Journal of Soft Computing and Information Technology (JSCIT)
Babol Noshirvani University of Technology, Babol, Iran
Journal Homepage: jscit.nit.ac.ir
Volume 12, Number 3, Autumn 2023, pp. 45-57
Received: 04/18/2023, Revised: 10/28/23, Accepted: 01/01/2024

A Method Based on Machine Reading Comprehension to Answer Multi-hop Complex Questions in Question Answering Systems

Arash Ghafouri [1], Hasan Naderi[2] and Behrouz Minaei Bidgoli [3*]
Computer Engineering, Iran University of Science and Technology, Tehran, Iran.
[1]aghafuri@comp.iust.ac.ir, [2]naderi@iust.ac.ir, [3*]b_minaei@iust.ac.ir

Corresponding author's address: Behrouz Minaei Bidgoli, Faculty of Computer Engineering, Iran University of Science and Technology ,Tehran, Iran.

Abstract- Question Answering Systems (QAS) are one of the most important intelligent systems that have the ability to provide instant and explicit answers to input questions. One of the new challenges of these systems is the ability to answer multi-hop complex questions that require information collection from multiple documents. In this paper, a method is proposed to solve the challenge of answering multi-hop complex questions. In this method, first, documents related to the question are retrieved in a two-step process. Then, to facilitate answering the question, knowledge is extracted from the retrieved documents and represented in the form of a graph. Finally, to find the answer to the question, reasoning is performed on the graph using a combination of heterogeneous graph neural network and transformer. To evaluate the effectiveness of the proposed method, this method and other related works have been tested on the HotpotQA open-domain dataset. The results obtained in answering questions based on F1 score and exact match are reported as 86.51 and 78.71 respectively, indicating superiority of this method over similar works.

Keywords- Question Answering Systems (QAS), Multi-hop Complex Questions, Machine Reading Comprehension (MRC), Heterogeneous Graph Neural Network (HGNN), Transformers.

مجله علمی پژوهشی رایانش نرم و فناوری اطلاعات
دانشگاه صنعتی نوشیروانی بابل
صفحه مجله: jscit.nit.ac.ir
جلد ۱۲، شماره ۳، پائیز ۱۴۰۲ ، صفحه ۴۵-۵۷
دریافت: ۱۴۰۲/۰۱/۲۹، بازنگری: ۱۴۰۲/۰۸/۰۶، پذیرش: ۱۴۰۲/۱۰/۱۱

ارائه روشی مبتنی بر درک مطلب ماشینی برای پاسخدهی به پرسش‌های پیچیده چندگامی در سامانه‌های پرسش و پاسخ

آرش غفوری[۱]، حسن نادری[۲] و بهروز مینایی بیدگلی*[۳]
دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران، ایران.
[1]aghafuri@comp.iust.ac.ir, [2]naderi@iust.ac.ir and [3]*b_minaei@iust.ac.ir

* نشانی نویسنده مسئول: بهروز مینایی بیدگلی، دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران، ایران.

چکیده- سامانه‌های پرسش‌وپاسخ به عنوان یکی از مهمترین سامانه‌های هوشمند توانایی پاسخگویی آنی و صریح به پرسش‌های ورودی را دارند. یکی از چالش‌های جدید این سامانه‌ها، قدرت پاسخگویی به پرسش‌های پیچیده چندگامی است که نیازمند جمع‌آوری اطلاعات از چندین سند است. در این مقاله یک روش برای حل چالش پاسخدهی به پرسش‌های پیچیده چندگامی ارائه می‌گردد. در این روش، ابتدا در یک فرایند دومرحله‌ای، اسناد مرتبط با پرسش بازیابی می‌شوند. سپس برای تسهیل در پاسخ‌دهی به پرسش، نسبت به استخراج دانش از اسناد بازیابی شده و بازنمایی آن در قالب یک گراف اقدام می‌گردد، در نهایت و برای یافتن پاسخ پرسش، استدلال روی گراف با ترکیب شبکه عصبی گرافی نامتجانس و مبدل انجام می‌پذیرد.

به منظور بررسی میزان کارآمدی روش پیشنهادی، این روش و سایر کارهای مرتبط بر روی مجموعه‌داده دامنه‌باز هات‌پات‌کیوای مورد آزمایش قرار گرفته است. نتایج بدست آمده در پاسخدهی به پرسش بر مبنای معیار اف‌وان و تطبیق دقیق به ترتیب مقادیر ۸۶٫۵۱ و ۷۸٫۷۱ گزارش شده است که نشانگر برتری این روش نسبت به سایرکارهای مشابه به خود است.

واژه‌های کلیدی: سامانه‌های پرسش و پاسخ، پرسش پیچیده چندگامی، درک مطلب ماشینی، شبکه عصبی گرافی نامتجانس، مبدل‌ها.

۱- مقدمه

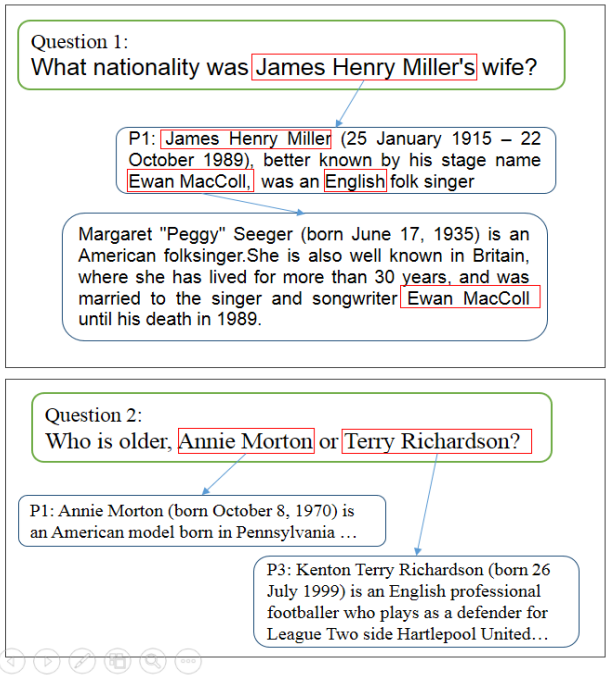
سامانه‌های پرسش‌وپاسخ امروزه به یکی از مهم‌ترین و کاربردی‌ترین سامانه‌های هوشمند تبدیل شده‌اند. هر زمینه‌ای که در آن تعامل به صورت پرسش و پاسخی بین انسان‌ها به‌منظور کسب اطلاعات باشد، سامانه‌های هوشمند پرسش‌وپاسخ می‌توانند این کار را به‌درستی و با اعتماد بیشتر انجام دهند. چنین سامانه‌هایی می‌توانند از تعامل انسانی نیز بهتر عمل کرده و مطمئن‌تر ظاهر شوند.

سامانه‌های پرسش و پاسخ نوع پیشرفته‌ای از موتورهای جست‌وجو هستند که سعی دارند پرسش‌های کاربر را در زبان طبیعی به عنوان ورودی دریافت نموده و پاسخی دقیق برگردانند. برخلاف موتورهای جست‌وجوی سنتی این سامانه‌ها با تعدادی عبارات کلیدی، لیستی از اسناد را برنمی‌گردانند بلکه با استفاده از درک زبان طبیعی مستقیماً پاسخی صحیح به پرسش کاربران داده می‌شود.

سامانه‌های پرسش و پاسخ علی رغم کارکرد خوب در پاسخدهی به پرسش‌های ساده، دارای مشکلات متعددی در پاسخدهی به انواع پرسش‌های پیچیده هستند. پرسش‌های پیچیده به انواع مختلفی تقسیم می‌شوند که از جمله مهمترین آن‌ها می‌توان به پرسش‌های مقایسه‌ای (تصدیقی و یا تاییدی)، چند گزینه‌ای، پل و نوع آزاد (بدون محدودیت) اشاره کرد. در پرسش‌های مقایسه‌ای، معمولاً چند موجودیت یا موضوع با یکدیگر مقایسه می‌شوند و ممکن است یک موضوع تصدیق یا رد گردد. در این حالت، پاسخ به پرسش می‌تواند بله یا خیر باشد.

در مورد پرسش‌های پل، آن‌ها از دو یا چند پرسش ساده تو در تو

تشکیل شده‌اند. به عبارت دیگر، یک پرسش پل قابل تجزیه به دو پرسش ساده‌تر است و پاسخ به پرسش ساده اول بخشی از پاسخ به پرسش ساده دوم خواهد بود. منظور از پرسش‌های پیچیده چندگامی همان پرسش‌های پیچیده مقایسه‌ای و پل است و در شکل ۱ نمونه ای از هر دو نوع پرسش قابل مشاهده است. پاسخدهی به پرسش‌های پیچیده چندگامی، سامانه‌های پرسش و پاسخ نیازمند بررسی حداقل دو پاراگراف متفاوت هستند. این سامانه‌ها در ادامه برای پاسخ‌دهی به پرسش‌ها، با استدلال روی پاراگراف‌ها و پرش از یک پاراگراف به پاراگراف‌های دیگر در همان سند یا اسناد دیگر اقدام می‌نمایند، که به این قابلیت استدلال چندگامی نیز گفته می‌شود.



شکل ۱: نمونه‌ای از پرسش‌های پیچیده چندگامی. پرسش ۱: پرسش پیچیده نوع پل. پرسش۲: پرسش پیچیده نوع مقایسه‌ای

در سامانه‌های پرسش و پاسخ سنتی مبتنی بر درک مطلب ماشینی، پرسش‌ها نیازمند یک استدلال ساده برای رسیدن به پاسخ بودند و برای پاسخگویی به اکثر این پرسش‌ها می‌توان با تطبیق پرسش و محتوا به پاسخ رسید. به همین خاطر شکافی بین عملکرد انسانی و ماشینی وجود داشت.

درک مطلب ماشینی[۱]، یک شاخه پژوهشی مهم در پردازش زبان طبیعی است و کاربردهای وسیعی در صنعت دارد. در سال‌های اخیر این شاخه پژوهشی، عمدتا به‌دلیل ظهور مجموعه‌داده‌هایی در مقیاس بزرگ و همچنین قرار گرفتن در عصر جدید الگوریتم‌های یادگیری عمیق پیچیده، پیشرفت‌های زیادی داشته است. درحال حاضر بسیاری از مدل‌های درک مطلب ماشینی، از عملکرد انسانی در مجموعه‌داده‌های مختلف پیشی گرفته‌اند و این درحالی است که شکاف بزرگ و آشکاری بین مدل‌های درک مطلب ماشینی و درک مطلب انسانی واقعی وجود دارد. این شکاف نیازمند مجموعه‌داده‌های غنی است که مدل‌های درک مطلب ماشینی را به سمت درک واقعی سوق دهد.

امروزه سامانه‌های پرسش و پاسخ نوین برای رفع چالش‌های خود ضمن ایجاد یا انتخاب داده‌های مناسب، برای رسیدن به پاسخ مناسب از روش‌های درک مطلب ماشینی استفاده می‌کنند تا ضمن پوشش قابلیت‌های سامانه‌های پرسش و پاسخ سنتی از پرسش‌های پیچیده چندگامی نیز پشتیبانی شود.

سازماندهی مقاله در ادامه به صورت زیر است. در بخش بعد کارهای مرتبط با پژوهش انجام شده بررسی می‌گردد. در بخش سوم روش پیشنهادی، معماری و اجزای آن به صورت کامل تشریح می‌گردد. بخش چهارم به آزمون روش پیشنهادی پرداخته و نتایج بدست آمده را گزارش می‌نماید. در ادامه و در بخش پنجم ضمن نتیجه‌گیری از پژوهش انجام شده، پیشنهاداتی به عنوان کارهای آتی برای ادامه پژوهش بیان می‌گردد.

۲- کارهای مرتبط

پژوهش‌های متعدد و با رویکردهای متفاوت برای توسعه سامانه‌های پرسش‌وپاسخ و پاسخدهی به پرسش انجام شده است. در این بخش پژوهش‌های انجام شده روی سامانه‌های پرسش و پاسخ که مبتنی بر درک مطلب ماشینی هستند و برای پاسخدهی به پرسش نیازمند استدلال روی چندین سند دارند مورد بررسی و مقایسه قرار می‌گیرند. در ادامه پژوهش‌های مرتبط در دو دسته مجموعه‌داده‌ها و روش‌های مشابه معرفی می‌گردند.

۲-۱- مجموعه‌داده‌های مرتبط

مهم‌ترین جنبه هوشمندی یک سامانه توانایی انجام استدلال و استنتاج زبان طبیعی است. پرسش‌وپاسخ وظیفه‌ای است که می‌تواند به‌صورت عینی و کمی توانایی استدلال یک سامانه هوشمند را بسنجد.

در بخش قبلی به شکاف میان سامانه‌های درک مطلب ماشینی با درک انسانی اشاره گردید و یکی از مهمترین ملزومات برای رفع این شکاف، مجموعه‌داده‌های غنی برشمرده شد. در ادامه انواع مجموعه‌داده‌های مرتبط با پژوهش انجام شده معرفی می‌گردد.

۲-۱-۱- درک مطلب ماشینی ساده

به عنوان اولین نوع اکثر مجموعه‌داده‌های درک مطلب ماشینی سنتی مثل اسکواد[۲][۱] با محوریت استدلال بر روی یک پاراگراف در یک سند یا به اصطلاح استدلال تک‌گامی ایجاد شده‌اند. اسکواد

و داده‌های مشابه برای پاسخگویی به پرسش‌ها به یک تک‌پاراگراف به‌عنوان محتوا نیاز دارند. بیشتر پرسش‌ها با تطبیق پرسش و تک‌جمله‌ای در پاراگراف، قابل پاسخگویی هستند و این داده‌ها محک خوبی برای هوشمندی سامانه‌ها نیستند. به رغم این‌که اسکواد توانایی استدلال در یک محتوای بزرگ و پیچیده را ندارد، جزو معروف‌ترین مجموعه‌داده‌ها است. با پیشرفت پردازش زبان طبیعی سامانه‌های ادراکی قوی‌تری نسبت به انسان بر پایه این مجموعه‌داده پدید آمده است.

داده‌های تریواکیوای[۳] [۲] و سرچ‌کیوای[۴] [۳] برای حل چالش مذکور روی بازیابی اسناد تمرکز کردند طوری‌که برای جواب دادن به یک پرسش نیاز به بازیابی چندین سند بود، با این حال برای پاسخ دادن به بیشتر پرسش‌ها، تطبیق دادن جملات نزدیک به هم در یک پاراگراف کافی بود.

۲-۱-۲- درک مطلب ماشینی پیچیده

نوع دوم مجموعه‌داده‌ها، داده‌های‌هایی مثل کیوای‌انگرو[۵][۴] و کامپلکس‌وب‌کوسشن[۶] [۵] هستند، که بر مبنای یک پایگاه‌دانش تشکیل یافته‌اند. سامانه‌های ساخته‌شده با این نوع مجموعه‌داده محدود به پایگاه‌دانش هستند و علاوه‌براینکه از تنوع پرسش و پاسخ‌های کمی برخوردارند، به‌روز ماندن‌شان نیز هزینه‌بر است.

مجموعه‌داده ویکی‌هاپ[۷] [۴] از مجموعه‌داده کیوای‌انگرو حاوی پرسش‌هایی به شکل سه‌گانه و بر اساس یال‌های یک گراف دانش ساخته شده است. با این حال، ویکی‌هاپ فاقد پرسش‌هایی با متن طبیعی و برچسب‌زنی در جملاتی است که می‌تواند برای پاسخ به این پرسش‌ها استفاده شود. کامپلکس‌وب‌کوسشن با تبدیل مسیرهای چندگامی در یک پایگاه دانش به یک پرسش متنی، ساخته شده است [۶]. با این نوع ساخت، پرسش‌ها را می‌توان با کمک یال‌های گراف دانش به پرسش‌های ساده‌تری تجزیه کرد.

دراپ[۸][۷] به استدلال گسسته روی متن (مانند شمارش یا جمع) نیاز دارد. تمرکز آن بر انجام عملیات گسسته ریاضی بر روی اطلاعات استخراج شده است.

مالتی آرسی[۹] [۸] از اسناد متنی برای ایجاد پرسش‌های چندگامی استفاده کرده است. با این حال، مالتی آرسی و دیگر مجموعه‌داده‌های تک‌سندی مثل ([۹] و [۱۰]) به‌جای ترکیب رابطه، تاکید بر ردیابی موجودیت[۱۰] دارند.

تمرکز پژوهش ما در این مقاله روی پرسش‌وپاسخ چندگامی است و مهمترین چالش آن پیچیدگی استدلال است. برای یک سامانه پرسش و پاسخ چندگامی پیدا کردن تکه‌هایی از اطلاعات به‌عنوان شواهد و استدلال روی این شواهد چالش‌برانگیز است و ردیابی موجودیت یا نگاشت جملات به تنهایی جواب نخواهد داد.

هات‌پات‌کیوای[۱۱] [۱۱] یک مجموعه‌داده مناسب برای روش‌های پرسش و پاسخ چندگامی مبتنی بر متن[۱۲] است که امکان یادگیری بانظارت از جملات دارای شواهد را برای پاسخدهی به پرسش، فراهم نموده است. هات‌پات‌کیوای تنها مجموعه‌داده ارائه شده برای پژوهش‌های پرسش و پاسخ چندگامی همچون پژوهش ما است. پیدا کردن شواهد در این داده‌ها به‌علت اینکه سرنخ‌های رسیدن به جواب مستقیماً در پرسش ذکر نشده‌اند، آسان نیست. از این‌رو نیازمند استدلال پیچیده‌تر و در سطح معنا است.

بیشتر مجموعه‌داده‌هایی مثل کیوای‌انگرو، هایبریدکیوای[۱۳] [۱۲] ،اپن‌بوک کیوای[۱۴] [۱۳] و دراپ حاوی جملات برچسب‌دار نیستند که بتوانند برای تولید پاسخ به‌عنوان شواهد استفاده شوند. این کاستی، مدل‌ها را مجبور به استدلال بدون نظارت مستقیم می‌کند.

۲-۲- روش‌های پژوهشی مرتبط

در استدلال مبتنی بر درک مطلب ماشینی که موضوع این پژوهش است، فرآیند رسیدن به پاسخ مهم است و در پژوهش‌های مختلف روش‌های گوناگونی بدین منظور ارائه شده است که در ادامه به آن اشاره می‌گردد.

در روش‌پایه[۱۵] یا اولین روش پژوهشی که مجموعه‌داده هات‌پات‌کیوای را مورد محک‌زنی قرار داده است یک دوباره‌سازی از مدل آقای کلارک و همکاران [۱۴] انجام شده است. تکنیک‌های برجسته‌ای همچون مدل‌های سطح کاراکتر، خودتوجه[۱۶] و توجه دوطرفه[17] در این مدل استفاده شده است و ترکیب این تکنیک‌ها موجب کسب نتایج خوبی روی داده‌های مذکور شده است. برای تشخیص جواب در آن روش بعد از یک شبکه همگشتی[۱۸]، طبقه‌بند سه مسیره برای تولید احتمال جواب تعبیه شده است.

در روشی دیگر معماری و مدلی به نام دی‌اف‌جی‌ان[۱۹] [۱۵] مبتنی بر برت[۲۰] [۱۶] ارائه گردید که با گرفتن پرسش و پاراگراف‌های کاندید، ضمن ساخت گراف موجودیت مبتنی بر شواهد استخراج‌شده از پاراگراف‌ها مرتبط‌ترین پاسخ را بر می‌گرداند. هر چند این روش با تقلیل محتوای چندسندی به نمونه‌های تک‌سندی توانسته است، مشکل محدود کننده مبدل را حل کند ولی می‌توان گفت ارتباطات بین سندی درنظر گرفته نمی‌شوند.

در روشی دیگر مدلی مبتنی بر مبدل‌ها[۲۱] به نام اس‌ای‌ایی[۲۲] [۱۷] ارائه شده است. این مدل هر پاراگراف کاندید و پرسش را به برت داده و بردارهای جانمایی آن‌ها را به‌دست می‌آورد. این مدل از یک خودتوجه چندواحدی استفاده می‌کند تا بتواند دو به دو مرتبط بودن اسناد را سنجیده و دو پاراگراف طلایی را تشخیص دهد. این

مدل با وجود دخالت دادن ارتباطات بین سندی در قالب سازوکار توجه، هزینه محاسباتی بالای درجه دو داشته که زمان استدلال را بالا می‌برد.

اس‌آرال‌جی‌آران[۲۳] [۱۸] مدل بر پایه مبدل دیگری است که روش کار مشابهی با دی‌اف‌جی‌ان دارد. در این معماری هر یک از پاراگراف‌های کاندید به همراه پرسش به مدل طبقه‌بندی با کدگذار برت داده می‌شود و بالاترین پاراگراف مرتبط به عنوان اولین پاراگراف انتخاب می‌شود. سپس با حذف پاراگراف انتخاب شده از لیست کاندیداها این فرآیند را تکرار می‌کند تا دومین بالاترین پاراگراف را هم انتخاب کند. مدل مذکور برای پیش‌بینی جواب بعد از گذر پرسش از یک طبقه‌بند مبتنی‌بر برت استفاده نموده است. در ادامه برای پرسش‌ها نیز با یک مدل مبتنی‌بر برت دیگر، ابتدا و انتهای گستره پاسخ در محتوا را پیش‌بینی کرده است.

مدل کیواف‌ای[۲۴] [۱۹] شباهت بسیار زیادی به معماری روش‌پایه هات‌پات‌کیوای دارد. تفاوت آن با روش‌پایه این است که محتوای ورودی و پرسش را یکبار در سطح کاراکتر و بار دیگر در سطح کلمه جانمایی می‌کند. این مدل یک لایه برای تطبیق جانمایی‌های محتوا و پرسش دارد که از توجه دوجهته و شبکه عصبی همگشتی[۲۵] استفاده کرده است. درنهایت برای هرجمله احتمال مرتبط بودن با پرسش را محاسبه می‌شود تا محتوای لازم برای مراحل بعدی آماده شود.

مدل دی‌کامپ‌آرسی[۲۶] [۲۰] روش متفاوتی برای حل چالش پاسخگویی به پرسش‌های چندگامی طرح کرده است. از آنجایی که پرسش‌ها پیچیده هستند، پرسش به چند پرسش تک‌گامی تجزیه می‌شود. در آن روش از برت برای جانمایی و پیدا کردن گستره پاسخ استفاده شده است. همچنین برای پاسخگویی به زیر پرسش‌های از ترکیب مدل درک مطلب ماشینی مبتنی‌بر برت و رویکرد انتخاب پاراگراف کلارک و همکاران استفاده کرده است. هر تجزیه شامل زیرپرسش‌ها و پاسخ‌ها است. دی‌کامپ‌آرسی تجزیه‌ها را رتبه‌بندی کرده و پاسخ تجزیه با امتیاز بالا را خروجی می‌دهد.

در پژوهشی دیگر با عنوان درک مطلب ماشینی چندگامی بدون گراف یا جی‌اف‌ام‌اچ‌آرسی[۲۷][۲۱]، برخلاف مدل‌های سنتی که مبتنی بر گراف هستند، از رویکرد جدیدی با نام استراتژی انتخاب به عنوان راهنما یا اس‌توجی[۲۸] استفاده می‌کند. در این رویکرد، به جای استفاده از گراف برای درک مطلب ماشینی، یک سازوکار توجه از درشت به ریز[۲۹] برای شناسایی مجموعه‌ای از پاراگراف‌های کاندید که احتمالاً حاوی پاسخ به یک پرسش هستند، به کار می‌رود. سپس مدل از یک سازوکار توجه برای انتخاب جملاتی که از میان پاراگراف‌های کاندید مرتبط‌تر استفاده می‌کند. در نهایت، مدل از یک شبکه نقطه‌گذاری برای پیش‌بینی پاسخ از جملات انتخاب شده استفاده می‌کند.

مدل گات[۳۰] [۲۲] دو توسعه به مدل شبکه گراف سلسله‌مراتبی برای پاسخ به پرسش‌های چندپرش در هات‌پات‌کیوای ارائه می‌دهد. توسعه اول شامل افزودن یال‌های جدید بین گره‌های پرسش و جملات متن می‌شود تا ساختار گراف سلسله‌مراتبی را تکمیل کند. توسعه دوم، سازوکار نوآورانه‌ای به نام گات معرفی می‌کند که می‌تواند نمایش‌های گره‌ها را به ترتیب بر اساس سطوح سلسله‌مراتبی به روز کند.

در پژوهشی دیگر آقای ین و همکاران[۲۳] به مطالعه در مورد استفاده از هموارسازی برچسب در پاسخدهی به پرسش‌های چندگامی پرداخته‌اند. نویسندگان یک تکنیک نرم‌سازی برچسب جدید به نام نرم‌سازی اف‌وان را پیشنهاد می‌کنند. این تکنیک یا سی‌توف‌ام‌اف‌وان[۳۱] مبتنی بر معیار اف‌وان، یک معیار عملکرد مشترک برای وظایف درک مطلب ماشینی است. این تکنیک با توزیع احتمالی برچسب‌های صحیح را با توجه به معیار اف‌وان آنها نرم‌سازی می‌کند.

در جدول ۱ روش‌های پژوهشی مرتبط یا روش‌های مبتنی بر درک مطلب ماشینی برای پاسخدهی به پرسش‌های پیچیده چندگامی در سامانه‌های پرسش و پاسخ معرفی و مقایسه شده است.

۲-۳- جمع‌بندی پژوهشی های مرتبط

مدل و معماری‌های معرفی‌شده، نمونه متنوع و مرتبطی از پژوهش‌های انجام‌شده روی داده‌های پرسش و پاسخ چندگامی هات‌پات‌کیوای بودند.

در بازیابی اسناد که اکثر مدل‌های مبتنی‌بر مبدل‌ها با محدودیت تعداد توکن ورودی مواجه بودند، معماری‌های دی‌اف‌جی‌ان و اس‌آرال‌جی‌آران مشکل محدودیت توکنی را حل کردند ولی ارتباطات بین‌سندی را به‌خوبی دخالت ندادند. مدل اس‌ای‌ایی با درنظر گرفتن ارتباطات میان اسناد، نتایج خوبی داشته است ولی با مقایسه دوبه‌دو این روش برای تعداد اسناد زیاد زمان‌بر و غیرمعقول می‌باشد. مدل کیواف‌ای و مدل روش‌پایه در مقاله مجموعه‌داده هات‌پات‌کیوای از کدگذاری کاراکترها و کلمات باهم استفاده می‌کند که یک کار زمان‌بری است و پیشرفت‌های اخیر در جانمایی محتوایی با استفاده از مبدل‌ها، می‌تواند اعمال شود تا سربار کدگذاری بیشتری ایجاد نشود.

جدول ۱: معرفی و مقایسه روش‌های پژوهشی مرتبط

نام مدل	سال ارائه	نقطه قوت	نقطه منفی	تکنیک کلی
روش‌پایه	۲۰۱۸	استفاده از تکنیک‌های مدل‌های سطح کاراکتر، خودتوجه[۳۲] و توجه دوطرفه[۳۳]	عملکرد ضعیف مدل به طرز قابل توجهی نسبته به عملکرد انسانی و عدم بهره‌گیری از بازنمایی های محتوایی جدید	شبکه‌های عصبی عمیق همگشتی سه مسیره
کیواف‌ای	۲۰۱۸	جانمایی محتوای ورودی و پرسش را یکبار در سطح کاراکتر و بار دیگر در سطح کلمه	عملکرد ضعیف مدل به طرز قابل توجهی نسبته به عملکرد انسانی و عدم بهره‌گیری از بازنمایی های محتوایی جدید	شبکه‌های عصبی عمیق همگشتی و توجه دوطرفه
دی‌اف‌جی‌ان	۲۰۱۹	ساخت گراف موجودیت مبتنی بر شواهد استخراج‌شده از پاراگراف‌ها	در نظر نگرفتن ارتباطات بین سندی	ساخت گراف و استفاده از مبدل‌ها
اچ‌جی‌ان	۲۰۱۹	ساخت گراف سلسله مراتبی شامل پرسش‌ها و پاسخ‌ها و پاراگراف‌ها	ساخت گراف غیر بهینه و ایجاد نویز با اضافه شدن گره‌های موجودیت به گراف	ساخت گراف سلسله مراتبی و استفاده از مدل زبانی برت
دی‌کامپ‌آرسی	۲۰۱۹	تجزیه پرسش پیچیده و به پرسش‌های ساده برای تسهیل در پاسخگویی به پرسش پیچیده	عملکرد بسیار ضعیف حتی نسبت به مدل پایه	تجزیه پرسش پیچیده و استفاده از مدل زبانی برت
اس‌آرال‌جی‌آران	۲۰۲۰	ساخت گراف سلسله مراتبی و دخالت دادن ارتباطات بین سندی	وجود تعداد زیاد گره و رابطه بی‌تاثیر در استدلال و پاسخدهی به پرسش و عدم تعیین اولویت برای گره‌ها	ساخت گراف و استفاده از برچسب زن نقش معنایی
اس‌ای‌ایی	۲۰۲۰	ساخت گراف از موجودیت‌های نامدار و عبارت های اسمی و دخالت ارتباطات بین سندی	داشتن مقایسه زیاد دو به دو برای لحاظ نمودن ارتباطات بین سندی و عدم امکان استفاده از روش در اسناد زیاد	ساخت گراف و استفاده از شبکه‌های عصبی گرافی پیچشی
جی‌اف‌ام‌اچ‌آرسی	۲۰۲۱	استفاده از مدل از شبکه اشاره‌گر برای پاسخدهی به پرسش‌ها	عدم امکان شناسایی همه روابط بین موجودیت‌ها و مفاهیمی لازم برای پاسخدهی به پرسش‌های پیچیده	سازوکار توجه درشت به ریز
گات	۲۰۲۳	استفاده از رویکرد سلسله‌مراتبی و یادگیری اهمیت هر موجودیت در زمینه پرسش	نقص در یافتن پاسخ در صورت به دلیل توزیع ناهمگون موجودیت‌ها در گراف	سازوکارتوجه گرافی برای سلسله مراتب
سی‌توف‌ام‌اف‌وان	۲۰۲۳	درک روابط بین چندین جمله	پیش‌بینی برچسب‌های نادرست و کاهش سرعت یادگیری	نرم‌سازی برچسب

مدل دی‌کامپ‌آرسی پرسش‌ها را به پرسش‌های ساده تجزیه می‌کند و سپس با مدل‌های از پیش‌آموش‌دیده برای پرسش‌های ساده، جواب را تولید می‌کند. این روش علاوه‌بر سربار اضافی تجزیه پرسش‌ها، وظیفه را روی این داده‌ها بیش‌برازش کرده است و توانایی تعمیم ندارد.

پژوهش‌های منتشر شده در سال اخیر نیز علی رغم استفاده از تکنیک‌های نوین همچون نرم‌سازی برچسب برای معیار اف‌وان و یا استفاده از تکنیک های جانمایی گرافی جدید همچنان در ساخت زیرساخت گرافی بهینه دچار مشکل هستند. آنها معمولا بدلیل برخورداری از گره ها و روابط زیاد دارای سربار زیاد و یا گره ها . روابط کم دارای ضعف در یافتن پاسخ دقیق هستند. از سویی دیگر در گراف‌های متنی موجود معمولا بدون داشتن هیچ اولویتی برای یافتن پاسخ از مدل‌های استدلال گرافی استفاده می‌شود و این منجر به کاهش دقت در پاسخدهی به پرسش‌های پیچیده می‌گردد. در ادامه سعی می‌گردد با ارائه روش پیشنهادی برای رفع چالش‌های ذکر شده تلاش گردد.

۳- روش پیشنهادی

در سامانه های پرسش و پاسخ، برای پاسخدهی دقیق به پرسش‌های چندگامی، نیاز به یک سازوکار مشخص متشکل از گام‌ها مختلف است. انجام وظایف درست در هریک از این گام منجر کارآمدی می‌گردد. از همین جهت معماری برای روش پیشنهادی اراده شده است که شکل ۲ نمایانگر شمای کلی آن است. روش پیشنهادی شامل سه مولفه انتخاب پاراگراف‌های مرتبط، ساخت گراف نامتجانس و پیش‌بینی پاسخ است. همان‌طور که در معماری روش پیشنهادی مشخص است با داشتن پرسش Q و مجموعه‌ای Nتایی از پاراگراف‌هایی که پاسخ پرسش به‌وسیله آن‌ها قابل بازیابی است می‌توان ضمن انجام استدلال ماشینی به پاسخ پرسش رسید. برای رسیدن به پاسخ ابتدا لازم است پاراگراف‌های مرتبط شناسایی گردد. سپس ضمن ساخت زیرساخت گرافی برای رسیدن به پاسخ، گستره دقیق پاسخ به‌وسیله ترکیب یک شبکه عصبی گرافی و مبدل‌ها پیش‌بینی و مشخص گردد.

روش پیشنهادی به منظور استدلال مبتنی بر درک مطلب ماشینی

از مدل ترکیبی شامل مبدل‌ها و شبکه عصبی گرافی استفاده نموده است و توانسته نتایج خوبی کسب نماید.

۳-۱- مشارکت علمی

در این پژوهش در یک سازوکار یکپارچه سعی گردیده است تا با انجام سه گام موثر حداکثر کارآمدی در پاسخدهی به پرسش‌های پیچیده چندگامی حاصل شود. در گام اول سعی گردیده است تا با استفاده از بهترین مبدل‌ها و در دو مرحله مرتبط‌ترین پاراگراف با پرسش پیچیده چندگامی شناسایی شود. در گام دوم با یک ایده جدید نسبت به تبدیل متن پاراگراف‌ها به یک گراف سلسله مراتبی نامتجانس برای پاسخدهی بهینه به پرسش‌ها اقدام گردید. در این مسیر و برای غنی کردن هرچه بیشتر این گراف از تکنیک‌هایی همچون تشخیص موجودیت‌های گراف، یافتن روابط بین موجودیت‌ها[۳۴] و یافتن موجودیت‌های هم مرجع استفاده[۳۵] می‌شود. ایده اصلی در گام دوم تشخیص نوع پاسخ با استفاده از خود پرسش مطرح شده است که این امر مانع از ساخت گراف بزرگ با تعداد گره‌ها و ارتباطات اضافی می‌شود. از سویی دیگر سعی شده است تا با شناسایی ارتباطات انواع گره‌ها در دانه‌بندی های مختلف یعنی پاراگراف، جمله و موجودیت بایکدیگر در امر استدلال محتوایی و شناسایی پاسخ تسهیل گردد. در گام سوم با استفاده از تکنیک‌های مختلف جانمایی گرافی سعی در بهره‌گیری کامل از زیرساخت گرافی برای پاسخدهی به پرسش‌های پیچیده چندگامی گردید. در ادامه روش پیشنهادی با جزییات بیشتر تشریح می‌گردد.

۳-۲- مولفه انتخاب پاراگراف‌های مرتبط

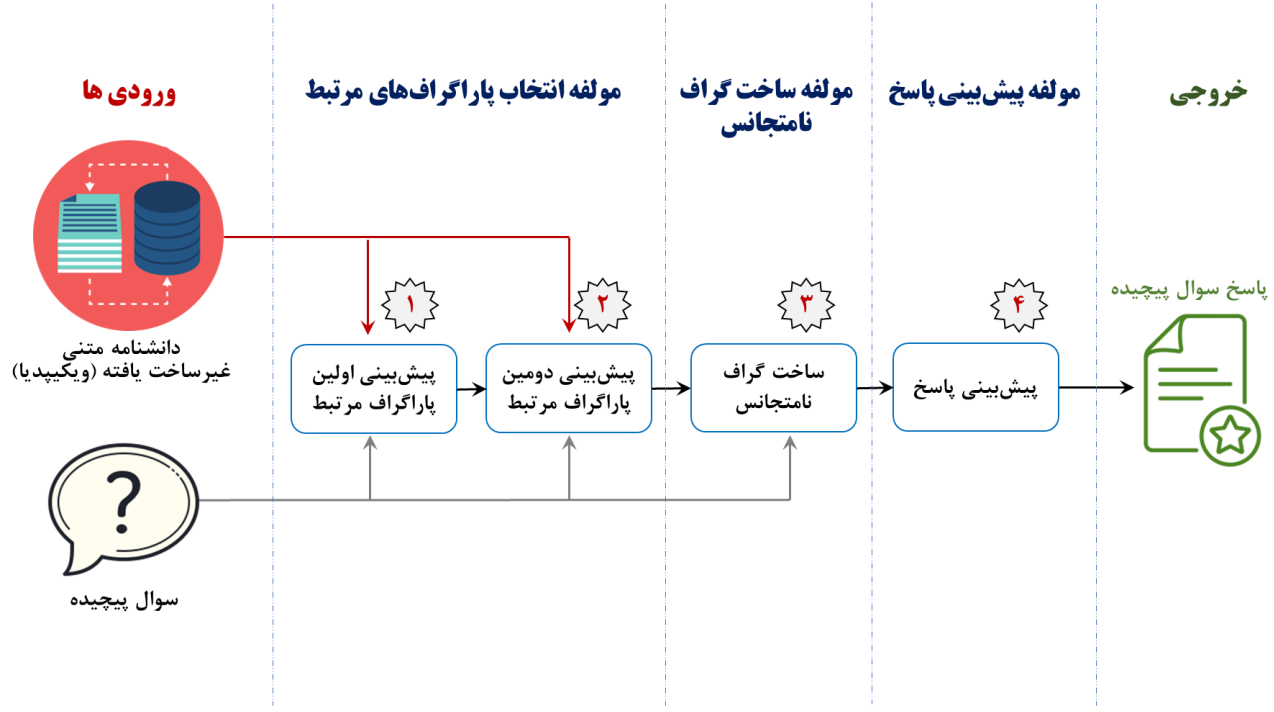
اولین گام در یک سامانه پرسش و پاسخ، بازیابی محتوای کامل و مرتبط با پرسش است که قابل استدلال و نتیجه‌گیری برای رسیدن به پاسخ صحیح باشد. انتخاب اسناد یا پاراگراف‌های مرتبط می‌تواند اطلاعات پرت و غیرلازم را به حداقل رسانده و کارآیی و بهره‌وری سامانه‌های پرسش‌وپاسخ را ارتقا ببخشد.

باتوجه به اینکه در این پژوهش روی مسئله پاسخدهی به پرسش‌های چندگامی تمرکز شده است، سامانه‌ای نیاز هست که بتواند اطلاعات بین سندی را دخالت داده و با حداقل منابع محاسباتی، به نتیجه خوب برسد. بدین منظور در این بخش یک روش دو مرحله‌ای مبتنی‌بر مبدل‌ها برای رسیدن به پاسخ ارائه شده است که ضمن درنظر گرفتن ارتباطات بین‌سندی بر محدودیت‌های مبدل نیز غلبه نماید.

لازم به‌ذکر است که این مدل به‌صورت رمزگذار متقابل[۳۶] آموزش دیده است.

۳-۲-۱- انتخاب اولین مرتبط‌ترین پاراگراف

در این مرحله هدف پیدا کردن پاراگرافی است که بیشترین ربط را به پرسش داشته باشد. از این رو یک شبکه عصبی عمیق مبتنی بر مبدل طراحی شده است تا به‌ازای هر پاراگراف کاندید و پرسش مربوطه که به مدل داده می‌شود، میزان مرتبط بودن پرسش و محتوا را بدست آورد. ورودی مدل مطابق رابطه ۱ خواهد بود:



شکل ۲: شمای کلی معماری روش پیشنهادی

$$I_1 = [CLS]\ \text{question}\ [SEP]\ \text{title}\ [CONT]\ \text{paragraph}\ [SEP] \quad (۱)$$

که در آن [CLS] بردار خلاصه کل محتوای توالی و [SEP] جدا کننده زیرتوالی‌های مختلف، از توکن‌های مخصوص مبدل هستند و همچنین [CONT]توکن اضافه‌ای برای جدا کردن عنوان title و paragraph از یکدیگر است. ورودی I_1 درحالی به کدگذار دبرتا پایه[۳۷] [۲۴] با یک شبکه تماماً متصل داده می‌شود که ۸ لایه پایه از ۱۲ لایه پشته کدگذار منجمد شده و ۴ لایه باقی‌مانده تنظیم می‌شوند. در نهایت با استفاده از تابع ضرر کراس انتروپی[۳۸] دودویی به‌صورت رابطه ۲، شبکه آموزش داده می‌شود.

$$L_1 = -\sum_{i=1}^{N} \Sigma_{j=1}^{M} \left(t_{ij} \log P(d_{ij} \mid q_i)\right) + (1 - t_{ij}) \log \left(۱ - P(d_{ij} \mid q_i)\right) \quad (۲)$$

در رابطه ۲، q_i پرسش iاُم در داده‌هایN نمونه‌ای و d_{ij} سند jاُم در مجموعه اسناد کاندید M تایی است. همچنین t_{ij} برچسب مرتبط‌بودن پرسش با سند است.

۳-۲-۲- انتخاب دومین مرتبط‌ترین پاراگراف

بعد از آن‌که مرتبط‌ترین پاراگراف انتخاب شد، با داشتن پرسش و سند بازیابی‌شده اول، نسبت به انتخاب دومین مرتبط‌ترین پاراگراف اقدام می‌گردد. برای انجام این کار، همانند مرحله اول از یک شبکه عصبی عمیق مبتنی‌بر مبدل دبرتا پایه استفاده می‌گردد تا پاراگراف طلایی بعدی را نیز بدست آورده شود. توالی ورودی به کدگذار مبدل مطابق رابطه ۳ است:

$$I_2 = [CLS]\ \text{question}\ [SEP]\ \text{title}\ [CONT] P_1 [SEP]\ \text{title}\ [CONT]\ \text{paragraph}\ [SEP] \quad (۳)$$

بعد از به‌دست آوردن جانمایی‌های معنایی به‌وسیله کدگذار مبدل، از یک شبکه عصبی روبه‌جلو برای محاسبه مرتبط بودن توالی‌ها به هم استفاده می‌گردد، و بالاترین امتیاز به عنوان پاراگراف دوم یعنی P_2 بازیابی می‌شود.

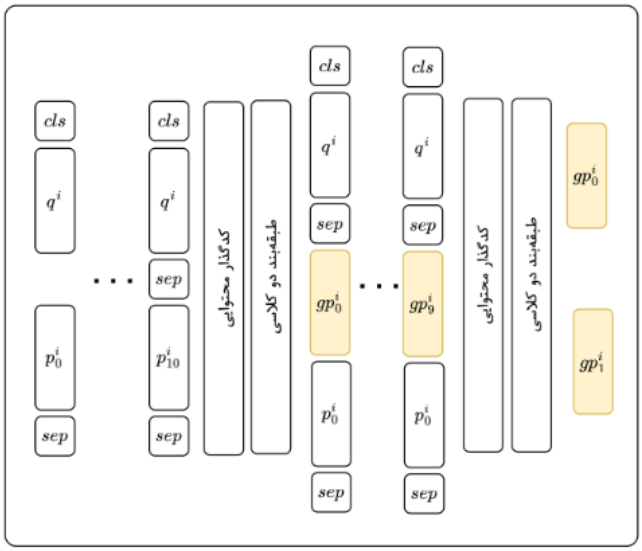
با استفاده از تابع ضرر کراس انتروپی دودویی مطابق رابطه ۴، شبکه آموزش داده می‌شود.

$$L_1 = -\Sigma_{i=1}^{N} \Sigma_{j=1}^{M} \left(t_{ij} \log P\left(d_{ij} \mid q_i, p_{i1}\right)\right) + \left(1 - t_{ij}\right) \log \left(۱ - P\left(d_{ij} \mid q_i, p_{i1}\right)\right) \quad (۴)$$

در رابطه ۴ i_1 پاراگراف به‌دست‌آمده از سند i ام در مرحله قبل می‌باشد. روش دومرحله‌ای ارائه شده دارای مزیت‌هایی به شرح زیر می‌باشد:

۱. سازگاری با تعداد محدودیت توکنی مبدل
۲. دخالت ارتباطات بین سندی
۳. قابلیت تعمیم به بازیابی پاراگراف‌های بیشتر
۴. هزینه محاسباتی به‌صرفه

در شکل ۳ معماری مدل عصبی مولفه بازیابی اسناد نمایش داده شده است. هر پرسش به‌همراه ده سند کاندید، از کدگذارهای محتوایی و یک طبقه‌بند دو کلاسه عبور داده می‌شوند تا بالاترین سند مرتبط بازیابی شود.



شکل۳: معماری خلاصه‌شده مدل عصبی مولفه بازیابی پاراگراف‌های مرتبط

در ادامه این پاراگراف طلایی با دیگر پاراگراف‌های باقی‌مانده و پرسش از کدگذارهای محتوایی و یک طبقه‌بند دو کلاسه عبور می‌کنند. در نهایت این امر منجر به بازیابی دومین مرتبط‌ترین پاراگراف می‌گردد.

۳-۳- مولفه ساخت گراف نامتجانس

برای پاسخدهی دقیق به پرسش نیاز به یک شمای ساختاریافته‌ای هست تا به استدلال ماشینی کمک کند. از سویی دیگر متن‌ها، داده‌های بدون ساختار هستند و شامل پرسش‌ها و محتوای اسناد (پاراگراف‌ها) که پاسخ پرسش‌ها را نیز دربر می‌گیرد هستند. با ترکیب نوع داده ساختار یافته و بدون ساختار انتظار می‌رود، به استدلال چندگامی برای رسیدن به پاسخ کمک شود.

تبدیل متن به یک گراف، هزینه محاسباتی برای فرآیند استدلال خواهد داشت، بدین‌منظور برای استفاده از مزیت ساختار گرافی، گراف نامتجانس سبک‌وزنی طراحی شده است تا ضمن کاهش هزینه محاسباتی تا جایی که به کارآیی آسیب نرسد، استدلال را برای پاسخدهی به پرسش انجام دهد. در ادامه انواع گره‌های گراف

نامتجانس و ارتباطات میان آن‌ها معرفی می‌گردد.

۳-۳-۱- انواع گره‌های گراف نامتجانس

در این گراف دو نوع گره وجود دارد که در ادامه تشریح می‌گردد.

۱) موجودیت یا عبارت اسمی

یکی از انواع اشیایی که درون گراف قرار می‌گیرند، موجودیت‌ها یا عبارات اسمی هستند، که از درون محتوا با استفاده از یک مدل تشخیص موجودیت و مدل استخراج عبارات اسمی، بازیابی می‌شوند و به‌عنوان گره درون گراف قرار می‌گیرند.

۲) جملات و پرسش

نوع دیگری از گره‌هایی که به گراف اضافه می‌شوند، جملات و پرسش‌ها می‌باشند. علاوه بر پرسش اصلی، جملات پاراگراف طلایی نیز به عنوان گره در گراف قرار داده می‌شوند.

۳-۳-۲- انواع یال یا ارتباط میان گره‌ها

انواع مختلفی از ارتباطات میان گره‌ها وجود دارد. با توجه به وجود دو نوع گره در گراف، سه نوع ارتباط میان این دو نوع گره شامل ارتباط موجودیت-موجودیت، موجودیت-جمله و جمله-جمله در قالب چهار عنوان یال در گراف مطرح می‌شود که در ادامه به صورت کامل تشریح می‌گردد.

۱) ارتباطات وجودی موجودیت در جمله

اگر یک موجودیت درون یک جمله ظاهر شود، بین گره جمله و گره موجودیت یالی افزوده می‌شود.

۲) ارتباطات هم‌رخدادی موجودیت

اگر دو موجودیت درون یک پاراگراف پدیدار شوند، بین دو گره یالی افزوده می‌شود.

۳) ارتباطات جملات با موجودیت مشترک

اگر دو جمله دارای یک موجودیت مشترک باشند بین آن دو یالی افزوده می‌شود.

۴) ارتباطات هم‌رخدادی جملات

اگر دو جمله درون یک پاراگراف باشند، بین دو گره یالی افزوده می‌شود.

برای ساخت چنین گرافی از برت تنظیم دقیق‌شده برای وظیفه تشخیص موجودیت‌های نام‌دار با داده‌های کو ان ال ال ۲۰۰۳ [۳۹] استفاده شده است.

۳-۴- مولفه پیش‌بینی پاسخ

برای رسیدن به پاسخ پرسش، نیازمند استدلال روی پاراگراف‌های بازیابی‌شده است. از این‌رو در این بخش رویکردی تلفیقی از شبکه عصبی گرافی و مبدل مورد استفاده قرار گرفته است.

بعد از ساخت گراف نامتجانس که در بخش قبل توضیح داده شد، سه‌گانه‌های گراف با استفاده از کدگذاری که روی همین وظیفه تنظیم دقیق شده‌اند، مقداردهی می‌شوند.

سپس با استفاده از سازوکار انتقال پیام گراف و شبکه‌های عصبی گرافی، روی گراف نامتجانس جانمایی‌های ساختاری ایجاد می‌شوند.

بعد از استدلال ماشینی روی گراف، جانمایی ساختاری بدست‌آمده از هر جمله با جانمایی‌های اولیه جملات که از کدگذارهای مبدل گذرانده شده‌اند، تجمیع می‌شود. درنهایت بهترین گستره برای پاسخ انتخاب می‌شود.

در ادامه با استفاده از روابط به درک بهتر این بخش پرداخته خواهد شد:

$$S = [CLS]\ \text{question}\ [SEP] T_1 [SEP] P_1 [SEP] T_2 [SEP] P_2 [SEP] \quad (۵)$$

ورودی مدل به‌صورت رابطه ۵ می‌باشد که در آن P_1 و P_2 پاراگراف‌های بازیابی شده هستند که جملات هرکدام تعریف و تفکیک شده است.

T_1 و T_2 هم عنوان پاراگراف‌های بازیابی شده هستند. از مبدل دبرتا و به‌دنبال آن یک شبکه عصبی روبه‌جلو مثل روشی که در مقاله برت معرفی شده است، برای پیش‌بینی محدوده ابتدا و انتهای گستره پاسخ استفاده می‌گردد.

$$L_{\text{start}} = -\Sigma_{i=1}^{N} (t_{si} \log P(IS_i)) + (1 - t_{si}) \log(۱ - P(IS_i))) \quad (۶)$$

$$L_{\text{end}} = -\Sigma_{i=1}^{N} (t_{ei} \log P(IS_i)) + (1 - t_{ei}) \log(۱ - P(IS_i))) \quad (۷)$$

$$L_{\text{span}} = L_{\text{start}} + L_{end} \quad (۸)$$

رابطه ۶ و ۷ به‌ترتیب تابع ضرر کراس انتروپی محدوده شروع و انتهای گستره پاسخ را نشان می‌دهد و ضرر کلی رابطه ۸، جمع این دو است. i ، IS_i امین توکن ورودی می‌باشد. t_{ei} و t_{si} به‌ترتیب برچسب شروع و پایان گستره در i امین توکن هستند. بعد از تنظیم دقیق مسئله روی مبدل و شبکه روبه‌جلو مدنظر، از کدگذار به‌دست‌آمده (F_{en}) برای مقداردهی اولیه گراف و تنظیم دقیق دوباره مدل با افزودن شبکه‌های عصبی گرافی استفاده می‌گردد.

در بخش‌های قبل گراف نامتجانس معرفی گردید و در ادامه نحوه استفاده آن را در روند پاسخگویی به پرسش شرح داده می‌شود.

بعد از ساخت گراف نامتجانس، وقت آن است که با استفاده از شبکه‌های عصبی گرافی، استدلال گرافی انجام گردد.

با استفاده از گراف نامتجانس می‌توان برای هر نوع رابطه‌ای، شبکه عصبی گرافی تعریف کرده و استدلال را انجام داد.

روی سه‌نوع رابطه میان گره‌های گراف نامتجانس سه شبکه گرافی عصبی [۲۵] به‌صورت زیر تعریف شده است:

$$X_{ss} = GCNConv_{\text{sent_sent}} \quad (۹)$$

در رابطه ۹ با استفاده از شبکه پیچشی گرافی[۴۰] گره‌های جملات که باهم رابطه دارند و با کدگذار (F_{en}) مقداردهی شده‌اند، بروزرسانی می‌شوند. جانمایی اولیه جملات در گراف، بردار [CLS] بعد از کدگذاری جملات است.

$$X_{es} = SAGEConv_{\text{entity_sent}} \quad (۱۰)$$

به‌همین ترتیب در رابطه ۱۰ با استفاده از شبکه گراف‌سیج[۴۱] [۲۶] گره‌های جملات و موجودیت‌هایی که باهم رابطه دارند و با کدگذار(F_{en}) مقداردهی شده‌اند، بروزرسانی می‌شوند.

$$X_{ee} = GATConv_{\text{entity_entity}} \quad (۱۱)$$

در ادامه و در رابطه ۱۱ با استفاده از گات کانو[۴۲] [۲۶] گره‌های موجودیتی که باهم ارتباط دارند و با کدگذار (F_{en}) مقداردهی شده‌اند، بروزرسانی می‌شوند.

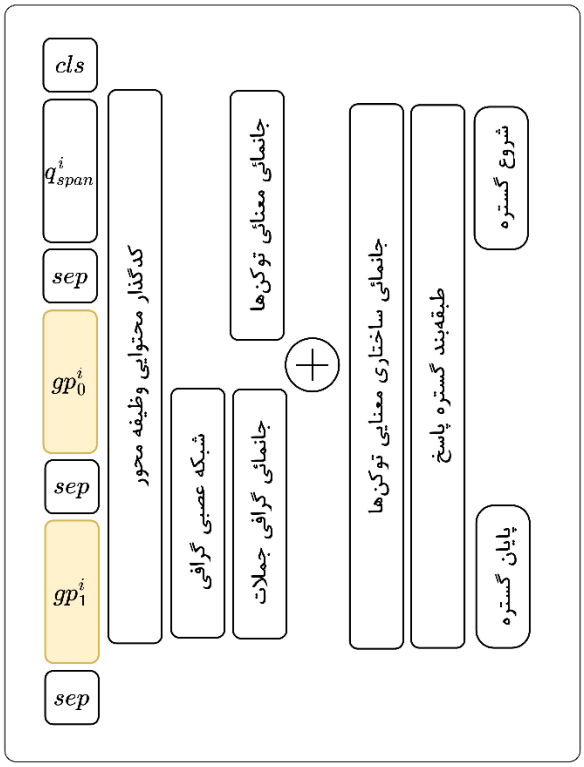
$$X_{\text{sent}} = SUM(X_{ss}, X_{se}, X_{ee})_{sent} \quad (۱۲)$$

بعد از ساخت شبکه گراف نامتجانس با ترکیب سه زیرشبکه گرافی مذکور، جانمایی‌های X_{sent} جملات را برای مرحله بعدی استدلال استخراج می‌گردد.

درنهایت، جانمایی جملات استخراج شده توسط شبکه گراف نامتجانس با جانمایی‌های مبدل تک تک توکن‌های جملات مطابق رابطه ۱۲ پیوند داده می‌شود تا ترکیبی از بازنمایی ساختاری و محتوایی حاصل گردد.

برای پیش‌بینی گستره پاسخ کافی است انتها و ابتدای گستره با تابع ضرر رابطه ۸ پیش‌بینی گردد.

شکل ۴ معماری مدل عصبی مولفه پیش‌بینی پاسخ را نشان می‌دهد. در این معماری بعد از دریافت پرسش برای به‌دست آوردن جانمایی محتوایی، پرسش و دو سند طلایی به یک کدگذار محتوایی داده می‌شوند. همچنین برای به‌دست آوردن جانمایی ساختاری از یک کدگذار وظیفه محور برای مقداردهی اولیه گراف نامتجانس استفاده می‌شود. در ادامه جانمایی‌های ساختاری و محتوایی باهم تجمیع شده و با استفاده از یک طبقه‌بند گستره پاسخ، ابتدا و انتهای پاسخ پیش‌بینی می‌شود.



شکل۴: معماری خلاصه‌شده مدل مولفه پیش‌بینی پاسخ

۴- ارزیابی روش پیشنهادی

بعد از معرفی روش پیشنهادی در این بخش آزمایش‌های انجام‌شده در طی پژوهش گزارش خواهند شد.

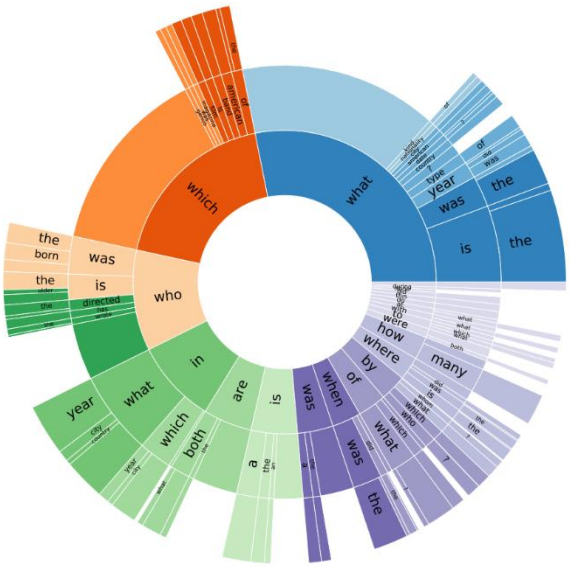
همان‌طور که در بخش‌های قبل بیان گردید، در این پژوهش روی پرسش و پاسخ مبتنی بر درک مطلب ماشینی چندگامی تمرکز شده است و برای دستیابی به این مهم، روشی مبتنی بر گراف نامتجانس طراحی شده است. در ادامه با گزارش نتایج آزمایش‌ها و مقایسه آن با کارهای مرتبط، راجع به میزان کارآمدی روش پیشنهادی در پاسخگویی به پرسش‌های پیچیده چندگامی بحث می‌گردد. ارزیابی روش پیشنهادی، نیازمند داده‌هایی بود که علاوه‌بر داشتن پرسش‌های متنوع و دامنه‌باز، شامل پرسش‌های چندگامی باشد. از این‌رو داده‌های هات‌پات‌کیوای با پیکربندی دیسترکتور[۴۳] یا منحرف‌کننده انتخاب گردید.

در ادامه ابتدا ضمن معرفی مجموعه‌داده آزمون، معیارهای ارزیابی معرفی می‌گردد. سپس میزان کارآمدی روش پیشنهادی و برخی مولفه‌های آن برحسب معیارهای معرفی‌شده گزارش شده و با پژوهش‌های مشابه مقایسه خواهد شد.

۴-۱- مجموعه داده آزمون

مجموعه‌داده آزمون برای محک‌زنی روش پیشنهادی هات‌پات‌کیوای است و متشکل از بیش ۱۱۲ هزار پرسش و پاسخ

پاسخ است. همان طور در بخش های قبل نیز اشاره شد این مجموعه داده انواع مختلفی از پرسش‌های مقایسه‌ای و پل را شامل می‌شود. توزیع انواع پرسش‌ها در این مجموعه داده مطابق شکل ۵ است.



شکل۵: توزیع انواع پرسش‌ها در مجموعه‌داده هات‌پات‌کیوای[۱۱]

از آنجا که مجموعه‌داده هات‌پات‌کیوای از روی دانشنامه ویکیپدیا استخراج شده است، برای پرسش و پاسخ دامنه باز مناسب است و طیف گستردهای از موضوعات را پوشش می‌دهد.

این مجموعه‌داده در دو حالت فول‌ویکی و دیسترکتور آماده‌سازی شده است. در حالت فول‌ویکی پرسش‌ها و عناوین اسناد مرتبط به همراه پاسخ نهایی پرسش قرار داده شده و روش پیشنهادی باید از طریق بازیابی اطلاعات روی کل اسناد ویکی‌پدیا توانایی پیدا کردن پاسخ را داشته باشد. در حالت دیسترکتور برای هر پرسش ده پاراگراف شامل دو پاراگراف طلایی در نظر گرفته شده است که پاسخ پرسش مورد نظر با استدلال روی تعدادی از این پاراگراف‌ها قابل استنتاج است. دسته‌بندی مجموعه‌داده هات‌پات‌کیوای مورد استفاده در روش پیشنهادی به صورت جدول ۲ است.

جدول ۲: دسته‌بندی انواع داده‌ها در مجموعه‌داده هات‌پات‌کیوای [۱۱]

نوع داده	نوع استفاده	آمار تعداد پرسش‌ها
داده تک گامی(آسان)	آموزش مدل	۱۸.۰۸۹
داده چندگامی گامی(متوسط)	آموزش مدل	۵۶.۸۱۴
داده چندگامی گامی(سخت)	آموزش مدل	۱۵.۶۶۱
داده آزمون در حالت دیسترکتور	آزمون	۷.۴۰۵

۴-۲- معیارهای ارزیابی

به منظور ارزیابی روش پیشنهادی از دو معیار اف‌وان[۴۴] و تطابق دقیق[۴۵] استفاده شده است. دلیل انتخاب این معیارها، معرفی آن‌ها توسط مجموعه‌داده آزمون برای ارزیابی روش‌ها و مدل‌های استفاده کننده از مجموعه‌داده مذکور است.

۴-۳- جزییات پیاده‌سازی

پیاده‌سازی روش پیشنهادی مبنی بر کتابخانه ترنسفورمرز[۴۶] و پای‌ترچ[۴۷] است. همچنین از پای‌جی[۴۸] در بخش گراف استفاده شده است. هیچ محدودیتی روی تعداد موجودیت و جملات در یک نمونه در بخش ساخت گراف اعمال نشده است.

۴-۳-۱- توصیف مدل

برای بخش انتخاب پاراگراف دومرحله‌ای از مبدل دبرتا با تعداد لایه پنهان ۱۲ استفاده شده است که ۸ لایه آن منجمد شده است. در مولفه‌های دیگر از این مبدل با ۲۴ لایه پنهان شامل ۱۸ لایه منجمد برای آموزش مدل استفاده شده است.

۴-۳-۲- جزییات پارامترها

مدل‌های عصبی در ۵ ایپاک[۴۹] مورد آموزش قرار گرفته‌اند. تعداد دسته[۵۰] در همه مولفه‌ها برابر ۸ و مقدار نرخ یادگیری[۵۱] 1e-5 است. نرخ گرم کردن[۵۲] ۰.۱ و کاهش نرخ یادگیری[۵۳] در آزمون های انجام شده برابر ۰.۰۱ است. همچنین از بهینه‌ساز آدام دبلیو[۵۴] با بکارگیری زمان‌بند استفاده شده است. برای آزمایش از ۲ پردازنده گرافیکی انویدیا ای۱۰۰[۵۵] در مجموع با ۶۲ گیگابایت حافظه موقت استفاده شده است و زمان آموزش حدود ۱۲۰ ساعت بوده است.

۴-۴- پیچیدگی زمانی

پیچیدگی زمانی در سامانه‌های پرسش و پاسخ به مولفه‌های سازنده سامانه و نوع وظایف انجام شده توسط هر مولفه و همچنین جزییات پیاده‌سازی آنها مربوط می‌شود. پیچیدگی زمانی مولفه انتخاب پاراگراف‌های مرتبط با توجه به جزییات پیاده سازی که در بخش قبل به آن پرداخته شد$(O(12 * \mathrm{n} * \mathrm{d} * h^2 + 8 * m)$ است که n تعداد گره‌ها و یا حجم داده‌های آموزشی، h تعداد سری توجه، d ابعاد نمایش در هر گره و m تعداد پارامترهای منجمد است که تاثیری در پیچیدگی زمانی مدل ندارند.

پیچیدگی زمانی مولفه ساخت گراف نامتجانس با توجه به بکارگیری و ترکیب سه نوع تکنیک جانمایی گرافی گراف‌سیج، گات‌کانو و جی‌سی‌ان و همچنین جزییات پیادسازی برابر با

$O(\mathrm{n} * k * 24)$ است و در آن n برابر با تعداد گره های گراف و k تعداد همسایگان است. پیچیدگی زمانی مولفه پیشبینی پاسخ که در آن از مبدل دبرتا و به‌دنبال آن یک شبکه عصبی روبه‌جلو استفاده شده است دارای پیچیدگی زمانی با پارامترهای مشابه مولفه انتخاب پاراگراف مرتبط برابر $O(24 * \mathrm{n} * \mathrm{d} * h^2 + 18 * m)$است. به عبارت دیگر، پیچیدگی زمانی کل سامانه پرسش و پاسخ با سه مولفه مذکور، برابر $O(\mathrm{n} * \mathrm{d} * h^2) + O(\mathrm{n} * k * 24) + O(m)$است.

۴-۵- نتایج آزمایش‌ها

در ادامه و در این بخش سعی شده است نتایج بدست آمده برای دو مولفه انتخاب پاراگراف‌های مرتبط و پیش‌بینی پاسخ با نتایج گزارش شده توسط بهترین پژوهش‌های مرتبط مقایسه گردد.

۴-۵-۱- مولفه انتخاب پاراگراف‌های مرتبط

اولین مولفه ای که در مقایسه با پژوهش‌های مرتبط مورد ارزیابی قرار گرفته است انتخاب اسناد مرتبط است و نتایج آن مطابق جدول ۳ است.

جدول ۳ نتایج مقایسه مولفه انتخاب پاراگراف‌های مرتبط با پژوهش‌های مشابه

روش پژوهشی	معیار مقایسه	
	تطابق دقیق	اف‌وان
اچ جی ان	-	۹۴.۵۳
اس‌ای‌ایی	۹۱.۷۸	۹۵.۷۶
دی اف جی ان	۹۱.۷۸	۹۵.۷۶
جی اف ام اچ آر سی	۹۵.۷۷	۹۷.۸۲
روش پیشنهادی-پیکره‌بندی اول	۹۱	۹۷.۶
روش پیشنهادی-پیکره‌بندی دوم	**۹۵.۹**	**۹۸.۹**

در پیکربندی تک‌مرحله‌ای مولفه انتخاب پاراگراف‌های مرتبط در روش پیشنهادی، میزان تطبیق پاراگراف و پرسش را به‌دست آورده و دو تا از بهترین تطبیق‌ها را به‌عنوان پاراگراف‌های طلایی خروجی ارائه می‌دهد. نتایج بدست آمده در آزمایش‌ها برای گام اول بازیابی پاراگراف برای معیارهای اف‌وان و تطابق دقیق به ترتیب ۹۷.۶ و ۹۱ است.

در پیکربندی دو مرحله‌ای مولفه انتخاب پاراگراف‌های مرتبط یک مدل اضافه نیز برای جبران اشتباهات احتمالی مدل اول طراحی شده است. در این مدل میزان ارتباط دو پاراگراف با پرسش چندگامی به‌صورت همزمان سنجیده می‌شود. در مرحله اول پاراگرافی که احتمال کمتری نسبت به دیگری دارد، کنار گذاشته شده و پاراگراف مرتبط‌تر با پاراگراف‌های دیگر نمونه به مدل داده می‌شود و بالاترین شباهت به‌عنوان پاراگراف دوم نهایی انتخاب می‌شود.

این مدل دومرحله‌ای کارکرد خوبی داشته است و نتایج بدست آمده در آزمایش‌ها برای گام دوم بازیابی سند برای معیارهای اف‌وان و تطابق دقیق به ترتیب ۹۸.۹ و ۹۵.۹ است. نتایج بدست آمده برای این مولفه نشانگر بهبود ۱.۳ و ۴.۱۲ درصدی به ترتیب در معیار اف‌وان و تطبیق دقیق نسبت به سایر بهترین کارهای مشابه موجود است.

۴-۵-۲- مولفه پیش‌بینی پاسخ

پس از شناسایی پاراگراف‌های طلایی و بازنمایی آنها در قالب زیرساخت گرافی، نوبت پیش‌بینی پاسخ است. برای پاسخدهی به پرسش‌ها یک کدگذار مرتبط با وظیفه با مبدل دبرتا طراحی شده است و سپس با ترکیب گراف نامتجانس و مبدل، گستره پاسخ شناسایی می‌گردد.

جدول ۴: نتایج مقایسه مولفه پیش‌بینی پاسخ با پژوهش‌های مشابه

روش پژوهشی	معیار مقایسه	
	تطابق دقیق	اف‌وان
بیس‌لاین	۴۴.۴۴	۵۸.۲۸
دی‌کامپ‌آرسی	۵۵.۲۰	۶۹.۶۳
اس‌ای‌ایی	۶۱.۳۲	۷۴.۸۱
دی اف جی ان	۵۵.۶۶	۶۹.۳۴
اس‌آرال‌جی‌آران	۶۲.۶۲	۷۶.۱۴
کیواف‌ای	۵۳.۷۰	۶۸.۷۰
ال‌کیوآرنت[۵۶]	۶۰.۲۰	۷۳.۷۸
گات	۶۶.۷	۸۰.۷
جی‌اف‌ام‌اچ‌آرسی	۷۰.۷۲	۸۳.۵۳
سی‌تواف‌ام‌اف‌وان	۷۲.۰۷	۸۴.۳۴
روش پیشنهادی	**۷۸.۷۱**	**۸۶.۵۱**

جدول ۴ آزمایش‌های انجام شده روی وظیفه پیش‌بینی پاسخ را نشان می‌دهد که با مدل‌های مشابه دیگر مقایسه شده است. مدل مربوط به مولفه مذکور در روش پیشنهادی نتیجه کل معماری خط‌لوله‌ای روش پیشنهادی را نشان می‌دهد. این مدل در پیش‌بینی پاسخ به معیار اف‌وان ۸۶.۵۱ (۲.۱۷ بهبود نسبت به بهترین روش) و همچنین تطابق دقیق ۷۸.۷۱ (۶.۶۴ درصد بهبود نسبت به بهترین روش) دست یافته‌است.

۵- نتیجه‌گیری

در این پژوهش، تمرکز بر روی پاسخدهی به پرسش مبتنی بر درک مطلب ماشینی چندگامی بود. شکاف بین درک مطلب انسانی و ماشینی در سامانه‌های پرسش‌وپاسخ سنتی، مقدمه‌ای برای شروع

کار روی درک مطلب ماشینی با قابلیت استدلال پیچیده چندگامی گردید. هدف ارائه روشی بود که علاوه‌بر پشتیبانی از پرسش‌های ساده، بتواند پرسش‌هایی را که نیاز به چندین گام و چندین سند یا پاراگراف برای پاسخگویی دارند، نیز پشتیبانی نماید.

در همین راستا و جهت تحقق اهداف مذکور، معماری خط‌لوله‌ای برای سامانه پرسش و پاسخ متشکل از سه مولفه اصلی پیشنهاد و پیاده‌سازی گردید. به منظور بررسی میزان اثربخشی پژوهش انجام شده، روش پیشنهادی با کارهای مشابه مبتنی بر معیارهای افوان و تطبیق دقیق مورد مقایسه قرار گرفت و نتایج حاصل از ارزیابی حاکی از برتری روش پیشنهادی نسبت به سایر رقبای خود بود.

۵-۱-۱- کارهای آتی

ازآنجایی‌که مسیر علم نامحدود است، لازم است نگاهی به آینده داشته و کار خود را به سمت بهبود بیشتر پیش ببریم. در همین راستا و برای بهبود و ارتقا روش پیشنهادی به عنوان کارهای آتی می‌توان گام‌هایی را برداشت که در ادامه به آن اشاره می‌گردد.

امروزه مدل‌های پیشرفته مثل مبدل‌ها، بعد از آموزش منابع سخت‌افزاری زیادی برای استنتاج نیاز دارند و در بعضی مواقع کند هستند. لازم است موارد زیر نیز از لحاظ بهینه‌سازی و کارایی لحاظ گردد:

۱)کمی‌سازی مدل[۵۷]

با انجام کمی‌سازی روی مدل خروجی، به‌جای محاسبات سنگین اعداد شناور، روی اعداد صحیح عملیات انجام گردد. این امر منجر به کاهش زمان استنتاج (افزایش سرعت) با کمترین کاهش دقت می‌گردد و استفاده از منابعی سخت‌افزاری مانند دیسک و حافظه تصادفی را نیز کاهش می‌دهد.

۲)هرس کردن مدل[۵۸]

با هرس کردن مدل‌های ساخته‌شده می‌توان تا جای ممکن، بخش‌های مختلف از شبکه را حذف کرده و سرعت را بهبود ببخشید.

۳)تقطیر دانش[۵۹]

با تقطیر دانش می‌توان دانش را از یک مدل بزرگ به مدل کوچک‌تر انتقال داد.

مراجع

[1] Rajpurkar P, Zhang J, Lopyrev K, Liang P. SQuAD: 100,000+ Questions for Machine Comprehension of Text. InProceedings of the 2016 Conference on Empirical Methods in Natural Language Processing 2016 Nov (pp. 2383-2392).

[2] Joshi M, Choi E, Weld DS, Zettlemoyer L. TriviaQA: A Large Scale Distantly Supervised Challenge Dataset for Reading Comprehension. InProceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) 2017 Jul (pp. 1601-1611).

[3] Dunn M, Sagun L, Higgins M, Guney VU, Cirik V, Cho K. Searchqa: A new q&a dataset augmented with context from a search engine. arXiv preprint arXiv:1704.05179. 2017 Apr 18.

[4] Welbl J, Stenetorp P, Riedel S. Constructing Datasets for Multi-hop Reading Comprehension Across Documents. Transactions of the Association for Computational Linguistics.2018 ;6:287-302.

[5] Talmor A, Berant J. Repartitioning of the complexwebquestions dataset. arXiv preprint arXiv:1807.09623. 2018 Jul 25.

[6] Talmor A, Herzig J, Lourie N, Berant J. CommonsenseQA: A Question Answering Challenge Targeting Commonsense Knowledge. InProceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) 2019 Jun (pp. 4149-4158).

[7] Dua D, Wang Y, Dasigi P, Stanovsky G, Singh S, Gardner M. DROP: A Reading Comprehension Benchmark Requiring Discrete Reasoning Over Paragraphs. InProceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) 2019 Jun (pp. 2368-2378).

[8] Khashabi D, Chaturvedi S, Roth M, Upadhyay S, Roth D. Looking beyond the surface: A challenge set for reading comprehension over multiple sentences. InProceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers) 2018 Jun (pp. 252-262).

[9] Dalvi B, Huang L, Tandon N, Yih WT, Clark P. Tracking State Changes in Procedural Text: a Challenge Dataset and Models for Process Paragraph Comprehension. InProceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers) 2018 Jun (pp. 1595-1604).

[10] Weston J, Bordes A, Chopra S, Rush AM, Van Merriënboer B, Joulin A, Mikolov T. Towards AI-complete question answering: A set of prerequisite toy tasks. In4th International Conference on Learning Representations, ICLR 2016 2016.

[11] Yang Z, Qi P, Zhang S, Bengio Y, Cohen W, Salakhutdinov R, Manning CD. HotpotQA: A Dataset for Diverse, Explainable Multi-hop Question Answering. InProceedings of the 2018 Conference on Empirical Methods in Natural Language Processing 2018 (pp. 2369-2380).

[12] Chen W, Zha H, Chen Z, Xiong W, Wang H, Wang WY. HybridQA: A Dataset of Multi-Hop Question Answering over Tabular and Textual Data. InFindings of the Association for Computational Linguistics: EMNLP 2020 2020 Nov (pp. 1026-1036).

[13] Mihaylov T, Clark P, Khot T, Sabharwal A. Can a Suit of Armor Conduct Electricity? A New Dataset for Open Book Question Answering. InProceedings of the 2018 Conference on Empirical Methods in Natural Language Processing 2018 (pp. 2381-2391).

[14] Clark C, Gardner M. Simple and Effective Multi-Paragraph Reading Comprehension. InProceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) 2018 Jul (pp. 845-855).

[15] Qiu L, Xiao Y, Qu Y, Zhou H, Li L, Zhang W, Yu Y. Dynamically fused graph network for multi-hop reasoning. InProceedings of the 57th Annual Meeting of the Association for Computational Linguistics 2019 Jul (pp. 6140-6150).

[16] Kenton JD, Toutanova LK. Bert: Pre-training of deep bidirectional transformers for language understanding. InProceedings of naacL-HLT 2019 Jun 2 (Vol. 1, p. 2).

[17] Tu M, Huang K, Wang G, Huang J, He X, Zhou B. Select, answer and explain: Interpretable multi-hop reading comprehension over multiple documents. InProceedings of the AAAI conference on artificial intelligence 2020 Apr 3 (Vol. 34, No. 05, pp. 9073-9080).

[18] Zheng C, Kordjamshidi P. SRLGRN: Semantic Role Labeling Graph Reasoning Network. InProceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP) 2020 Nov (pp. 8881-8891).

[19] Nishida K, Nishida K, Nagata M, Otsuka A, Saito I, Asano H, Tomita J. Answering while Summarizing: Multi-task Learning for Multi-hop QA with Evidence Extraction. InProceedings of the 57th Annual Meeting of the Association for Computational Linguistics 2019 Jul (pp. 2335-2345).

[20] Min S, Zhong V, Zettlemoyer L, Hajishirzi H. Multi-hop Reading Comprehension through Question Decomposition and Rescoring. InProceedings of the 57th Annual Meeting of the Association for Computational Linguistics 2019 Jul (pp. 6097-6109).

[21] Wu, B., Zhang, Z., & Zhao, H. (2021). Graph-free multi-hop reading comprehension: A select-to-guide strategy. arXiv preprint arXiv:2107.11823.

[22] He, Y., Gorinski, P. J., Staliunaite, I., & Stenetorp, P. (2023). Graph Attention with Hierarchies for Multi-hop Question Answering. arXiv preprint arXiv:2301.11792..

[23] Zhangyue, Y., Yuxin, W., Xiannian, H., Yiguang, W., Hang, Y., Xinyu, Z., ... & Xipeng, Q. (2023, August). Rethinking Label Smoothing on Multi-hop Question Answering. In Proceedings of the 22nd Chinese National Conference on Computational Linguistics (pp. 611-623).

[24] He P, Liu X, Gao J, Chen W. DEBERTA: DECODING-ENHANCED BERT WITH DISENTANGLED ATTENTION. In 9th International Conference on Learning Representations 2021 May.

[25] Kipf TN, Welling M. Semi-Supervised Classification with Graph Convolutional Networks. 5th International Conference on Learning Representations, ICLR 2017 April.

[26] Hamilton W, Ying Z, Leskovec J. Inductive representation learning on large graphs. Advances in neural information processing systems. 2017;30.

پاورقی‌ها:

[1] Machine Reading Comprehension (MRC)
[2] SQuAD
[3] TrivaQA
[4] SearchQA
[5] QAngaroo
[6] ComplexWebQuestions
[7] WikiHop
[8] DROP
[9] MultiRC
[10] Entity Tracking
[11] HotpotQA
[12] Text Based Multi-Hop Question Answering
[13] HybridQA
[14] OpenBookQA
[15] Baseline
[16] Self-Attention
[17] Bi-Attentions
[18] Recurrent Neural Network (RNN)
[19] DFGN
[20] BERT
[21] Transformers
[22] SAE
[23] SRLGRN
[24] QFE
[25] Recurrent Neural Networks (RNNs)
[26] DecompRC
[27] GFMHRC
[28] Select-to-Guide (S2G)
[29] Coarse-to-Fine
[30] GATH
[31] C2FM-F1
[32] Self-Attention
[33] Bi-Attentions
[34] Relation Extraction
[35] Coreference Resolution
[36] Cross-Encoder
[37] DeBERTa $_{base}$
[38] Cross-Entropy
[39] CoNLL 2003
[40] Graph Convolution Network (GCN)
[41] GraphSAGE
[42] GATConv
[43] Distractor
[44] F1 Measure
[45] Exact Match
[46] Transformers
[47] PyTorch
[48] PyG
[49] Epoch
[50] Batch
[51] Learning Rate
[52] Warm-up Rate
[53] Learning-wise Learning (L2) Rate Deacy
[54] AdamW
[55] NVIDIA A100
[56] LQR-Net
[57] Model Quantization
[58] Model Pruning
[59] Knowledge Distillation