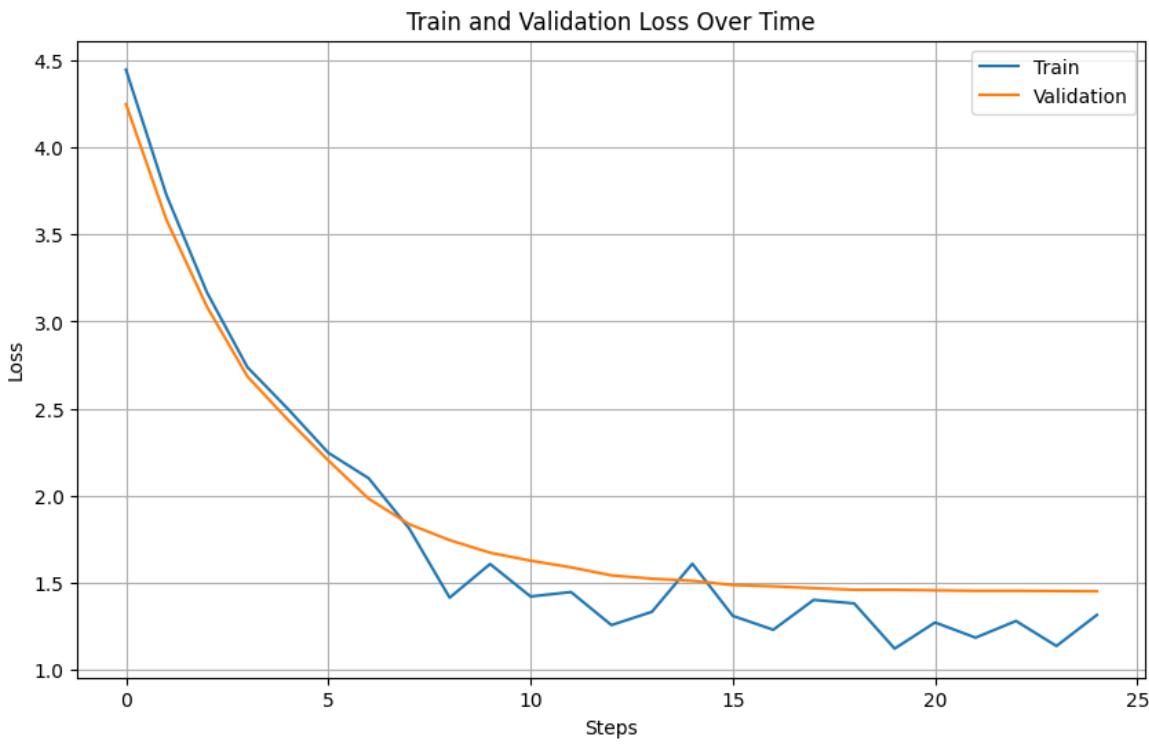
در این روش یک فاکتور rank stabilization برای اسکیل معروفی میشود که برای مشکل افزایش رنک آداتر هایی که مدل اضافه شده اند به کار میرود. آداتر ها همان لایه هایی هستند که توسط lora ایجاد شدند و رنک تعداد سطر های مستقل در یک ماتریس است، در این روش ماتریس ها به جذر رنکشان تقسیم میشوند و دوباره کاهش بعد پیدا میکنند تا سرعت کاهش پیدا نکند.

آموزش:

برای آموزش از روش LoRA استفاده کردیم چون متدهای بوده و هم نتیاج خوبی دارد و هم منابع خوبی برای آموزش دارد.

با استفاده از کتابخانه PEFT کانفیگ لورا با رنک 16 و 0.1 dropout (جهت overfit نشدن) استفاده کردیم. تعداد پارامتر های قابل آموزش 3,194,880 گزارش شد که حدود 0.1 درصد از کل پارامتر هاست. فرآیند آموزش تقریبا بدون دردرس بود و با استفاده از پارامتر های مشابه قسمت قبل انجام شد و نتیجه ی به مرتب بهتری داشت. به علت بیشتر بودن تعداد پارامتر های قابل آموزش، همانطور که در تصویر 1.3 مشاهده میشود روند کاهش خطابه مرتب سریعتر از راه قبلی است. شبیب نمودار نشان میدهد که در 5 ایپاک مدل تقریبا فاین تون شده در حالی که در حالت قبلی شبیب همچنان در حال کاهش بود.

استفاده از data collator همانطور که قبلا گفته شد در این بخش نتیجه بهتری نسبت به peft داشت.



شکل 1.3. نمودار خطا برای مدل LoRA

تحلیل نتایج:

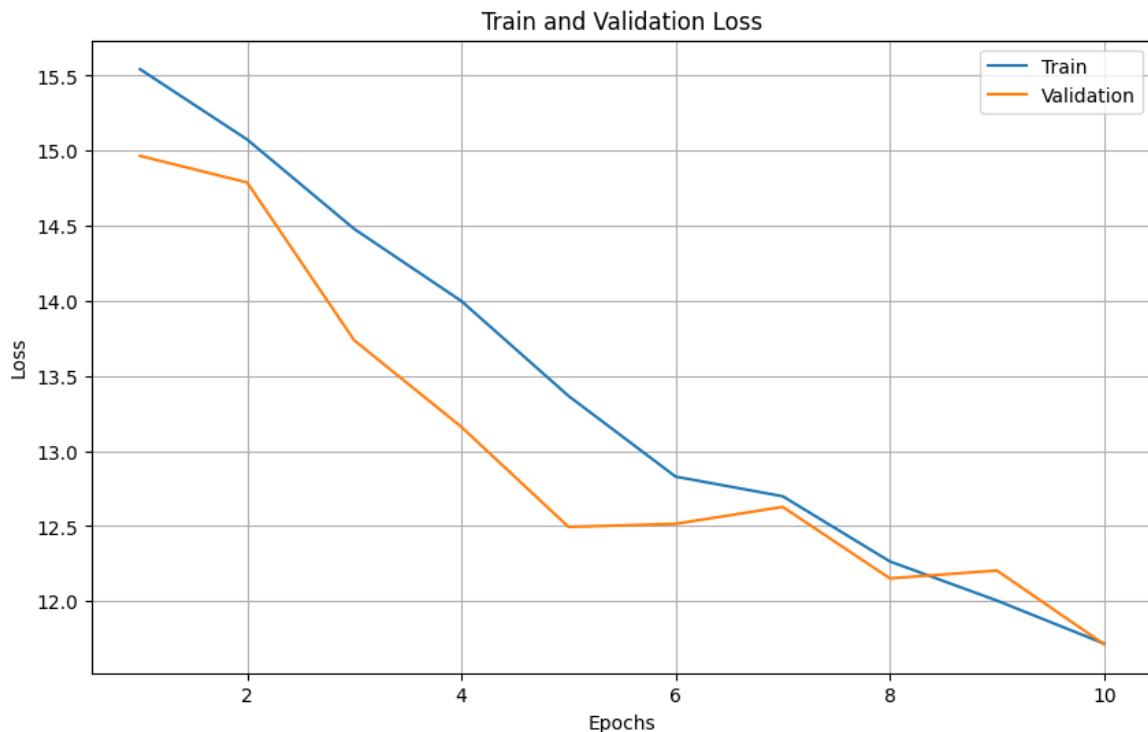
بهترین نتایج در این بخش حاصل شد. نتایج لورا کاملاً مربوط بودند (با یک استثنای کلمه‌ی قرار را نتوانست معنی کند) و هم در کار تحلیلی، هم در داستان و شعر، هم یادآوری وقایع و هم ترجمه متن عملکرد خوبی داشت. جمله‌بندی‌ها اشکال گرامری قابل توجهی نداشت و متون بهتر از مدل پایه بودند.

4-1. تغییر وزن برخی لایه‌ها

در این بخش سعی کردیم بدون استفاده از روش‌های PEFT و با راه سنتی فاین تیون تعداد لایه را فریز کنیم و تنها لایه اول و آخر را آموزش دهیم. لایه اول و آخر 264,192 پارامتر داشتند قابل آموزش داشتند که به مراتب کمتر از حالت lora است.

در این بخش از آپتیمايز AdamW استفاده کردیم و مدل را مانند تمرین‌های قبلی فاین تیون کردیم. برای evaluate کردن مدل فانکشن جدا گانه‌ای نوشته شد و بدون تغییر وزن لایه‌ها مدل را کردیم. با این که تعداد پارامتر‌ها کمتر بود، زمان تقریباً با متدهای دیگر یکسان بود و هم چنین شبیه نمودار خطا همانگونه که در تصویر 1.4 قابل مشاهده است، کمتر از متدهای قبلی بود و شبیه کمتر نشده بود، یعنی به مدت زمان بیشتری برای آموزش نیاز داشت.

از جمله چالش‌های پیش آمده در این بخش قابل آموزش کردن لایه‌ها بود که چون تنها لایه‌هایی که تعریف شده بودند میتوانستند از گرادیان استفاده کنند این شرط را در آزاد سازی لایه‌ها قرار دادیم.



شکل 4. نمودار خط‌با برای مدل سنتی

نتایج این بخش به صورت جزئی از مدل پایه بهتر است اما فرق زیادی بین این دو دیده نمی‌شود. برای مثال مشکل تغییر زبان در این بخش دیده نشد و جملات کمی ساختار یافته‌تر بودند. این رفتار مورد انتظار ما بود چرا که روش‌های سنتی در فاین تیون مدل‌های زبانی به داده‌ها و زمان زیادی نیاز دارند و دلیل استفاده از lora و prompt tuning همین ضعف است.

جدول 1.1 پاسخ مدل ها با فاین تیون مختلف به پرس های یکسان

| <i>Method</i> | <i>prompt</i> | <i>output</i> |
|-----------------|--|--|
| <i>original</i> | یک داستان کوتاه بگو | ###'درون غروب آتش غروب با سرعت فرا می رسد. با این حال، در غروب یک روز مغز آن چیزی که در دیوار یونانی خفته بود به صورت یک گپ بازار بود. او با تفاوت از خواب نمی انسنت که این غروب در نفس شکل کجا می تواند باید باشد. در آن حالت عجیب در مغز آن غروب از یک گپ بازار نیاز به یک نقطه نگاه می بازار بود که به صورت یک پرده بازار می شودپرده بازار با هر گپ بازار نیاز به یک نقطه نگاه می داد. شکل یونانی به سادگی جای یک صورت تکرار شده بود. با هر گپ بازار به سادگی جای یک بصورت تکرار شده بوداو با تفاوت از خواب می داشت که این غروب در نفس شکل کجا می تواند باید باشد. در آن حالت عجیب در مغز آن غروب از یک گپ بازار بود که به صورت یک پرده بازار می شود. پرده بازار با هر گپ بازار نیاز به یک نقطه نگاه می داد. شکل یونانی به سادگی جای یک بصورت تکرار شده بود.'. |
| | در صورتی که من بخواهم به شهر رشت در ایران سفر کنم، به نظرت چه لباس هایی باید بردارم؟ | 'سلام\n! برای سفر به رشت، لباس هایی که بردارید به عوامل مختلفی بستگی دارد، از جمله*\n*: زمان سفر:*** چه فصلی را انتخاب کردید؟ در فصل تابستان، نیاز به لباس های سبک و راحت تر دارد، اما در فصل زمستان، لباس های گرم و بُشنده ای ضروری خواهد بود***\n*غذا و تفریح:***در رشت، چه نوع تفریحاتی انجام خواهید داد؟ اگر به دریاچه ها و beaches می روید، لباس های وزشی و swimwear را با خود ببرید. اگر به art galleries و museums می روید، لباس های راحت و مناسب برای تماشا و بازدید از آثار هنری را انتخاب کنید***\n*مدت زمان سفر:***چه مدت زمان در رشت هستید؟ در صورت کوتاه مدت، لباس های مختصر و ساده را انتخاب کنید. در صورت سفر طولانی مدت، لباس های متنوع و مناسب برای هر روز را انتخاب کنید***\n*لباس هایی که معمولاً برای سفر به رشت استفاده می شود***در فصل تابستان * \n***لباس های ساده و راحت *\n*کلاه و عینک آفتابی *\n*کفش های راحت *\n*ashlolar و دامن *\n*لباس های تیشرت و شورت ***در فصل زمستان *\n*لباس های گرم و بُشنده ای *\n*ashlolar و دامن *\n*کلاه *\n*و عینک آفتابی *\n*کفش های مناسب برای باران و برف *\n*لباس هایی مناسب برای فکر بر نکات مهم *\n***لباس های سپیار روند و مخلوط استفاده می شود***در رشت، می توانید با لباس های سپیار روند و مخلوط استفاده شود***لباس هایی برای می شود***در رشت، می توانید با لباس های برای اسفاده می شود***کفش های سپیار می شود***در رشت، می توانید با کفش های می شود\n*با توجه به این اطلاعات، لباس هایی را انتخاب کنید که با هوا و شرایط در رشت سپیار می شود!' |
| | یک شعر فارسی بنویسید | ###'شبی در رقص\nashbi در رقص، با کوتaran \nاز خورشید شب بر سفر\nادر خشم نفس بر مغز\nادر آسمان سگ و مروارید\nاسنگ بر سنگ کوبید در تردید\nاعشق در آسمان زبان در خورشید\nابا تنهایی در باران خسته\nاو با غم در سفر\nادر خورشید شب کوتار بر سفر\nادر خشم نفس بر مغز\nادر آسمان سگ و مروارید\nA\ntranslation:**\n#\nA Night in Dance\nA night in dance, with the robins\nFrom the sun's night flight,\nacross the land\nIn the heart of the storm, a breath in the mind\nIn the sky, a dog and a diamond\nStones against stones, a doubt is born\nLove in the sky,\nlanguage in the sun\nWith loneliness in the rain, tired\nAnd with grief in the journey,\nIn the sun's night, a robin across the land\nIn the heart of the storm,\na breath in the mind\nIn the sky, a dog and a diamond\nExplanation:**\n\nThe poem uses imagery and metaphor to paint a picture of a night full of contrasts and emotions.\n\n***"Dance"*** represents a sense of movement and energy, while also being symbolic of life's journey.\n\n***"Robins"*** symbolize hope and joy, despite the darkness surrounding them.\n\n***"Sun's night flight"*** evokes a sense of escape and longing.\n\n' |

"Heart of the storm" represents a difficult and intense period in life.\n* ***"Dog and diamond"*** are symbols of duality, strength and fragility.\n* ***"Stones against stones"*** symbolizes the struggle and conflict within oneself.\n* ***"Love in the sky"*** represents a yearning for something more, while ***"language in the sun"*** suggests the need for communication and understanding.\n* ***"Loneliness in the rain"*** evokes a sense of isolation and melancholy.\n\nThe poem's

ملکه الیزابت کیست؟

البیزابت علاوه بر ان الیزابت الز
الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز
الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز
الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز
الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز
الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز
الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز
الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز
الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز الز

این جمله را به انگلیسی
ترجمه کن: من یک مدل
زبانی با قابلیت های
پیشرفته هستم

'شما یک مدل زبانی هستید که قابلیتها پیشرفته ای دارید.'

PEFT Prompt Tuning

یک داستان کوتاه بگو

The Last Book

The old bookstore was tucked away in a quiet corner of the city, its worn wooden shelves holding the weight of countless stories. Inside, a young woman named Elara sat, her fingers tracing the spines of forgotten books. She was a collector, not of rare first editions or signed copies, but of stories. Stories that whispered of forgotten dreams, of lost loves, of the quiet magic of words.

در صورتی که من بخواهم
به شهر رشت در ایران سفر
کنم، به نظرت چه لباس
هایی باید بردارم؟

It's great you're planning a trip to Shiraz! Here are some clothing ideas for your trip to Iran:\n**For the weather:** Shiraz has a pleasant climate, but it can get a bit chilly in the evenings. \n**What to pack:**\n**Comfortable pants or jeans:** These are"

یک شعر فارسی بنویسید

در دل در دل، در دل امی رو، می رو*\n*در قلب، در قلب،\n*در قلب، در قلب،*\n*این شعرها به زبان فارسی، درک می باشد.\n*می می*\n*توضیح*\n*این شعرها به زبان فارسی، درک می باشد.\n*

ملکه الیزابت کیست؟

I'm sorry, I don't understand. Can you please rephrase your question? 😊

I am here to help you, so please let me know what you need! 😊

این جمله را به انگلیسی
ترجمه کن: من یک مدل
زبانی با قابلیت های
پیشرفته هستم

I am a sophisticated language model with advanced capabilities. \n\nLet me know if you'd like to try a specific task or have any questions. "

معنی کلمه قرار در جمله
روبرو چیست؟ خودکار
روی میز قرار دارد

The sentence is: ***"The chair is on the table."*** \n\nThe sentence translates to:
"\n\"The table is on which is the chair." \n\n"

LoRA finetuning

یک بار، در یک جزیره دور در اقیانوس، یک زن جوان به نام ماریا زندگی می‌آورد که به شدت به دنیال یک شغل جدید بود. هر روز، در حین انجام وظایفی که به او داده می‌شود، به خود فکر می‌کند و آرزو می‌کند 'اگر'

برای سفر به رشت، توصیه می‌کنم لباس‌های سبک و راحت انتخاب کنید. هوا در رشت عمولاً گرم و مرطوب است، بنابراین لباس‌هایی با رنگ‌ها روش انتخاب کنید که به راحتی

در شب‌های تاریک، گهگاهی یک ستاره، ادر آسمان غروب، یک شمعی می‌سوزد.
این یک شعر فارسی بنویسید.

ملکه الیزابت، که در سال ۱۹۵۲ به عنوان ملکه انگلستان منصوب شد، آخرین ملکه بریتانیا است. وی در تاریخ ۲۳ مارس ۱۹۲۵ به دنیا آمد و در تاریخ ۲۱ مارس ۲۰۲۲ در سن ۹۶ سالگی درگذشت.

این جمله را به انگلیسی ترجمه کن: من یک مدل زبانی با قابلیت‌های پیشرفته هستم

Here's the translation: I am a state-of-the-art language model."

کلمه "قرار" در این جمله به معنای "در جایی که قرار می‌گیرد" است. به این معنی است که خودکار در جایی قرار دارد که به آن اجازه داده شده تا قرار گیرد.

Classic finetuning

در گوشه‌ای تاریک و عمیق کتابخانه، یک کتاب با جلد سبز خسته و خسته در گوشیده از سال ۲۰۰۰ می‌توانید آن را در هوا گرم و خشک در این شهر بپوشید. یک غبار بی‌نظم و شکننده تبدیل شده بود. در پشت جلد، یک نوشته کوچک و خسته در زیر یک تنہ ای نازک و فرسوده بود که زبان یکدست خط گرفته بود: "فاراهم کرده."

به شما پیشنهاد می‌کنم که برای سفر به رشت در ایران، به فکر لباس‌هایی باشید که به راحتی می‌توانید آن را در هوا گرم و خشک در این شهر بپوشید. معمولاً هوا کمی گرم است. برای اینکه به خوبی در این هوا گرم و خشک احساس راحتی کنید، لازم نیست که لباس‌های سنتی و خسته کننده بپوشیدید. initComponents است. در زیر آمده است:*

آن را در حال رخ می‌دانید و به زمین می‌رسد. یک کودک در یک اتاق کوچک و تاریک، سعی در انجام یک کار پیچیده دارد.

یک بخش از ایده‌ها برای یک شعر فارسی به شما می‌دانم که می‌توانید از آن را استفاده کنید:
آن را طوفان در حال رخ می‌دانید و به زمین می‌رسد.

این جمله را به انگلیسی ترجمه کن: من یک مدل زبانی هستم که قابلیت‌های پیشرفته را دارا هستید.

معنی کلمه قرار در جمله خودکار روی میز قرار می‌گیرد.

1-5. جمع بندی

تعداد پارامتر های مورد تغییر در هر بخش در جدول 1.2 مقایسه شده است.

بیشترین زمان آموزش توسط prompt tuning مصرف شد، در حالی که لورا با نتیاج خیلی بهتر مدت زمان یک سوم و فاین تون عادی یک دوم آن زمان برداشت (البته لورا نصف ایپاک های سایر ترین شد ولی چون به وضعیت پایدار در آمد از ادامه آموزش صرف نظر کردیم). با بررسی جدول 1.2 متوجه میشویم که LoRA تعداد پارامتر بسیار بیشتری دارد اما به علت استفاده بهینه از پارامتر ها که با استفاده از Rank ماتریس ها که قبلاً توضیح داده شد انجام پذیر است، سرعت آموزش بالاتری داشته است.

همچنین سرعت inference مدل ها متغیر بود و روش prompt tuning و LoRA حدود 30 ثانیه الى یک دقیقه برای اینفرنس زمان نیاز داشتند در حالی که روش کلاسیک چون overhead پارامتر های اضافی نداشت، حدود دو برابر سریعتر عمل میکرد.

عملکرد مدل ها در بخش خودشان تحلیل شد.

مزیت روش کلاسیک آسانتر بودن و سریار کمتر است، ولی نتایج جالبی تولید نمیکند و به آموزش طولانی مدت نیاز دارد.

مدل لورا پیاده سازی نسبتاً پیچیده ای دارد اما در عوض با منابع کم، نتایج بسیار خوبی تولید میکند.

مدل prompt tuning ساده تر از لورا پیاده میشود اما دقیق تر کمتری دارد و برای آموزش منابع بیشتری مصرف میکند.

بهترین روش همانطور که از نتایج مشخص است، LoRA بود.

جدول 1.2 مقایسه تعداد پارامتر های قابل آموزش هر مدل

| percentage | parameters | model |
|------------|------------|---------------|
| 0.0018% | 46,080 | Prompt Tuning |
| 0.1221% | 3,194,880 | LoRA |
| 0.0165% | 264192 | Classic |

پرسش 2 – تولید کپشن برای تصاویر (Image Captioning)

2-2. آماده‌سازی دیتاست

انتخاب مجموعه داده

برای این قسمت، دیتاست را از سایت Kaggle دانلود می‌کنیم. دیتاست‌های متفاوت و فراوانی با این نام توسط توسعه‌دهنگان متفاوتی آپلود شده‌است. برای این پرسش، دیتاستی که توسط توسعه‌دهنده اصلی آپلود شده‌است را استفاده می‌کنیم. به علت استفاده از دو محیط کگل و کولب، توابع مختلفی برای لود کردن دیتاست استفاده شده‌است. حال، به خواست سوال، ۵ تصویر به همراه کپشن متناظر با آن آپلود شده‌است. شکل 2.1 و 2.2 این تصاویر را نشان می‌دهد.



شکل 2.1 ۵ تصویر تصادفی به همراه کپشن آنها

پیش‌پردازش تصاویر

برای پیش‌پردازش تصاویر، در این قسمت تنها سایز عکس را تغییر می‌دهیم و از تغییر میزان شدت پیکسل‌ها خودداری می‌کنیم. زیرا که در سایت keras برای مدل pre-trained EfficientNetB0 شده نوشته شده بود که مقادیر ورودی شدتی بین 0 تا 255 دارند. در نتیجه نرم‌افزاری NLP انجام نمی‌دهیم. البته در بخش CNN-Transformer میانگین و واریانس را استانداردسازی می‌کنیم.

پیش‌پردازش متن (Captions)

تمام فعالیت‌های ذکر شده در تابع preprocess_captions انجام شده‌است. در این تابع، خط به خط فایل txt خوانده می‌شود، سپس طبق نحوه نوشته شدن فایل txt، دیکشنری image_captions ساخته می‌شود. در این دیکشنری، نام عکس‌ها به عنوان key های دیکشنری و کپشن‌ها به عنوان value های دیکشنری قرار دارند.

سپس پس از تولید captions_dictionary، کلمات را به شناسه عددی تبدیل می‌کنیم. این عملیات به کمک تابع tokenize_and_map_with_special_tokens انجام می‌گردد. در این تابع توکن‌های ویژه در نظر گرفته می‌شوند. همچنین، در انتهای فایل json با نام word_to_id ذخیره می‌گردد. این فایل در پیوست آمده است.

دلیل استفاده از طول ثابت برای کپشن‌ها این است که مدل زبانی ما نمی‌تواند از طول کپشن متغیر استفاده کند. زیرا که دیکودر این مدل‌ها یک خروجی با اندازه ثابت می‌دهد؛ در نتیجه باید این مشکل را به نحوی هندل کنیم. برای این کار، از اسپشیال توکن‌های ذکر شده در پروژه استفاده می‌کنیم.

تقسیم داده‌ها

برای آن که ترتیب عکس‌ها و کپشن‌ها یکی شود، این دو دیکشنری را مرتب می‌کنیم تا ترتیب آن‌ها یکی شود و عملیات تقسیم آن‌ها به سه داده آموزش، ارزیابی و تست ساده گردد. در این صورت هیچ عکس تکراری میان این داده‌ها نخواهد بود. این عملیات توسط `sort_and_split_data` انجام می‌گردد.

نمایش داده‌های پردازش شده

شکل 2.2 این 5 تصویر به همراه یکی از کپشن‌ها را نشان می‌دهد.

دو سگ در یک زمین باریک نمایندگان را می‌بینند



ady in orange is running



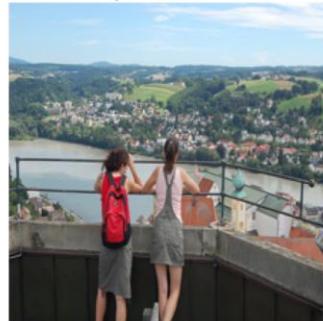
ady speaking in a microphone and another taking notes



dark skinned man standing outside a small yellow building which is setup to sell items

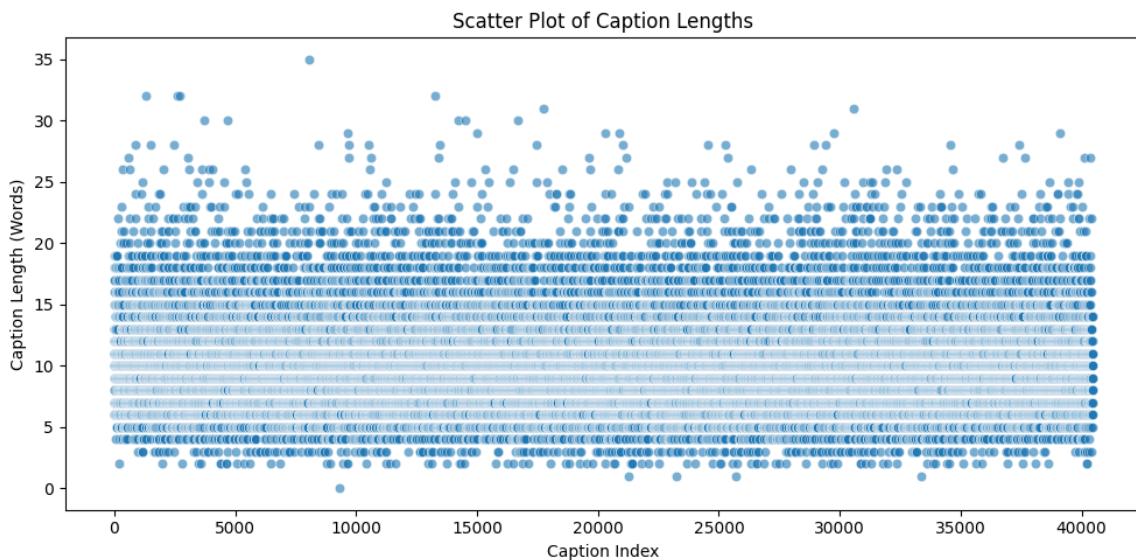


two girls take in the view



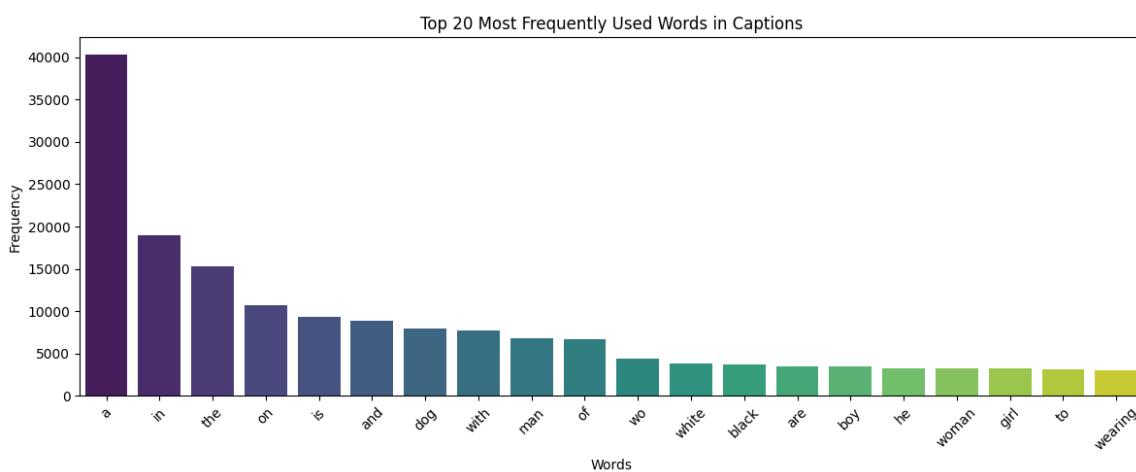
شکل 2.2.5 تصویر تصادفی به همراه یکی از کپشن آنها

شکل 2.3 نیز بیانگر نمودار پراکندگی (scatter plot) طول کپشن‌ها در دیتاست را نشان می‌دهد.



شکل 2.3. نمودار پراکندگی (scatter plot) طول کپشن‌ها در دیتاست

حال، هیستوگرام کلمات پرکاربرد در این دیتاست را بررسی می‌کنیم. شکل 2.4 این هیستوگرام را نشان می‌دهد.



شکل 2.4 هیستوگرام کلمات پرکاربرد

3-2. پیاده‌سازی CNN-RNN

طراحی مدل

پیاده‌سازی بخش رمزگذار (Encoder)

برای قسمت بخش رمزگذار از یک مدل پیش‌آموزش داده شده استفاده می‌کنیم. برای این که مدل قابلیت پیاده‌سازی و استفاده برای تسک ما را داشته باشد، لایه آخر را حذف می‌کنیم و لایه‌ای مطابق با تسک خود قرار می‌دهیم. خروجی بخش دیکودر، [1, 1280] می‌باشد. بخش دیکودر را غیرقابل آموزش می‌کنیم تا پارامترهای مدل pre-trained تغییر نکند.

پیاده‌سازی بخش رمزگشای (Decoder)

به جای استفاده از نمایش‌های تک بعدی (one-hot encoding) که بسیار پراکنده و کم کارا مد می‌باشند، از لایه Embeding استفاده می‌کنیم که کلمات را به بردارهای متراکم با ابعاد کمتر تبدیل می‌کند. این لایه روابط معنایی را یاد می‌گیرد و عملکرد مدل را بهبود می‌بخشد.

اتصال رمزگذار و رمزگشا (Encoder-Decoder)

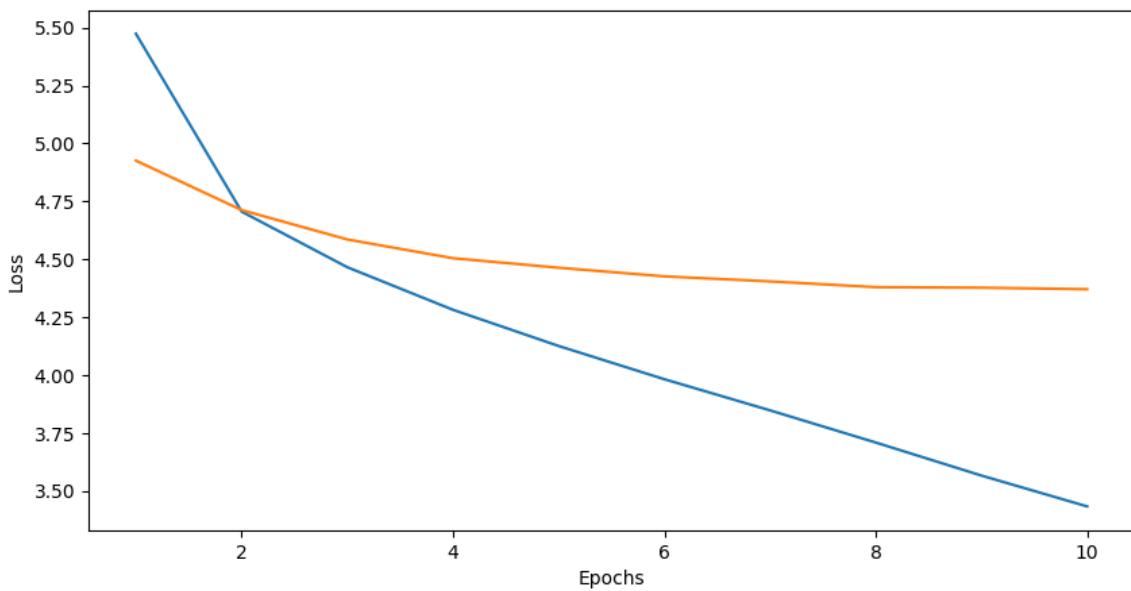
باید این دو کلاس را در یک کلاس مشترک استفاده کنیم که از nn.Module ارث بری کند. سپس باید تابع زیان مناسب را تعریف کنیم و سپس به آموزش این مدل پردازیم.

آموزش مدل

حال، مدل را آموزش می‌دهیم. به علت زمان بر بودن آموزش (حتی با gpu) دیتاست را محدود به تعدادی عکس می‌کنیم و به میزان تعداد معقولی ایپاک، مدل را آموزش می‌دهیم.

برای جلوگیری از بیش‌برازش، از early stopping، regularization، انتخاب معماری مناسب و ... استفاده می‌کنیم.

شکل 2.5 نمودار تابع هزینه را برای مدل CNN-RNN نشان می‌دهد.



شکل 2.5.تابع هزینه برای مدل CNN-RNN

همچنین، شکل 2.6 5 تا از تصاویر به همراه کپشن تولیدی را نشان می‌دهد.



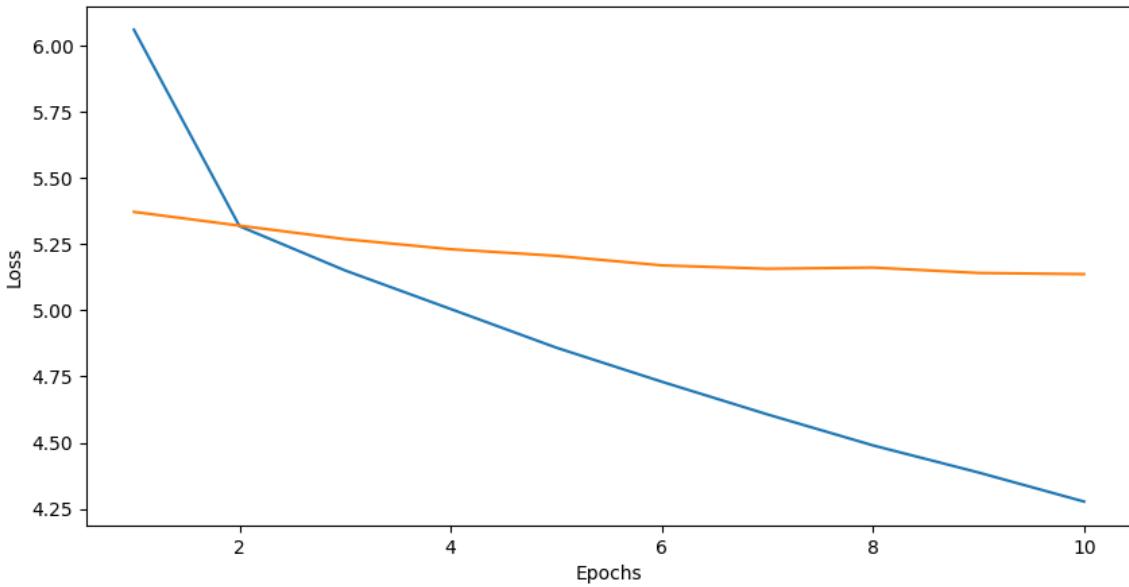
شکل 2.6. تصویر به همراه کپشن تولیدی مدل CNN-RNN

همانطور که مشاهده می‌شود، مدل قادر به جمله‌سازی درست نیست، اما می‌تواند عناصر درست داخل تصویر را شناسایی کند. همچنین، این مدل قادر به تشخیص حالات و ... نیست و تنها می‌تواند عناصر را به درستی تشخیص دهد. در نتیجه روابط در این مدل تشخیص داده نمی‌شوند.

4-4. پیاده‌سازی مدل Attention based CNN-RNN

در این قسمت خروجی قسمت انکورد سایزی برابر با [32, 38] دارد. در این قسمت از مکانیزم توجه استفاده شده است. مشابه مدل قبل، این مدل را نیز به علت زمان بر بودن آموزش (حتی با gpu) دیتابست را محدود به تعدادی عکس می‌کنیم و به میزان تعداد معقولی ایپاک، مدل را آموزش می‌دهیم.

شکل 2.7 نمودار تابع هزینه را برای مدل Attention based CNN-RNN نشان می‌دهد.



شکل 2.7. تابع هزینه برای مدل Attention based CNN-RNN

همچنین، شکل 2.8 ۵ تا از تصاویر به همراه کپشن تولیدی را نشان می‌دهد.



شکل 2.8 ۵ تصویر به همراه کپشن تولیدی مدل Attention based CNN-RNN

در این مدل خطاهای شناسایی بسیار کم و تقریباً صفر شده است. در مقایسه با قسمت قبل، متوجه می‌شویم که مدل بهبود عملکرد فوق العاده‌ای داشته است و پیشرفت چندانی کرده است.

2-5. پیاده‌سازی CNN-Transformer

برای این قسمت، از دو رویکرد و دو مدل متفاوت استفاده کرده‌ایم. در مدل اول، با رویکرد خودمان اقدام به پیاده‌سازی مدل دلخواه کرده‌ایم. به دلیل اینکه نتایج ما از نتایج بخش قبل بهتر نشد، از نوت‌بوک آپلود شده در ایلرن استفاده می‌کنیم و از آن کمک می‌گیریم.

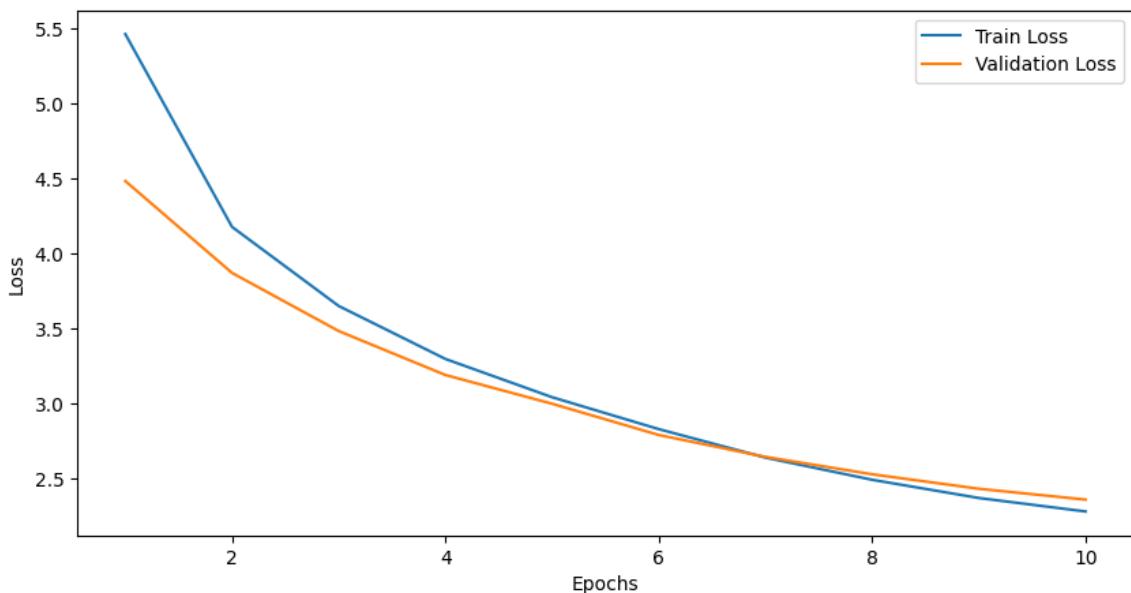
کلمات کمک می‌کند تا روابط میان کلمات را از طریق پیش‌بینی کلمات گمشده یاد بگیریم و از نشت اطلاعات جلوگیری کنیم. همچنین، masking کمک می‌کند تا مدل را بهتر تعمیم دهیم و همچنین مقاومت مدل را در برابر نویز بهبود ببخشیم.

برای سایر بخش‌ها، مطابق گزارش کار عمل می‌کنیم.

positionnal embedding به مدل کمک می‌کند تا بدون استفاده از ساختارهای تکراری یا همگام، اطلاعات موقعیتی را دریافت می‌کنیم و وابستگی‌های زمانی و ترتیبی را به درستی مدلسازی می‌کنیم.

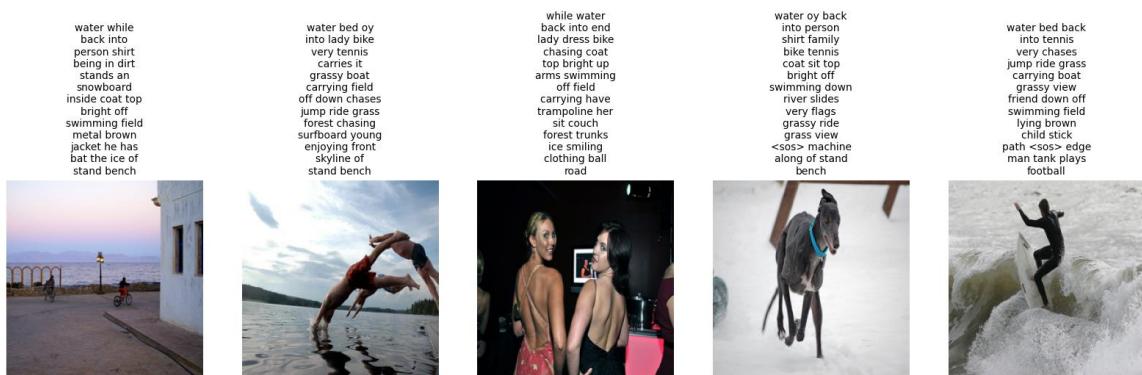
حال اقدام به آموزش مدل می‌پردازیم و نتایج را بررسی می‌کنیم.

شکل 2.9 نمودار تابع هزینه را برای مدل CNN-Transformer نشان می‌دهد.



شکل 2.9. تابع هزینه برای مدل CNN-Transformer

همچنین، شکل 2.10 5 تا از تصاویر به همراه کپشن تولیدی را نشان می‌دهد.

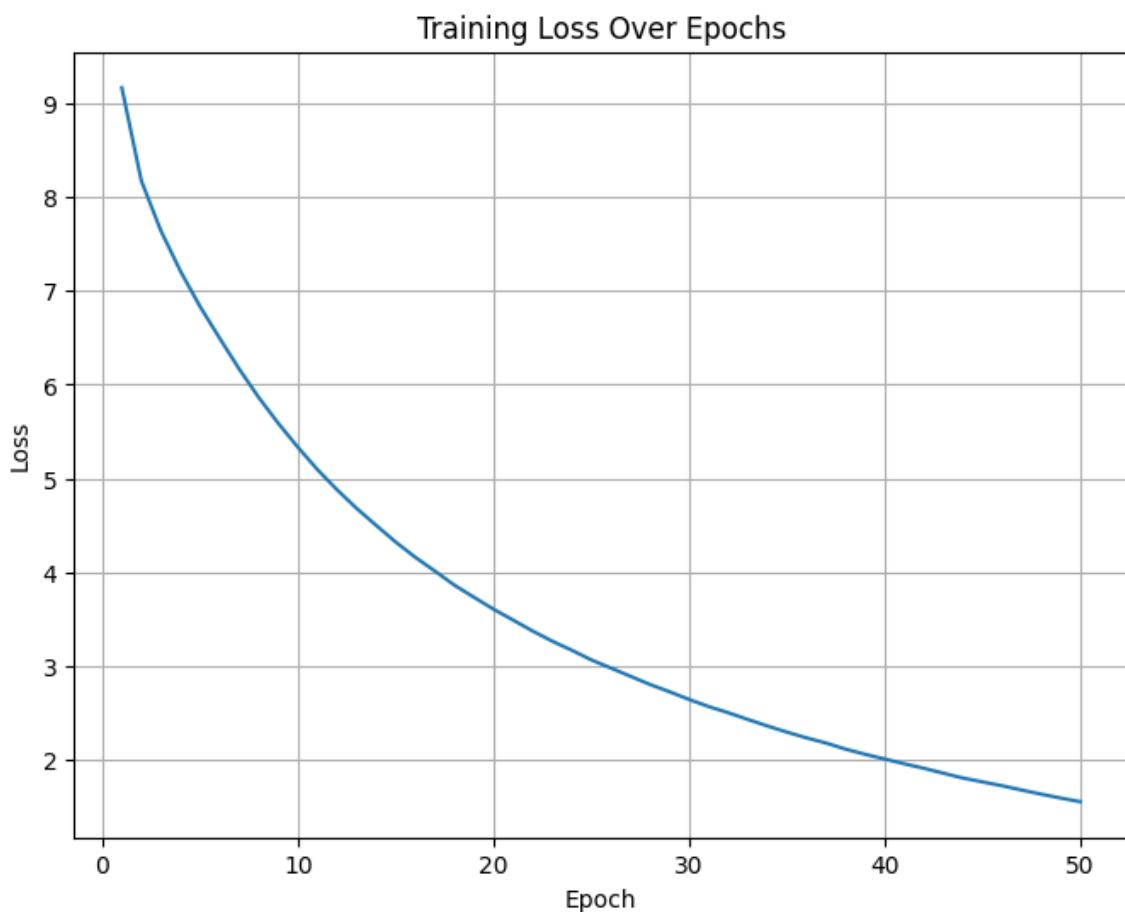


شکل 2.10. 5 تصویر به همراه کپشن تولیدی مدل CNN-Transformer

همانطور که مشاهده می‌شود، این مدل توصیفات بیشتری برای یک عکس به کار می‌برد، اما مشکل مدل ما این است که نتیجه خوبی در مورد کپشن‌ها نتیجه نمی‌دهد. برای این مشکل، با استفاده از آپلودشده تلاش می‌کنیم تا این مشکل را رفع کنیم. notebook

بدین منظور، یک notebook جدید و جدا می‌سازیم و در آن تمام مراحل را مجدداً انجام می‌دهیم. همچنین، تابع generate_caption را به نحوی تعریف می‌کنیم که از تولید کلمات تکراری جلوگیری شود. نتایج این مدل در پایین آمده است.

شکل 2.11 نمودار تابع هزینه را برای مدل جدید CNN-Transformer نشان می‌دهد.



شکل 2.11. تابع هزینه برای مدل جدید CNN-Transformer

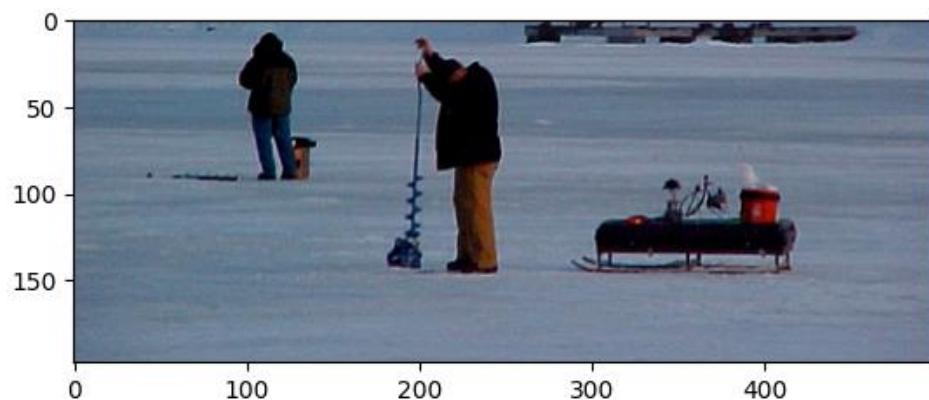
همچنین، شکل 2.12 تا 2.16 5 تا از تصاویر به همراه کپشن تولیدی را نشان می‌دهد.

کپشن تولیدی: rainy belongs grayhound bagpipes case pristine
glowers pigeon stomach ladie bowler british pretending netting dupont rusty avoid stepstool



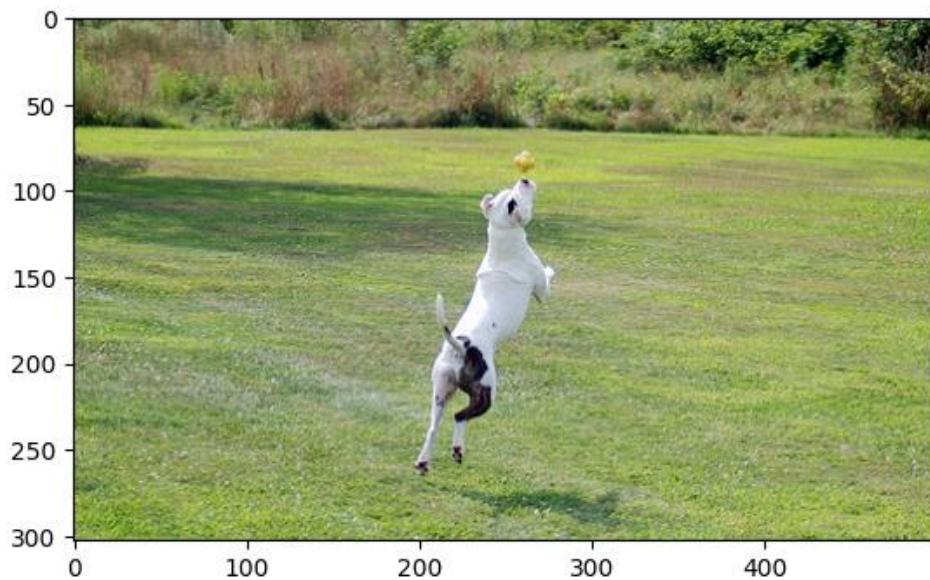
شکل 2.12. تصویر به همراه کپشن تولیدی مدل جدید CNN-Transformer

کپشن تولیدی: excited developing athelete pokes elf bicyclecross sharp terminal doorbell
twome skyscraper vinyl fortress rest setter fribee engulfed children



شکل 2.13. تصویر به همراه کپشن تولیدی مدل جدید CNN-Transformer

bangs representing crouches teenage spotted nike chasing alert stand end rugs
fixer misspelled headphone wrestles guessing



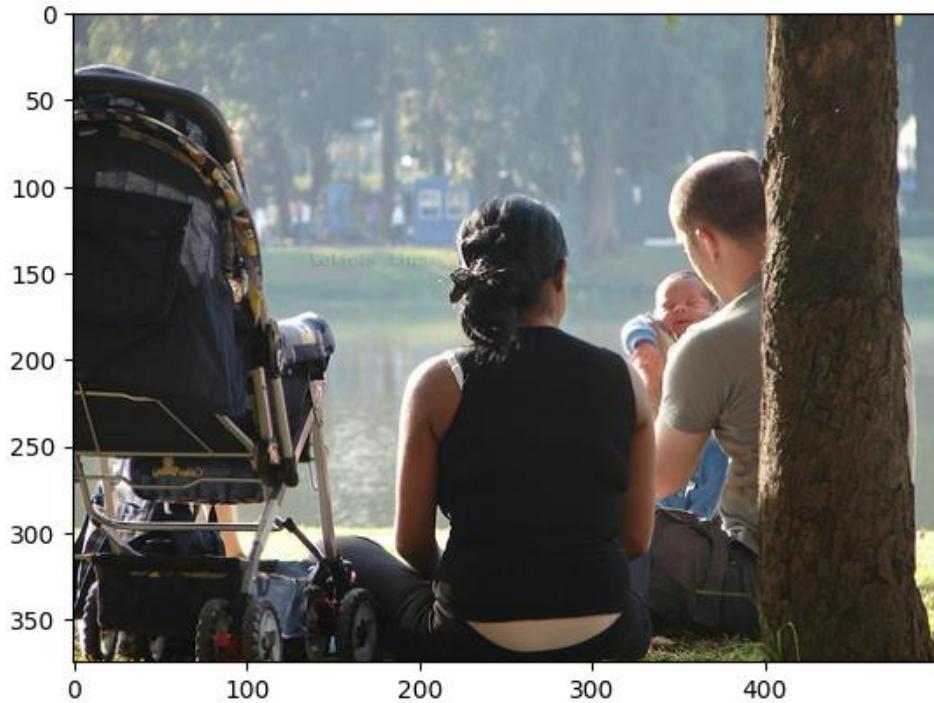
شکل 2.14. تصویر به همراه کپشن تولیدی مدل جدید CNN-Transformer

splashing kitten there jousting armstand pair herd subject cocacola highfive
creative places cam eagle give gleefully coasts bullbranded popping



شکل 2.15. تصویر به همراه کپشن تولیدی مدل جدید CNN-Transformer

کپشن تولیدی: winsuit soars gazing toetouch anklehigh badge juggles sacks stoppie sacks
hippie eye enjoyable backpacking choke play greens lame



شکل 2.16. تصویر به همراه کپشن تولیدی مدل جدید CNN-Transformer

مجدها مشاهده می‌شود علیرغم تغییرات انجام شده، باز هم کپشن‌های دلخواه نشده‌است.

6-2. بخش امتیازی

طبق مقاله‌هایی که داخل صورت گزارش وجود داشت، مهم‌ترین متريک برای ارزیابی اين ماشين‌های تولید کپشن، BLEU می‌باشد. اين کلمه مخفف bilingual evaluation understudy است که در تسک‌های زبان طبیعی استفاده می‌گردد. این متريک، کیفیت تولید متن را بررسی می‌کند. اين معیار ساده و کارآمد می‌باشد و در ارزیابی سистем‌های ترجمه ماشینی و تولید کپشن استفاده می‌گردد. اگرچه اين مدل محدودیت‌هایی دارد، اما همچنان يك معیار پراستفاده در حوزه پردازش زبان طبیعی است. در اين حوزه، معیارهای دیگری همچون CIDEr و METEOR نیز استفاده می‌گردد.