

گزارش تمرین دوم درس فهم زبان

موضوع تمرین:

پیاده‌سازی مدل‌های Bert و T5 در تشخیص قصد و پر کردن شکاف

نام استاد درس: دکتر حسین زینلی

نام دانشجو: مجید ادیبیان

شماره دانشجویی: ۴۰۰۱۳۱۰۷۸

پاییز ۱۴۰۱

۱. مقدمه

پر کردن شکاف و تشخیص قصد دو قسمت از سیستم‌های گفت‌وگو محور هستند. در تشخیص قصد، همان‌طور که از اسمش مشخص است، هدف تشخیص قصد گوینده متن مورد نظر است. در پرکردن شکاف متناسب با قصد گوینده باید slot هایی را که نیاز داریم را پرکنیم. به عنوان مثال اگر قصد گوینده تنظیم یک یادآور است باید slot مربوط به زمان آن یادآور را از جمله‌ای که گفته شده تعیین کنیم.

در این تمرین برای انجام این دو تسک از دو مدل Bert و T5 استفاده شده است. مدل Bert با گرفتن متن ورودی برای هر یک از کلمات متن یک بردار تعبیه می‌سازد و سپس باید از بردارهای ساخته شده برای هر کلمه خروجی دو تسک مورد نظر را به دست آورد. در مدل T5 باید تسک مورد نظر را به صورت یک تسک متن به متن تبدیل کرد. برای این کار متن ورودی را به عنوان ورودی مدل و دنباله slot ها در کنار intent جمله را به عنوان متن خروجی مدل در نظر می‌گیریم.

در ادامه روش پیاده‌سازی هر یک از دو مدل و نتایج به دست آمده از آن‌ها را بررسی خواهیم کرد.

۲. روش انجام کار

فرایند انجام کار شامل پیاده‌سازی مدل‌ها، پیش‌پردازش داده‌ها برای ساخت داده‌هایی مناسب برای آموزش مدل‌ها و در نهایت ارزیابی نتایج است که در ادامه با جزئیات به هر کدام می‌پردازیم.

۱.۲. آماده‌سازی داده‌ها

برای آموزش مدل‌ها باید داده‌ها را به صورتی که برای مدل قابل آموزش باشد تبدیل کنیم.

جهت آموزش مدل Bert با خواندن داده‌ها متن اصلی و slot ها و intent را برای هر ورودی به دست آورده‌ایم و سپس متن اصلی را به استفاده از Bert tokenizer به توکن‌های تشکیل دهنده شکسته و دنباله اعداد متناسب با آن‌ها را جایگزین کرده‌ایم. همچنین باید به ازای هر توکن ورودی برچسب slot متناظر آن را هم مشخص کنیم که به دلیل آن که برخی کلمات در قسمت قبل به چند توکن شکسته شده‌اند باید برچسب slot متناظر با آن‌ها را هم به همان تعداد تکرار کرد که در این کار برچسب‌های تکرار شده بعدی اگر با B- شروع شده‌اند آن‌ها را با I- جایگزین می‌کنیم. سپس هر یک از برچسب‌های slot ها و intent ها را یک id متناظر می‌کنیم تا از این اعداد در آموزش مدل استفاده شود.

در خصوص مدل T5 باید داده‌ها ورودی و خروجی مدل به صورت متنی باشند. برای آماده‌سازی داده‌های ورودی برای این مدل در ابتدای متن‌ها، تسک مورد نظر را اضافه می‌کنیم. به عنوان مثال برای اجرای هر دو تسک جمله «intent and slot» را به ابتدای همه جملات ورودی اضافه می‌کنیم. برای آماده‌سازی داده‌های خروجی باید برچسب‌ها را به صورت متنی بنویسیم که برای حالتی که هر دو تسک را بخواهیم همزمان انجام دهیم ابتدا intent قرار می‌گیرد و پس از علامت «;» دنباله برچسب slot ها قرار می‌گیرد و در صورتی که تنها یک تسک بخواهد انجام شود فقط برچسب همان تسک را به صورتی که گفتیم قرار می‌دهیم.

با بررسی‌هایی که انجام شده مشخص شد که تعداد زیادی از داده‌ها در هر یک از فایل‌ها آموزشی و ارزیابی و آزمون تکراری هستند که هنگام خواندن داده‌ها موارد تکراری در هر فایل را نادیده می‌گیریم که در انتها به ۲۳۰۱۶ جمله آموزشی و ۳۷۸۹ جمله ارزیابی و ۷۴۲۷ جمله آزمون می‌رسیم.

۲.۲. پیاده‌سازی مدل

برای این تمرین دو مدل Bert و T5 پیاده‌سازی شده است جزئیات پیاده‌سازی هر یک در ادامه گفته شده است.

۱.۲.۲. مدل Bert

این مدل یک کدگذار است که از لایه‌های ترنسفورمری استفاده می‌کند. در این تمرین این مدل به صورتی پیاده‌سازی شده که بتواند هر دو تسک مورد نظر را همزمان انجام دهد. در واقعاً چون خروجی این دو تسک می‌تواند بر روی یک‌دیگر اثر بگذارند و برای رسیدن به پیش‌بینی بهتر به هم کمک کنند، انجام همزمان این دو مدل می‌تواند باعث نتایج بهتری از آموزش دو مدل مجزا برای هر یک از تسک‌ها شود.

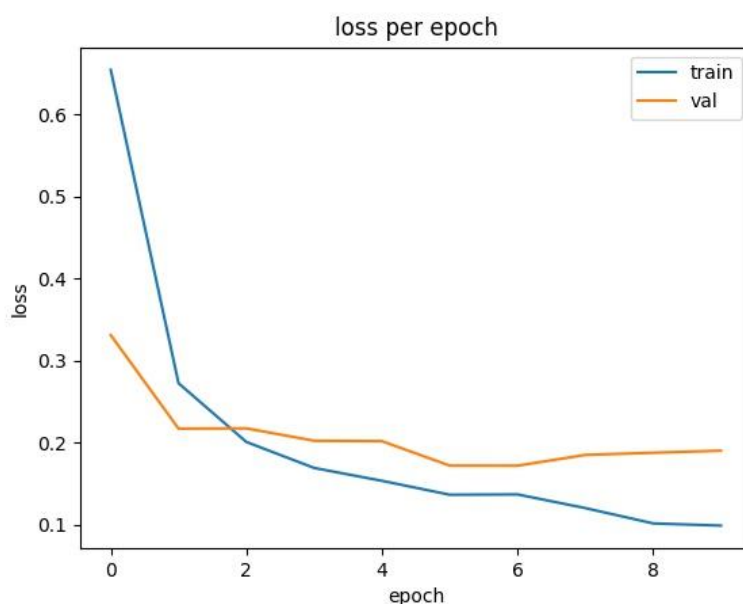
برای ساخت مدلی برای پیش‌بینی همزمان برای دو تسک مورد نظر کلاسی نوشته شده که بر روی خروجی هر یک از کلمات حاصل از Bert ابتدا یک drop out و سپس یک لایه خطی قرار می‌دهد تا بتواند برچسب مربوط به هر slot را پیش‌بینی کند. همچنین بر روی خروجی توکن cls هم همین دو لایه قرار گرفته شده تا بتواند intent متن را پیش‌بینی کند. برای محاسبه loss مربوط به آموزش مدل در هر دو تسک از cross entropy loss استفاده شده و نتیجه آن‌ها با هم جمع زده شده تا loss کلی مدل به دست آید. همچنین از Adam به عنوان تابع بهینه‌ساز استفاده شده است و با سعی و خطا مقدار نرخ یادگیری مناسب برای آن برابر $5e-5$ در نظر گرفته شده است.

۲.۲.۲. مدل T5

برای پیاده‌سازی این مدل از کتابخانه SimpleT5 استفاده شده است. ابتدا داده‌ها به صورت Source و Target در می‌آیند و سپس مدل بر روی این داده‌ها آموزش می‌بیند. در این قسمت ابتدا مدل از پیش آموزش دیده‌ی T5-base بارگذاری شده و سپس این مدل بر روی داده‌های جدید ساخته شده تا ۵ اپیاک آموزش می‌بیند.

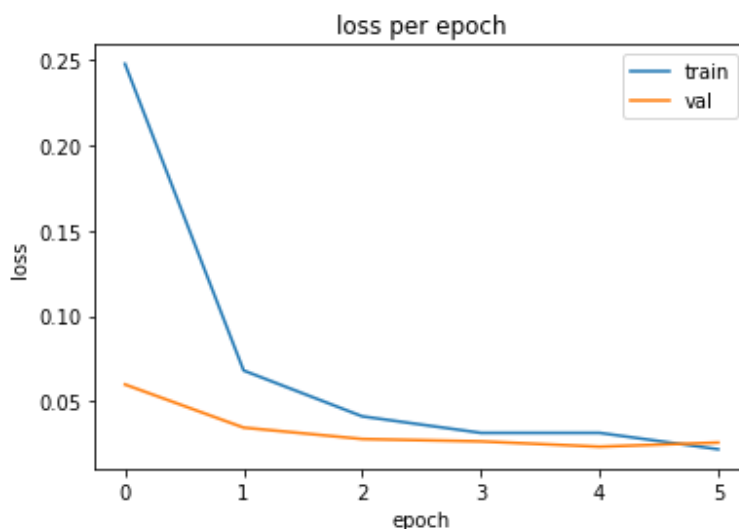
۳.۲. آموزش مدل

آموزش مدل Bert تا ۱۰ اپیاک بر روی داده‌های ساخته شده انجام شده است که تغییرات loss در آن برای داده‌های آموزشی و ارزیابی به صورت زیر است. زمان لازم برای آموزش مدل در هر اپیاک برابر ۲ دقیقه و ۱۵ ثانیه می‌باشد.



همان‌طور که از نمودار تغییرات loss مشخص است مقدار loss برای هر دو داده آموزشی و ارزیابی تا ۵ اپیاک کاهش یافته و پس از آن مقدار loss برای داده آموزشی همچنان کاهش بوده ولی برای داده ارزیابی مقداری افزایش یافته است.

آموزش مدل T5 تا ۶ اپیاک انجام شده است که تغییرات مقدار loss در داده‌های آموزشی و ارزیابی در آن به صورت زیر است:



۴.۲. نحوه اجرا

برای اجرای آموزش و ارزیابی مدل‌ها فایل نوت‌بوکی در کنار گزارش قرار گرفته شده است که دستورات اجرای کدها و برخی خروجی‌ها در آن قرار دارد.

۳. نتایج:

برای ارزیابی مدل‌های آموزش دیده، خروجی داده‌های آزمون را از مدل به دست می‌آوریم و این نتایج را مقادیر صحیح مقایسه می‌کنیم.

جهت ارزیابی نتایج در خروجی‌های slot filling از کد قرار گرفته در پروژه استفاده شده است که معیارهای precision و recall و F1 را به دست می‌آورد. برای ارزیابی خروجی‌های intent detection از معیار accuracy استفاده می‌کنیم. همچنین خود نتایج به دست آمده را در فایل در پوشه results ذخیره می‌کنیم که قابل بررسی و مقایسه انسانی باشد.

در مدل T5 جهت بررسی آن که انجام دو تسک به صورت همزمان یا مجزا چه تفاوتی دارد، مدل را به دو صورت joint و مجزا آموزش داده‌ایم که در ادامه نتایج را بررسی می‌کنیم.

در جدول زیر نتایج حاصل از مدل Bert و دو حالت گفته شده از مدل T5 قرار داده شده است:

T5 for Intent Detection	T5 for Slot Filling	Joint T5	Joint Bert	
-	Precision: 72	Precision: 70	Precision: 91.06	Slot Filling
	Recall: 76.5	Recall: 75.6	Recall: 91.5	
	F1: 74.15	F1: 72.4	F1: 91.28	
Accuracy: 96.5	-	Accuracy: 95.5	Accuracy: 98.8	Intent Detection

همچنین جهت بررسی ویژگی‌های دو مدل جدول زیر فراهم شده است که بتوان از جهت سرعت و حجم مدل‌ها این دو مدل را مقایسه کرد:

سرعت آموزش در هر اپیک (batch size=32)	تعداد پارامترها	
135 s	110 M	Bert
557 s	222 M	T5

نتیجه‌گیری:

با توجه به نتایج به دست آمده می‌توان نتیجه گرفت که مدل Bert نسبت به مدل T5 توانسته در هر دو تسک slot filling و intent detection و به خصوص در تسک slot filling به نتایج بهتری برسد. دلیل این موضوع می‌تواند آن باشد که مدل Bert با ساخت بردار embedding برای هر کلمه و سپس یک لایه‌ی دسته‌بندی بر روی آن، برای این تسک مدل بهتری است. البته مدل T5 را می‌توان به روش‌های مختلفی تست کرد و نتایج بهتری گرفت که از جمله این موارد آن است که فرمت‌های مختلفی برای متن ورودی و خروجی بررسی کنیم. یکی از این فرمت‌هایی که می‌توان بررسی کرد آن است که در slot filling به جای آن که دنباله برچسب‌ها را قرار دهیم، در متن خروجی ابتدا خود کلمه را بیاوریم و سپس برچسب آن را مقابل آن قرار دهیم.

همچنین مدل Bert نسبت به مدل T5 به لحاظ تعداد پارامترها و سرعت آموزش بسیار بهتر است.

در بررسی تاثیر آموزش مدل T5 به صورت مجزا برای هر تسک و به صورت ترکیبی از نتایج به دست آمده مشخص شد که آموزش مدل‌ها به صورت مجزا توانسته به دقت بیشتری برسد. دلیلی که برای این مورد می‌توان بیان کرد آن است که برای آموزش مدل به صورت ترکیبی، فرمت متن خروجی پیچیده‌تر می‌شود و مدل باید بیاموزد که متنی را تولید کند که شامل دو قسمت مجزا است و هر قسمت معنای متفاوتی را نشان می‌دهد. در حالی که در آموزش مدل برای هر تسک به صورت مجزا این پیچیدگی وجود ندارد.