

**چکیده:** سیستم‌های پرسش‌و‌پاسخ روشی برای کسب اطلاعات از ساختارهای مختلفی همچون ساختار‌های منظم و ساختار‌های غیرمنظم به انسان ارائه می‌کنند و بخش قابل توجهی از هوش مصنوعی مکالمه‌ای را تشکیل می‌دهند که همین مورد باعث شده شاخه‌ی تحقیقاتی جدیدی با عنوان پرسش‌و‌پاسخ مکالمه‌ای در علم پردازش زبان طبیعی ایجاد شود که در آن سیستم‌ها باید قادر به فهم متون بوده و در مکالمات چند مرحله‌ای با کاربر تعامل کرده و به درخواست‌های وی پاسخ دهد. در این پژوهش ابتدا به بررسی پیش‌زمینه‌های مورد نیاز برای درک این چالش پرداخته و سپس تعدادی از مدل‌های روز ارائه شده در این حوزه را از نظر معماری بکار رفته، نوآوری‌ها و دقت بدست آمده بررسی می‌کنیم.

**کلمات کلیدی:** فهم زبان طبیعی ، پرسش‌‌و‌پاسخ مکالمه‌ای ، مدل برت ، شباهت معنایی ، عبارات اسمی هم‌مرجع

**1 مقدمه**

طراحی یک سیستم دیالوگ هوشمند که بتواند همانند یک انسان مکالمه­ها را درک کرده و بتواند به پاسخ دادن سؤالات در زمینهها و موضوعات مختلفی همچون بیوگرافی افراد مشهور و یا اخبار جدید بپردازد، همواره یکی از مهمترین اهداف حوزه‌ی هوش مصنوعی بوده است که باعث شده در سالهای اخیر توجه محققین بیشتری از حوزه‌ی دانشگاه و صنعت را به خود جذب کند. این حوزه از هوش مصنوعی که با نام هوش مصنوعی مکالمه‌ای شناخته میشود، خود به سه زیرشاخه تقسیم میشود. شاخه اول سیستم‌های مکالمه‌ای گفتگو‌محور‌ (chat-oriented) میباشد که هدف اینگونه سیستم‌ها ایجاد یک مکالمه‌ی طبیعی و دوستانه با کاربر میباشد. سیستم‌هایی از قبیل آلکسا (Alexa)، سیری (Siri) و یا کورتانا (Cortana) را میتوان جزو این شاخه حساب کرد. شاخه‌ی دوم سیستم‌های مکالمه‌ای وظیفه‌محور (task oriented) میباشند که وظیفه‌ی اجرای اعمال بخصوصی مانند رزرو یک رستوران و یا برنامه‌ریزی یک رخداد را پس از فهم مکالمه دارند. سومین شاخه سیستم‌های سؤال و جواب مکالمه‌محور (QA dialog systems) میباشند که هدف ارائه‌ی پاسخی صحیح و روان در پی سؤالات کاربر را دنبال میکنند.[1] در این پژوهش به روش‌ها و چالش‌های ایجادِ سیستم‌های سؤال‌ و جواب مکالمه‌محور میپردازیم. ابتدا تعدادی از مفاهیم مورد نیاز را شرح میدهیم و سپس به بررسی مدل‌های روز و پژوهش‌های انجام شده در این حوزه میپردازیم. در نهایت آزمایشات انجام شده و نتایج بدست آمده را بررسی خواهیم کرد و یک زمانبندی برای ادامه‌ی پژوهش ارائه می‌دهیم.

**1.1 مفاهیم اولیه**

در این قسمت به بررسی مسائل مرتبط با مسأله‌ی پرسش و پاسخ مکالمه‌ای میپردازیم و روش‌های ارزیابی دقت مدل‌های ایجاد شده برای هر یک از این چالش‌ها را از نظر میگذرانیم.

۱.۱.۴ مسأله‌ی پرسش و پاسخ مبتنی بر متن (Question Answering in Context)

**تعریف:** در تاریخچه علم پردازش زبان طبیعی، آموزشِ خواندن و فهم معنای متون به کامپیوترها همواره یکی از مهمترین اهداف بوده است. برای نیل به این هدف محققین سعی در ارائه روش‌هایی برای اندازه‌گیری مقدار فهم کامپیوتر از متون کرده‌اند. در این روش‌ها با ارائه یک قطعه متن به ماشین و پرسیدن سؤالات مرتبط با همان متن سعی می‌شود میزان فهم کامپیوتر از متن ارزیابی شود. این روش‌ها به دسته‌های مختلفی از قبیل چند‌گزینه‌ای (multiple choice)، پاسخ با فرم آزاد (free-form answer)، پرکردن جای خالی (cloze style) و پیش‌بینی محدوده (span prediction) تقسیم میشوند که تمرکز ما در این پژوهش روی مورد آخر میباشد. در چالش پرسش‌و‌پاسخ مبتنی بر متن با پیش‌بینی محدوده، ماشین بایستی پاسخ سؤال پرسیده شده را با مشخص کردن یک محدوده – که دارای یک نقطه شروع و یک نقطه پایان میباشد - داخل همان متن ورودی، ارائه نماید. به صورت ریاضی اگر به ماشین یک متن به شکل C = {c­0, …, ck, …, ct} داده شود که در آن ci ها کلمات واژه­نامه میباشند، پاسخ بایستی به شکل A = {cm, …, cn} تولید شود که 0 ≤ m ≤ n ≤ t میباشد. [2]

**مجموعه‌های دادگان.** در پی تعریف این چالش مجموعه دادگانهای مختلفی نیز ارائه شد که از مهم‌ترین آنها میتوان به مجموعه دادگان‌های SQuAD که در دو نسخه SQuAD1.0 و SQuAD2.0 ارائه شده، اشاره کرد.[3] نکته قابل توجه در نسخه دوم این دادگان وجود سؤالاتی است که پاسخ آنها در متن داده‌شده وجود ندارد و این باعث میشود تا سیستم‌های طراحی‌شده مجبور باشند در صورتیکه پاسخ در متن ورودی موجود نیست از ارائه حدس‌های بی‌ربط خود‌داری کنند.[4]

**ارزیابی.** لازم به ذکر است که برای ارزیابی مدل‌ها وابسته به مجموعه‌ی ‌دادگان مورد نظر روش‌های مختلفی وجود دارد. به عنوان مثال روش‌های ارزیابی مدل ارائه شده برای دادگان SQuAD شامل موارد زیر می‌باشد:

* *تطبیق کامل (exact match):* در این حالت پاسخ ارائه‌شده بایستی با یکی از پاسخ‌های موجود برای سؤال مطابقت کامل داشته باشد و درصد این تطابق‌های کامل گزارش میشود.
* *معیار F1 در سطح توکن(token-level F1):* بر اساس این معیار هر پاسخ ارائه شده مدل در سطح کلمات یا توکن‌ها با پاسخ‌های طلایی مقایسه میشود و دو معیار صحت (precision) و یاداوری (recall) محاسبه شده و مقادیر F1 برای هر پاسخ بدست می‌آید.

۱.۱.۵ مسأله‌ی پرسش و پاسخ مکالمه‌ای مبتنی بر متن(conversational question answering in context)

**تعریف.**  همانطور که امر پرسش و پاسخ از طریق مکالمه یکی از روش‌های اصلی جمع‌آوری اطلاعات برای انسان‌هاست، چالش پرسش و پاسخ مکالمه‌ای به این منظور ارائه شده تا معیاری باشد برای فهم ماشین از مکالمات. این مسأله تا حد خوبی شبیه به مسأله‌ی پرسش و پاسخ است و تنها موردی که باعث تفاوت این دو مورد از یکدیگر میشود وجود تاریخچه‌ی مکالمه یا همان مراحل قبلی سؤال و جواب در مکالمه‌ی فعلی است. به بیان ریاضی اگر به ماشین یک متن به شکل C = {c­0, … ck …, cd} به همراه یک مجموعه تاریخچه به صورت H = {(Q0, A0), (Q1, A1), …, (Qs-1, As-1)} داده ‌شود که در آن ci ها کلمات واژه­نامه میباشند و مجموعه H شامل تاریخچه مکالمه به صورت زوج‌های سؤال و جواب است، پاسخ بایستی به شکل A = {cm, …, cn} تولید شود که 0 ≤ m ≤ n ≤ d. [5]

**مجموعه‌های دادگان.** از مهمترین مجموعه دادگان‌های ارائه شده برای این چالش می‌توان به مجموعه دادگان CoQA و QuAC اشاره کرد. ساختار هر دو این مجموعه دادگان تقریبا مشابه یکدیگر است و در هر دو تعدادی مکالمه وجود دارد که تعداد مراحل هر مکالمه به طور میانگین شامل ۱۰ مرحله میباشد و در هر مکالمه سؤالاتی مرتبط با متن مرجع مربوط به آن مکالمه پرسیده میشود. سؤالات دارای ترتیب خاصی می‌باشند و ممکن است به سؤالات قبل از خودشان در همان مکالمه ارجاع داده باشند که این هم یکی از چالش‌های مهم این مجموعه دادگان هاست. همچنین می‌توان گفت QuAC از CoQA چالش برانگیزتر است که کارایی مدل‌های روز بر روی هر یک از این مجموعه دادگان‌ها نیز مؤید این امر می‌باشد.[6, 7]

**ارزیابی.** ارزیابی مدل‌های ارائه شده بر‌روی این دو دادگان شامل همان روش‌های مذکور در ۱.۱.۴ می‌شود منتها دادگان QuAC دو معیار دیگر نیز ارائه کرده است که ارزیابی کارایی مدل‌ها در مقایسه با کارایی انسان را میسر میکند. بر اساس این معیار که با HEQ (کوتاه‌شده‌ی عبارت نمره‌ی معادل انسانی)(human equivalence score) شناخته می‌شود، درصدی از پاسخ‌های مدل که دقت F1 آنها با دقت F1 انسانی برابری میکند و یا از آن پیشی میگیرد گزارش می‌شود.[7]

1.1.1 مسأله‌ی شباهت معنایی

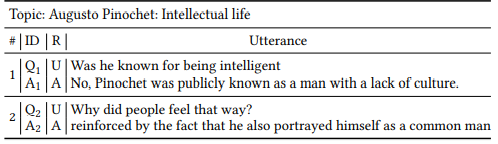
**تعریف.** در این مسأله که یکی از قدیمی‌ترین مسائل در حوزه پردازش زبان طبیعی است، هدف مقایسه کمّی شباهت دو قطعه‌ی متنی میباشد. به عبارت دیگر، روش‌های ارائه‌شده معمولا بایستی قادر باشند تا درصدی از شباهت بین دو قطعه ارائه دهند و صرفا ارائه پاسخ به صورت بله یا خیر کافی نیست. همچنین لازم به ذکر است که این مسأله معمولا به عنوان مترادف با مسأله‌ی ارتباط معنایی بکار می‌رود.[8]

**مجموعههای دادگان.** شناخته‌شده‌ترین دادگان‌های موجود در این حوزه مجموعه دادگان‌های شباهت متنی بین جملات (STS)، مجموعه دادگان عبارات معادل مایکروسافت (MSRP)، مجموعه سؤالات تکراری وب‌سایت کوئرا (Quora duplicate questions) و مجموعه جملات مرتبط از نظر دانش عمومی(SICK) می‌باشند.[8]

**ارزیابی.** در برخی از دادگان‌ها همانند نسخه‌های مختلف STS، زوج‌های مختلف از جملات توسط نیروی انسانی به عنوان شباهت دو نمونه نمره‌گذاری عددی شده اند و بر اساس این برچسب‌ها کارایی مدل‌ها توسط معیار همبستگی پیرسون (Pearson correlation) سنجیده می‌شود. در عین حال برخی از دادگان‌ها همانند دادگان کوئرا و دادگان مایکروسافت، این مسأله را به یک مسأله‌ی دسته‌بندی دودویی (binary classification) تبدیل کرده‌اند که در آن مدل باید شباهت و یا عدم شباهت دو قطعه‌ی متنی را اعلام کند و بنابراین همه‌ی معیارهای ارزیابی برای دسته‌بندها، همانند معیار F1، در اینجا نیز قابل استفاده می‌باشند.[9]

۱.۱.۳ مسأله‌ی حل عبارات اسمی هم‌مرجع (coreference resolution)

**تعریف.** این مسأله که یکی از مسائل پایه‌ای علم پردازش زبان است، هدفِ یافتن تمام ارجاعات به موجودیت‌های داخل یک پیکره‌ی متنی را دنبال می‌کند و می‌تواند تاثیر چشمگیری در بهبود وظایف سطح بالاتر پردازش زبانی از قبیل خلاصه‌سازی، آنالیز احساسات و ... داشته باشد به طوریکه معمولا حل این مسأله توسط ماشین را معادل پاس کردن تست تورینگ (Turing test) دانسته‌اند، چرا که حل هم‌مرجعی نیازمند دانش عمومی می‌باشد. حل این مسأله معمولا در دوفاز دسته‌بندی و سپس خوشه سازی انجام می‌شود. در ابتدا موجودیت‌های اسمی شناسایی می‌شوند و سپس در فاز خوشه بندی به دسته‌های هم‌مرجع قرار داده می‌شوند. به علت‌ چالش‌های زیاد، این مسأله خود به مسائل کوچکتری مثل حل عبارات ضمیری هم‌مرجع، آنافورا (anaphora) و کاتافورا (cataphora) تقسیم میشود.[11]



شکل 1- نمونه ای ‌از عبارات اسمی هم مرجع در دادگان QuAC

**مجموعههای دادگان.** از مهم‌ترین مجموعه‌دادگان‌های این چالش می‌توان به دادگان CoNLL، دادگان MUC و دادگان WikiCoref اشاره نمود. همچنین از دادگان چالش وینوگراد (WSC) که یکی از محبوب‌ترین دادگان‌ها برای مسأله حل عبارات ضمیری هم‌مرجع است میتوان نام برد که گرفتن امتیاز بالا در این چالش نیازمند بهره‌مندی مدل از دانش عام می‌باشد. [12]

**ارزیابی.** هم‌مرجعی معمولا توسط معیار F1 کانل (CoNLL) سنجیده می‌شود که ترکیبی از سه معیار MUC ، B­3 و CEAF­­­e می‌باشد.[11]

۱.۱.۶ مدل برت(BERT)

یکی از مهمترین و شناخته‌ترین مدل‌هایی که امروزه در امر پردازش زبان طبیعی بکار میرود مدل برت می‌باشد که در سال ۲۰۱8 توسط گروه تحقیقات هوش مصنوعی شرکت گوگل ارائه شد. این مدل که نام آن مختصر‌شده‌ی نمایش‌های کدگذار دو طرفه از ترانسفورمرهاست، همانند ترانسفورمر از مکانیزم توجه استفاده میکند و به صورت پیش‌فرض روی مجموعه‌های حجیمی از متون مختلف از قبیل وب‌سایت‌های خبری و علمی برای یادگیری مدل زبانی آموزش داده شده و معمولا به عنوان یک مدل هسته‌ای برای قسمت کدگذار‌ (encoder) در مدل‌های پردازش زبانی به کار میرود و میتواند با توجه به متن ورودی برای هر کلمه یک بازنمایی مناسب تولید کند. برای استفاده از این مدل دیگر نیازی به ساعت‌ها آموزش روی دادگان حجیم نمی‌باشد و صرفا نیاز به دو مرحله‌ی پیش‌آموزش (pre-training) و تعدیل‌سازی (fine-tuning) می‌باشد. در مرحله‌ی اول مدل زبانی برت را با ماسک کردن حدود ۱۵ درصد از کلمات دادگان آموزش و به صورت بدون نظارت (unsupervised) آموزش میدهند و در مرحله بعدی با اضافه کردن تعدادی لایه شبکه‌ عصبی، پیش و یا پس از مدل برت، می‌توان مدلی برای وظیفه‌ی مورد نظر ایجاد کرد و آن را آموزش داد. با توجه به ساختار استفاده شده و تعداد لایه‌ها و یا تعداد پارامتر‌های استفاده شده داخل مدل و حتی دادگان استفاده شده هنگام مرحله آموزش مدل زبانی، مدل برت به نام‌ها و دسته‌های مختلفی تقسیم می‌شود که برای نمونه می‌توان به مدل برت پایه (BERTBASE) و مدل برت بزرگ (BERTLARGE) اشاره نمود که در حالیکه اولین مورد دارای ۱۲ لایه رمزگذار می‌باشد مدل برت بزرگ دارای ۲۴ عدد لایه رمزگذار است.[10]

**۱.۲ تعریف مسأله**

همانطور که اشاره شد، سیستم‌های پرسش‌و‌پاسخ مکالمه‌ای برای داشتن عملکرد مناسب، بایستی علاوه بر توانایی در درک متن منبع و استخراج پاسخ مناسب از آن، بتوانند در هنگام پاسخ به سؤال فعلی، تاریخچه‌ی پرسش‌ و پاسخ‌های پیشین مکالمه را نیز در نظر بگیرند و از آن‌ها نیز در هنگام پاسخ به سؤال بهره ببرند که همین مورد اساسی‌ترین مسأله‌ای است که در این حوزه به آن پرداخته میشود. برای مواجهه با این مسأله تا به حال روشهای مختلفی در نظر گرفته شده؛ در برخی از پیاده‌سازی‌ها به کلی تاریخچه‌ی مکالمه در نظر گرفته نمی‌شود، یا اینکه سعی شده است که تعداد مشخصی از مراحل قبلی و یا تمام تاریخچه در حین پاسخ به یک سوال در نظر گرفته شود؛ اما در یکی از پژوهشهای اخیر، سامانه به گونه‌ای طراحی شده که در هر نوبت تنها قسمتی از تاریخچه که مرتبط‌تر با سوال فعلی بنظر میرسند در نظر گرفته شود. این کار از طریق یک ماژول شبکه عصبی انجام شده و در هر نوبت ابتدا سؤالاتی از تاریخچه که با سوال فعلی مرتبط هستند انتخاب شده، سپس با سؤال فعلی ترکیب شده و یک نمایش از آنها به یک سامانه برت داده میشود و در نهایت پاسخ به صورت قسمتی از متن مرجع داده می‌شود که دقت مناسبی نیز بدست آورده است.[5] هدف ما در این پژوهش بهبود دادن واحد انتخاب تاریخچه و همچنین بهبود دقت این پژوهش است. در حالی‌که در مقاله‌ی مذکور برای انتخاب تاریخچه یک معیار شباهت معنایی از سؤال فعلی با سؤالات قبلی در نظر گرفته شده که برپایه‌ی یک آستانه، سؤالات مرتبط را انتخاب میکند و به عنوان تاریخچه در نظر میگیرد، هدف ما در این پژوهش یادگیری آستانه مناسب بر اساس الگوریتم‌های نزول در امتداد گرادیان خواهد بود و همچنین برای انتخاب تاریخچه علاوه بر معیار شباهت معنایی از عبارات اسمی هم‌مرجع نیز کمک خواهیم گرفت، چرا که در امر پرسش‌و‌پاسخ مکالمه‌ای اسامی هم‌مرجع به وفور یافت می‌شوند و یکی از چالش‌های اساسی این مسأله می‌باشند. براساس تحلیل‌های انجام شده، ذکر شده که در دادگان CoQA نزدیک به ۴۹ درصد از سؤالات حاوی ارجاعات به تاریخچه مکالمه هستند و این مورد برای دادگان QuAC چیزی حدود ۴۴ درصد می‌باشد. [6, 7]

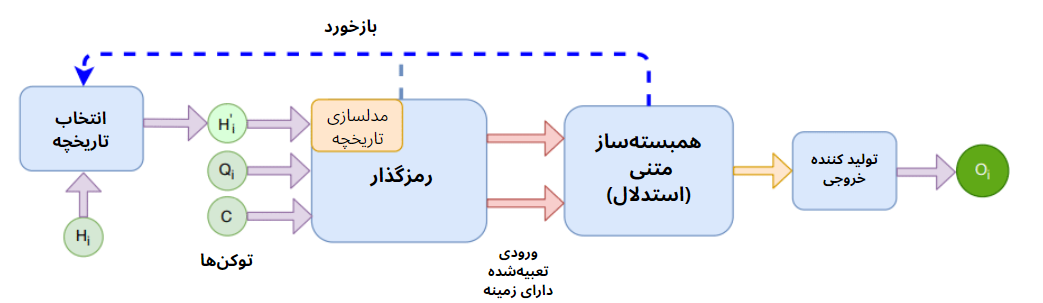
دادگان‌های مورد استفاده در این پژوهش مجموعه‌های QuAC و CoQA و معیارهای ارزیابی F1 و Q-HEQ و D-HEQ خواهند بود که همان معیارهای رسمی شناخته شده در حوزه پرسش‌و‌پاسخ مکالمه‌محور می‌باشند.

**۲ پژوهش‌های مرتبط**

افزایش توجه محققین به مسأله‌ی پرسش‌و‌پاسخ مکالمه‌ای در سال‌های اخیر، باعث نشر همزمان مقالات متعددی در این زمینه شده است که هر یک به شیوه‌ای متفاوت ولی از نظر ساختاری مشابه، سعی در ارائه روش خود داشته‌اند. در ادامه، ابتدا به بررسی معماری کلی این روش‌ها خواهیم پرداخت و سپس تعدادی از این روش‌ها را به طور مختصر بررسی خواهیم نمود.

**۲.۱ معماری کلی سیستم‌های پرسش و پاسخ مکالمه‌ای مبتنی بر متن**

با توجه به مشاهدات انجام شده و چالش‌های وظیفه‌ی پرسش‌و‌پاسخ مکالمه‌ای، معماری‌های مختلفی ارائه شده که اکثر آن‌ها دارای یک معماری کلی مشترک می‌باشند و در جزئیات با یکدیگر متفاوت اند. یک مدل پرسش‌و‌پاسخ مکالمه‌ای همانند شکل 2 دارای چهار بخش اساسی است که در ادامه به توضیح آنها می‌پردازیم.



شکل 2 - شمای کلی یک مدل پرسش‌و‌پاسخ مکالمه‌ای

۲.۱.۱ انتخاب تاریخچه

برای اینکه مدل ما بتواند پاسخ‌ها را به طور دقیق‌تری پیش‌بینی کند لازم است که تاریخچه مکالمه را به همراه پرسش فعلی و متن مرجع به مدل به عنوان ورودی بدهیم. در عین حال، علیرقم اینکه قسمت‌هایی از تاریخچه که به سؤال فعلی مرتبط می‌باشند می‌توانند مفید واقع شوند، قسمت‌های نامرتبط از تاریخچه می‌توانند همانند نویز عمل کنند و دقت مدل را کاهش دهند. بنابراین انتخاب دقیق قسمت‌های مناسب تاریخچه اهمیت اساسی در کارایی مدل خواهد داشت. بر این اساس استراتژی‌های انتخاب تاریخچه معمولا به دسته‌های زیر تقسیم می‌شود:

* *انتخاب K مرحله:*  در این نوع استراتژی تمام مراحل مکالمه وارد بخش رمزگذار می‌شود.
* *انتخاب تعدادی از آخرین مراحل*: در این استراتژی یک عدد ثابت در نظر گرفته می‌شود و آن تعداد از آخرین مراحل مکالمه وارد رمزگذار می‌شود. بطور مثال [14] پیشنهاد داده که این عدد را ۲ در نظر بگیریم، در حالیکه مدل BERT-HAE [15] پیشنهاد میدهد که در نظر گرفتن ۵ یا ۶ مرحله‌ی انتهایی مکالمه کافی است.
* *انتخاب تاریخچه به صورت پویا:*  با بررسی نمونه‌های دادگان‌های پرسش‌و‌پاسخ مکالمه‌ای مشاهده می‌شود که خواصی از یک مکالمه همانند تغییر موضوع و یا بازگشت به موضوع با استراتژی انتخاب آخرین مراحل در تعارض است، به همین علت یک استراتژی با نام مدلسازی پاسخ‌های تاریخچه (History answer modeling) به عنوان یک روش پویا معرفی شد که به مراحل قبلی تاریخچه وزن‌های مختلفی بر اساس ارتباطشان با سؤال فعلی می‌دهد و در نهایت تاریخچه‌ی وزندار با سؤال فعلی ترکیب شده و وارد مدل می‌شود.[16] همچنین در روشی دیگر که با نام Env-ConvQA شناخته می‌شود یک استراتژی برای انتخاب پویای چند مرحله از تاریخچه پیشنهاد شده که براساس استراتژی پاداش‌محورِ عقبگرد تقویتی (reward-based reinforced backtracking) کار میکند. این مدل پروسه‌ی انتخاب تاریخچه‌ی مرتبط را به عنوان یک پروسه‌ی تصمیم‌گیری پشت‌ سر هم در نظر میگیرد و همه‌ی تاریخچه را به صورت عقبگرد امتحان میکند تا بتواند تشخیص دهد مرحله با سؤال فعلی مرتبط است یا خیر.[17]

۲.۱.۲ رمزگذار

این قسمت وظیفه تبدیل توکن‌های متن منبع، سؤال فعلی، و قسمت انتخاب شده از تاریخچه به بردار‌های با طول ثابت را که به عنوان ورودی ماژول استدلال می‌باشند برعهده دارد. به منظور بهبود تاثیر تعبیه‌ی لغات معمولا خواص اضافی مانند برچسب ادات سخن (part of speech tag) و تعبیه پاسخ‌های تاریخچه (history answer embeddings) نیز به عنوان ورودی به این قسمت داده می‌شوند. روش‌های رمزگذاری معمولا به دو دسته تقسیم می‌شوند.

دسته‌ی اول روش‌های سنتی تعبیه لغات هستند که وظیفه رمزگذاری لغات ورودی به بردار‌های با ابعاد ثابت را بر عهده دارند که دو نمونه از معروف‌ترین آنها روش‌های GloVE و Word2Vec می‌باشند.

در حالیکه روش‌های سنتی بازنمایی نسبتا خوبی از لغات در یک فضای برداری ایجاد میکنند، هنگامی که لغات در زمینه‌های مختلفی قرار می‌گیرند تغییری در نمایش آن‌ها ایجاد نمی‌شود. به عبارت دیگر این گونه روش‌ها توانایی در بر گرفتن بازنمایی زمینه را دارا نیستند. به همین دلیل دسته دیگری از روش‌ها که به عنوان تعبیه‌ی زمینه‌ای (contextual embedding) معرفی شدند. این نوع از تعبیه‌سازی‌ها ابتدا توسط حجم عظیمی از متون پیش‌آموزش‌دهی می‌شوند و سپس به عنوان تعبیه‌ی لغات استفاده می‌شوند که به عنوان شناخته‌ترین آن‌ها می‌توان از برت نام برد.

۲.۱.۳ مدلسازی تاریخچه

مدلسازی تاریخچه پروسه‌ای است که در آن تاریخچه‌ی مکالمه با متن مرجع و سؤال فعلی ترکیب می‌شود تا یک ورودی کامل برای ماژول استدلال تشکیل شود. در حالیکه این پروسه معمولا بخشی از ماژول رمزگذار است، در اینجا به شرح جداگانه‌ی آن میپردازیم، چرا که مدل‌های مختلف از روش‌ها و تکنیک‌های متفاوتی برای مدلسازی تاریخچه استفاده کرده‌اند که هر یک ایده‌ی جداگانه‌ای بوده و حائز اهمیت می‌باشد. تعدادی از این روش‌ها به شرح زیر است:

* *چسباندن تاریخچه‌ی مکالمه به انتها:* در این روش مراحل انتخاب شده از تاریخچه‌ی مکالمه را به همراه سؤال فعلی به انتهای متن مرجع اضافه میکنیم و متن حاصل را به ماژول استدلال می‌دهیم. در برخی از روش‌های مشابه تنها سؤالات قبلی از تاریخچه به همراه رمزگذاری شماره‌ی هر سؤال به انتهای متن اضافه می‌شود.[17,18]
* *نشانه‌گذاری پاسخ‌ها داخل متن مرجع:* در این روش هر توکن از متن مرجع به گونه‌ای کدگذاری می‌شود که مشخص شود در یکی از پاسخ‌های پیشین وجود داشته یا خیر.[15] در روشی مشابه علاوه بر وجود یا عدم وجود هر توکن در پاسخ‌های پیشین، شماره‌ی مرحله‌ای از مکالمه که آن توکن در پاسخش حاضر است نیز کدگذاری می‌شود. این باعث می‌شود تا مدل درکی از تقدم و تاخر پاسخ‌های کدگذاری شده داشته باشد. [16]
* *ایجاد نمایش پنهان با استفاده از توکن‌های متن مرجع:* دسته‌ای از مدل‌ها که با نام مدل‌های جریان‌محور (flow-based) شناخته می‌شوند، سعی میکنند با استفاده از توکن‌های متن یک فضای پنهان و واسط ایجاد کرده و به این طریق جریان مکالمه را دنبال کنند.[18, 19]

۲.۱.۴ بخش استدلال

مدل‌های پرسش‌و‌پاسخ مکالمه‌ای براساس نحوه اجرای پروسه‌ی استدلال به دسته‌های مختلفی از قبیل روش‌های سنتی، مدل‌های زبانی پیش‌آموزش‌داده‌شده، و مدل‌های جریان‌محور تقسیم می‌شوند. علیرقم وجود این دسته‌ها، امروزه بیشتر از مدل‌های از پیش آموزش‌داده‌شده همانند انواع مدل‌های برت استفاده می‌شود چرا که یکی از مهم‌ترین مزایای استفاده از آن‌ها قابلیت ادغام ماژول‌های رمزگذار و استدلال با یکدیگر می‌باشد. این قابلیت باعث مخفی شدن طبیعت تعاملی سؤال فعلی با متن مرجع می‌شود. با این حال جا‌دادنِ تاریخچه‌ی قبلی به همراه متن مرجع یک مسأله‌ی چالشی در مدل‌های پیش‌آموزش‌داده‌شده‌ (علی‌الخصوص مدل برت) می‌باشد چرا که این مدل تنها اجازه‌ی ورود دو قطعه متن را می‌دهد و به علاوه داری محدودیت حداکثری ۵۱۲ روی تعداد توکن‌هاست. بنابراین اگر بخواهیم قسمت‌های بیشتری از تاریخچه را به متن اضافه کنیم نیاز داریم خود متن را کوتاه‌تر کنیم که این باعث نزول کارایی می‌شود. این نکته باعث اهمیت بیشتر مدلسازی مناسب تاریخچه و تاثیر مستقیم آن در بخش استدلال می‌شود.

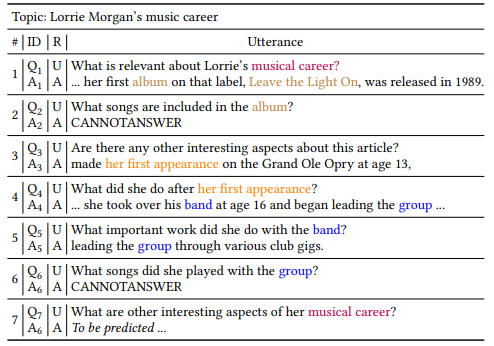
۲.۱.۵ بخش تولید خروجی

در نهایت برای پیش‌بینی پاسخ که قرار است یک بازه‌ی به هم پیوسته از متن مرجع‌ باشد، احتمال شروع پاسخ بودن و احتمال پایان پاسخ بودن به صورت جداگانه و برای هر توکن محاسبه می‌شود و توکن‌های با بیشترین احتمال برای آغاز و پایان بازه انتخاب شده و پاسخ بدست می‌آید. برای سؤالاتی که غیرقابل‌پاسخ (unanswerable) هستند، یک توکن به شکل UNANSWERABLE به انتهای متن اضافه می‌شود و مدل یاد میگیرد این توکن را در صورت غیرقابل پاسخ بودن سؤال پیش‌بینی نماید.

**۲.۲ مدل‌های مرجع پرسش و پاسخ مکالمه‌ای مبتنی بر متن**

برخی از مهم‌ترین چالش‌ها در امر پرسش‌و‌پاسخ مکالمه‌ای عبارتند از:

* مدل‌ها باید قادر باشند علاوه بر متن مرجع که همراه سؤالات داده می‌شود، تاریخچه را نیز تعبیه‌سازی (encode) نمایند.
* مشاهدات کلی از نحوه جمع‌آوری اطلاعات در مکالمات انسان بیانگر آن است که در شروع مکالمه تمرکز برروی قسمت‌های ابتدایی متن می‌باشد و این تمرکز در حین پیشرفت مکالمه به صورت تدریجی به قسمت‌های بعدی متن پیشرفت می‌کند. بنابراین نیاز است که مدل‌های ارائه شده به نحوی این جابه‌جایی تمرکز را در نظر بگیرند.
* در مکالمات معمولا به صورت صعودی استفاده از اسامی هم‌مرجع افزایش میابند و هر مرحله از مکالمه می‌تواند دنباله‌ای از مرحله‌‌ی قبلی، تغییر موضوع قبلی، بازگشت به یکی از موضوعات بحث شده در قبل و یا درخواست شفاف‌سازی باشد. بنابراین مدل باید به نحوی بتواند برای این موارد راه حلی ارائه کند.[13]



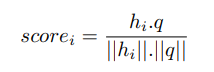
1شکل - نمونه ای از سه حالت تغییر تمرکز مکالمه در دادگان QuAC

هر یک از پژوهش‌های ارائه شده در امر پرسش‌و‌پاسخ مکالمه‌ای، سعی میکنند با استفاده از ارائه روشی برای چالش‌های بالا، مدل‌هایی با کارایی مناسب ارائه کنند. در این قسمت به بررسی برخی از مهم‌ترین مدل‌هایی که بر پایه‌ی مدل برت هستند می‌پردازیم.

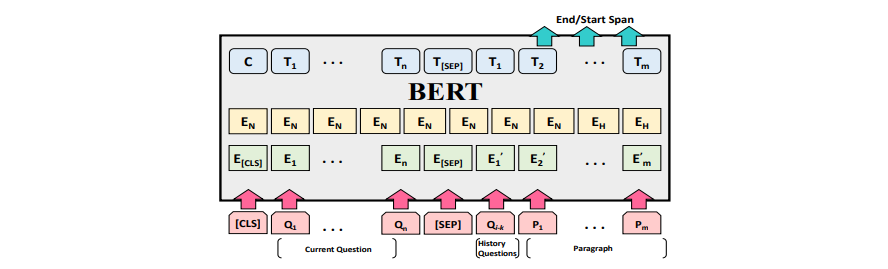
۲.۲.۱ مدل BERT-CoQAC

در مقاله [5] محققین سعی در انجام آزمایشاتی برای پی بردن به تأثیر انتخاب مناسب تاریخچه داشته‌اند که در نهایت به این نتیجه رسیده‌اند که با انتخاب ۵ مرحله از تاریخچه به دقت F1 بهینه می‌رسند که برابر با ۴/۶۳ درصد بر روی دادگان آزمایش QuAC است، به این معنا که در نظر گرفتن تمام تاریخچه می‌تواند باعث کاهش کارایی و دقت مدل شود.

مدل ارائه شده در این مقاله از برت به همراه یک ماژول انتخاب تاریخچه بر مبنای شباهت معنایی با مراحل قبلی تاریخچه استفاده میکند. این شباهت معنایی که به عنوان امتیاز از آن نام برده شده از طریق رابطه‌ی کسینوسی و به شکل زیر محاسبه می‌شود:



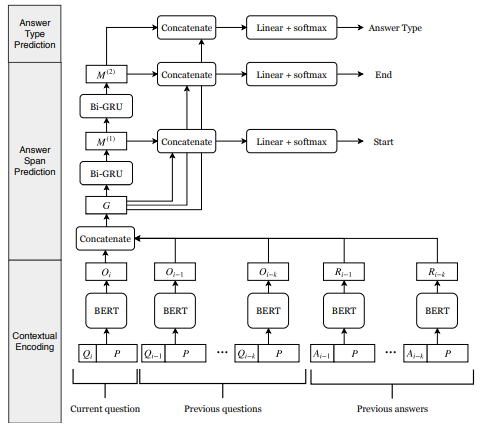
در این رابطه hi نمایانگر مراحل تاریخچه و q نماد سؤال فعلی است. در نهایت مراحلی با امتیاز بزرگتر از ۵/۰ به عنوان تاریخچه برای سؤال فعلی در نظر گرفته می‌شوند و تاریخچه به صورت روش نشانه گذاری پاسخ‌ها داخل متن مرجع کدگذاری می‌شود و سؤالات قبلی به همراه سؤال فعلی به ابتدای متن اضافه می‌شوند.



شکل 3- معماری مدل BERT-CoQAC و نحوه مدلسازی تاریخچه

۲.۲.۱ مدل BERT w/k-ctx

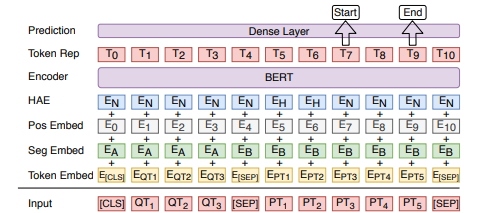
در مقاله [14] روشی پیشنهاد شده که اساس انتخاب تاریخچه آن بر پایه‌ی روش انتخاب تعدادی از آخرین مراحل می‌باشد. این مدل که از برت در لایه رمزگذاری‌اش استفاده میکند، برپایه‌ی یک عدد ثابت، مثلا ۲، مراحل انتهایی مکالمه را در نظر میگیرد و هر یک از سؤالات را به صورت جداگانه به ابتدای متن مرجع، پس از اضافه کردن توکن جداسازی برت [SEP] اضافه می‌کند و همین کار را برای هر یک از پاسخ‌های تاریخچه نیز به صورت جداگانه برای هر یک انجام میدهد. سپس هر یک از متون ایجاد شده را به تعدادی مدل برت میدهد که پارامترهای مشترک دارند و خروجی آنها را با هم ادغام میکند. در نهایت خروجی بدست آمده را در دو مرحله متوالی از واحد‌های دو طرفه‌ی GRU عبور می‌دهد و احتمال شروع و پایان را برای هر یک از توکن‌ها بدست می‌آورد. همچنین با استفاده از آخرین خروجی و یک شبکه عصبی سؤالاتی که پاسخ‌های خاص دارند (مثل بله/خیر/غیرقابل‌پاسخ) را از سؤالات بازه‌محور تشخیص داده و در‌صورت وجود، پاسخ مرتبط را به جای بازه‌ی بدست آمده به عنوان خروجی می‌دهد. در پژوهش مذکور تعدادی آزمایش برای بدست آوردن مقدار بهینه تاریخچه انجام شده که برای دادگان QuAC این عدد برابر با ۲ مرحله و برای دادگان CoQA عدد ۵ و ۷ مرحله بدست آمده. همچنین دقت بهینه‌ی F1 این مدل بر دادگان QuAC مقدار ۴/۶۵ درصد و برای CoQA مقدار ۷/۷۹ درصد بوده است.



شکل 4- معماری مدل BERT w/k-ctx

۲.۲.۱ مدل BERT-HAE

در مقاله [15] پژوهشگران یک روش نو برای مدلسازی تاریخچه ارائه کردند که همانطور که در ۲.۱.۳ ذکر شد این روش نشانه‌گذاری توکن‌های پاسخ در متن مرجع بود. این روش کمک شایانی به مدل‌های پرسش‌و‌پاسخ مکالمه‌ای کرد چرا که مدل برت را قادر می‌ساخت تا بتواند در چنین محیط‌هایی با شرایط مکالمه‌ای عملکردی طبیعی داشته باشد. بر طبق این روش توکن‌ها پیش از ورود به رمزگذار، علاوه بر تعبیه‌ی مکانی (positional embedding) ، تعبیه‌ی قسمی (segment embedding) ، و تعبیه‌ی کلیدی (token embedding) که در همه‌ی مدل‌های برت‌محور مشترک می‌باشند، داخل یک لایه شبکه عصبی تعبیه‌ی دیگری می‌شوند که نمایانگر وجود یا عدم وجود در پاسخ‌های تاریخچه است. روش ارائه شده دقت F1 برابر با ۴/۶۲ درصد برروی دادگان QuAC داشت و در شکل ۵ نشان داده شده.



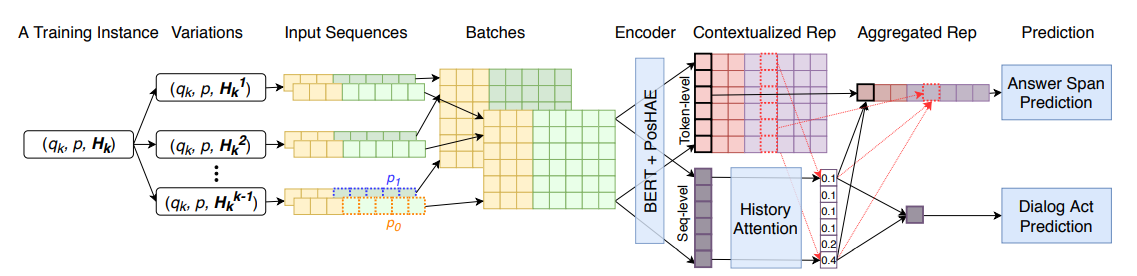
شکل 5-معماری مدل BERT-HAE

۲.۲.۱ مدل BERT-HAM

پژوهشگران مقاله‌ی [15] در مقاله‌ای که چند ماه پس از آن با عنوان انتخاب توجه‌محور تاریخچه [16] انتشار دادند دو نوآوری جدید ارائه نمودند. اولین نوآوری این بود که در اینجا علاوه بر نشانه‌گذاری پاسخ‌های تاریخچه که صرفا وجود و عدم وجود توکن‌ها را نشان میداد، یک نشانه‌گذاری دیگر ایجاد کردند که مرحله‌ای که توکن موردنظر در پاسخ آن سؤال ظاهر می‌شود را نیز در تعبیه‌ی خود داشته باشد. برای اجرای این روش یک مجموعه واژگان به صورت {0, 1, …, K}در نظر گرفته شد که هر واژه نشان‌دهنده‌ی یکی از مراحل قبلی است و واژه‌ی 0 نشان‌دهنده‌ی عدم وجود توکن در هیچ‌یک از پاسخ‌های مراحل قبلی می‌باشد. سپس یک تعبیه برای این مجموعه واژگان در یک شبکه عصبی آموخته می‌شود و توکن‌های خروجی از برت حاوی این نوع تعبیه خواهند بود.

علاوه بر این، برای انتخاب تاریخچه از یک روش وزن‌دار استفاده کردند که در آن هر یک از مراحل تاریخچه ابتدا در یک ماژول توجه به تاریخچه که بر مبنای مکانیزم توجه کار میکند، بر مبنای سؤال فعلی وزن‌دهی شده و در نهایت یک مجموع وزن‌دار از این مراحل بدست می‌آید. سپس این مقدار وارد یک شبکه عصبی شده و پیش‌بینی پاسخ بدست می‌آید.

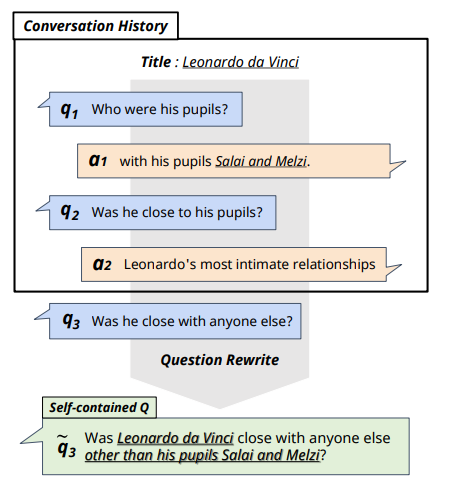
این مدل به دقت F1 برابر با ۷/۶۶ درصد برروی QuAC رسیده که بهبودی ۲ درصدی نسبت به مدل قبلی‌شان، BERT-HAE بوده است.



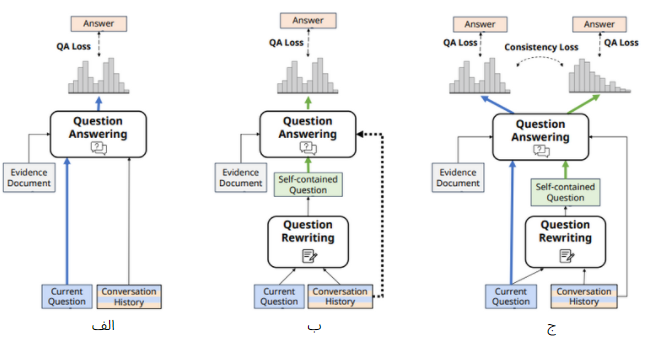
2معماری مدل BERT-HAM

۲.۲.۱ مدل ExCorD

علاوه بر روش‌هایی که در بخش ۲.۱ گفته شد روش‌هایی وجود دارند که سعی می‌کنند مسأله‌ی پرسش‌و‌پاسخ مکالمه‌ای را به یک مسأله‌ی پرسش‌و‌پاسخ ساده تبدیل کنند و به این ترتیب بخش‌های انتخاب تاریخچه و مدلسازی تاریخچه را حذف کرده و در نهایت از مدلهای موجود برای امر پرسش‌و‌پاسخ استفاده کنند. این نوع روش‌ها سعی می‌کنند با استفاده از حل چالش‌های هم‌مرجعی و پدیده‌ی آنافورا، هر یک از سؤالات مکالمه را به یک سؤال مستقل و یا به اصطلاح خود‌محور تبدیل کنند و سؤال ایجاد شده را به مدل بدهند.(شکل X) برای پیاده‌سازی این ایده، این روش‌ها با مسأله‌ای با نام بازنویسی سؤال (question rewriting)مواجه می‌شوند و برای این کار نیاز است یک مدل آموزش داده شود که امر مورد نظر را انجام دهد. بطور نمونه در مقاله‌ی [20] یک مدل برای بازنویسی سؤال برروی دادگان CANARD [21] آموزش داده شده و سپس از مدل حاصل در یک مدل سؤال‌و‌جواب استفاده شده. اما در مقاله‌ی [22] مؤلفین اذعان کردند که استفاده‌ی تنها از بازنویسی سؤال برای پرسش‌و‌پاسخ مکالمه‌ای باعث وابستگی این مدل‌ها به مدل‌های بازنویسی سؤال می‌شود و می‌تواند خطای مدل‌های بازنویسی را منتشر کند. به همین دلیل معماری دیگری پیشنهاد کرده‌اند که در آن هم سؤال اصلی و هم سؤالِ بازنویسی‌شده به مدل داده می‌شوند، اما تحت یک نوع منظم‌سازی سازگاری (consistency regularization)، مدل را تشویق می‌کنند که با استفاده از تنها سؤال اصلی پاسخ‌هایی را تولید کند که با سؤال بازنویسی‌شده تولید می‌شوند. این مدل توانست دقت ۷/۶۷ درصد را برروی دادگان QuAC بدست آورد.



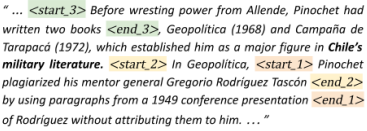
3نمونه ای از تبدیل سؤالات مکالمه به سؤال مستقل



4- (الف) معماری مدل‌های بر‌پایه انتخاب تاریخچه (ب) معماری مدل‌های برپایه بازنویسی سؤال (ج) معماری مدل ExCorD

۲.۲.۱ مدل MarCQAp

اخیرا در مقاله‌ای با عنوان مدلسازی مؤثر زمینه برای پرسش و پاسخ مکالمه‌ای، که هنوز به طور رسمی منتشر نشده دو نو آوری ارائه شده. اولین آنها استفاده از نوع جدیدی از مدلسازی تاریخچه در داخل متن است. در این نوع مدلسازی، حول قسمت‌هایی از از متن که در پاسخ‌های مراحل پیشین مکالمه وجود داشته‌اند دو توکن خاص گذاشته می‌شوند که مشخص کننده‌ی حضور آن بازه از متن در آن مرحله از مکالمه‌ است. همچنین این مدل از معماری خاصی از ترنسفورمر به نام لانگفورمر (Longformer) استفاده می‌کند که دارای مکانیزم توجه تُنک بوده و باعث کاهش پیچیدگی زمانی آن شده و اجازه‌ی پردازش متون بلند تا اندازه‌ی ۴۰۹۶ توکن را می‌دهد. مدل MarCQAp دقت ۸/۷۰ درصد را روی دادگان QuAC بدست‌ آورده که قابل توجه است.



5-نحوه مدلسازی تاریخچه داخل متن مرجع توسط مدل MarCQAp

**۳ راهکار پیشنهادی**

همانطور که در بخش ۱.۲ ذکر شد، هدف ما بهبود قسمت انتخاب تاریخچه‌ی مرتبط با سؤال فعلی در روش ارائه شده در مدل BERT-CoQAC است. این مدل تاریخچه را بر اساس شباهت معنایی که برپایه‌ی رابطه‌ی کسینوسی بدست می‌آید انتخاب می‌کند و آن را به روش نشانه‌گذاری پاسخ‌ها درون متن کدگذاری می‌کند. هدف ما بهبود معیار استفاده شده برای بررسی شباهت معنایی سؤال فعلی و تاریخچه است. همچنین برای انتخاب تاریخچه از ضمایر هم‌مرجع نیز کمک میگیریم. در حالیکه مدل BERT-CoQAC یک آستانه برای انتخاب تاریخچه بر اساس شباهت معنایی در نظر میگیرد، هدف ما یادگیری این آستانه از طریق روش‌های نزول در امتداد گرادیان است. به علاوه برای نشانه گذاری توکن‌های پاسخ داخل متن مرجع، از روش ارائه شده در مدل MarCQAp کمک میگیریم، چرا که نه تنها در این روش کدگذاری هر توکن حاوی اطلاعاتی راجع به شماره‌ی مرحله‌ی بکار رفته در آن خواهد بود،‌ بلکه هر توکن می‌تواند متعلق به چند مرحله باشد که مدلسازی مناسبی از محیط پرسش‌و‌پاسخ مکالمه‌ای است. به علاوه قصد داریم از مدل لانگفورمر استفاده کنیم، چرا که در صورت استفاده از مدل‌های دیگر مثل برت پایه، با محدودیت ۵۱۲ توکن مواجه خواهیم بود و برای رفع آن لازم است از تکنیک تقسیم متن به قسمت‌های کوچکتر استفاده کنیم، اما بررسی‌های انجام شده نشان می‌دهد این تکنیک دقت را اندکی پایین می‌آورد.

**۴ نتایج آزمایش‌ها**

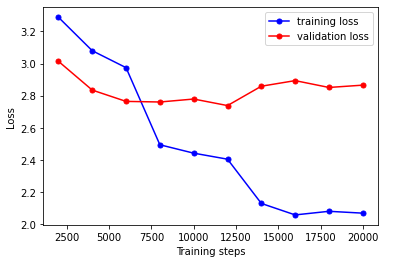
در این مرحله به پیاده‌سازی یک مدل ساده‌ی پرسش‌و‌پاسخ برروی دادگان CoQA پرداختیم. ابتدا دادگان CoQA را پیش پردازش کردیم و در مرحله‌ی اول از آنجا که داده‌ها به صورت مکالمه‌ای هستند و هر متن با تعدادی سؤال می‌آید، این وابستگی را از بین بردیم و متن مورد نظر را برای هر سؤال کپی کرده و یک جدول تخت تشکیل دادیم که دارای ستون‌های متن مرجع، سؤال، شروع و پایان پاسخ می‌باشد. در مرحله‌ی بعدی مقادیر نوشته‌شده در داده‌های آموزش برای شروع و پایان بازه‌ی پاسخ که بر حسب کاراکتر بودند را با توجه به توکنیزه‌کننده‌ی (tokenizer)خودمان به مقادیر برحسب توکن تبدیل کردیم. در نهایت 108647 نمونه آموزش و ۷۹۸۳ نمونه‌ی اعتبارسنجی بدست آمد.

در ادامه از یک مدل برت پایه (پاورقی:bert-base-uncased) از ‌پیش آموزش‌دیده استفاده نمودیم و ابتدا این مدل را با تنظیمات بدون نظارت و به صورت مدلسازی سازی زبانی علّی (causal language modeling)و سپس با مدلسازی زبانی ماسکه (masked language modeling) و برروی متون مرجع دادگان آموزشی پیش‌آموزش دادیم.

سپس از مدل پیش‌آموزش داده شده استفاده کردیم و با گذاشتن یک لایه شبکه عصبی با ابعاد 768×2 جلوی مدل برت شرایط آموزش مدل پرسش‌و‌پاسخ را آماده کردیم. سپس مدل مذکور را به اندازه‌ی ۳ دوره آموزش دادیم و در نهایت برروی دادگان اعتبار سنجی و با استفاده از قطعه برنامه‌ی ارائه‌شده توسط برای ارزیابی دادگان CoQA عملکرد مدل را بدست آوردیم که حاصل آن در جدول X قابل مشاهده است.

|  |  |  |  |
| --- | --- | --- | --- |
| **تعداد نمونه‌ها** | **%EM** | **%F1** | **موضوع** |
| 1425 | 45.6 | 69.6 | Children stories |
| 1630 | 34.7 | 57.3 | Literature |
| 1653 | 37.7 | 61.1 | Mid high school |
| 1649 | 35.2 | 61.8 | News |
| 1626 | 43.9 | 70.8 | Wikipedia |
| 7983 | 39.3 | 64.0 | Overall |

شکل 6- عملکرد مدل برروی دادگان اعتبارسنجی CoQA



شکل 7- نمودار تابع هزینه هنگام آموزش مدل

همچنین تلاش‌هایی که برای تست با مدل لانگفورمر صورت گرفت بدلیل زمان آموزش بالا و سخت‌افزار ناکافی، با ناکامی مواجه شد و نتیجه‌ای دربر نداشت، فلذا تست‌های این مدل به آینده موکول می‌شود.

**۵ کارهای آتی**

در این پژوهش به بررسی معماری و نوآوری‌های مدل‌های روز در حوزه‌ی پرسش‌وپاسخ مکالمه‌ای پرداختیم و دقت‌های بدست آمده توسط هریک از آنها را مرور کردیم. سپس روشی برای بهبود دقت این مدل‌ها ارائه کردیم. در نهایت در جدول X میزان پیشرفت و زمان‌بندی برای ادامه این پژوهش آورده شده است.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ردیف** | **عنوان** | **مدت زمان لازم** | **پیشرفت** | **زمان شروع** |
| ۱ | مطالعه پژوهش‌های پیشین | چهار ماه | ۸۰٪ | مرداد ۱۴۰۰ |
| ۲ | نگارش دو فصل پایان‌نامه | دو ماه | ۱۰۰٪ | آذر ۱۴۰۰ |
| ۳ | پیاده‌سازی اولیه | یک ماه | ۱۰۰٪ | دی ۱۴۰۰ |
| ۴ | پیاده‌سازی روش پیشنهادی | ۲ ماه | ۱۰٪ | اسفند ۱۴۰۰ |
| ۵ | نگارش فصل‌های پایانی پایان‌نامه | ۱ ماه | ۰٪ | اردیبهشت ۱۴۰۱ |
| ۶ | آماده شدن برای دفاع | ۱ ماه | ۰٪ | خرداد ۱۴۰۱ |

**6** جدول - میزان پیشرفت و زمان‌بندی ادامه‌ی پژوهش

مراجع

مراجع

واژه نامه