Journal of Soft Computing and Information Technology (JSCIT)
Babol Noshirvani University of Technology, Babol, Iran
Journal Homepage: jscit.nit.ac.ir
Volume 13, Number 1, Spring 2024, pp. 1-10
Received: 10/19/2023, Revised: 05/27/2024, Accepted: 08/14/2024

Improving the Effectiveness of Open-Domain Question-Answering Systems for Answering Multi-hop Questions in Persian Language

Arash Ghafouri[1], Mobina Taji[2], Hasan Naderi[3*], Behrouz Minaei Bidgoli[4] and Mohammad Reza Hasani Ahangar[5]
1- Computer Engineering Department, Iran University of Science and Technology, Tehran, Iran.
2- Computer Engineering Department, Iran University of Science and Technology, Tehran, Iran.
3*- Computer Engineering Department, Iran University of Science and Technology, Tehran, Iran.
4- Computer Engineering Department, Iran University of Science and Technology, Tehran, Iran.
5- Faculty and Research Institute of Artificial Intelligence and Cognitive Sciences, Imam Hossein University, Tehran, Iran.
[1] aghafuri@comp.iust.ac.ir, [2] mobina_taji@comp.iust.ac.ir, [3*] naderi@iust.ac.ir, [4] b_minaei@iust.ac.ir, and [5]mrhasani@ihu.ac.ir

Corresponding author's address: Hasan Naderi, Faculty of Computer Engineering, Iran University of Science and Technology, Tehran, Iran.

Abstract- Today, one of the most popular and challenging tasks in natural language processing is answering complex user questions. Question-answering systems, as a new generation of search engines, receive user questions in natural language without semantic limitations and provide precise answers. In recent years, most research in the field of question-answering systems has been focused on the English language, and not much effort has been made in languages with limited resources, such as Persian. This limitation prevents these systems from efficiently handling questions in languages like Persian.
In this regard, this article aims to enhance the efficiency of question-answering systems in the Persian language by creating a dataset for answering complex multi-turn questions. Multi-hop questions, require at least two steps of reasoning to reach an answer. This dataset, called PersianMHQA, is the first open-domain question-answering dataset in Persian and includes 7,000 multi-hop questions. It was generated using the Persian Wikipedia as a knowledge source. To evaluate and benchmark this dataset, it has been fine-tuned on the latest pre-trained language models that support the Persian language.
The best results obtained based on F1 score and exact match on this dataset are 75.92% and 71.73%, respectively. These results indicate that this dataset is a powerful starting point for improving multi-hop complex question-answering for Persian language systems.

Keywords- Open Domain Question Answering, Multi-hop Question, Dataset, Persian Languages, Persian Wikipedia Encyclopedia.

مجله علمی رایانش نرم و فناوری اطلاعات
دانشگاه صنعتی نوشیروانی بابل
صفحه مجله: jscit.nit.ac.ir
جلد ۱۳، شماره ۱، بهار ۱۴۰۳، صفحه ۱-۱۰
دریافت: ۱۴۰۲/۰۷/۲۷، بازنگری: ۱۴۰۳/۰۳/۰۷، پذیرش: ۱۴۰۳/۰۵/۲۴

بهبود کارآمدی سامانه‌های پرسش و پاسخ دامنه‌باز برای پاسخ‌دهی به پرسش‌های چندگامی در زبان فارسی

آرش غفوری[۱]، مبینا تاجی[۲]، حسن نادری*[۳]، بهروز مینایی بیدگلی[۳] و محمدرضا حسنی آهنگر[۵]
۱- دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران، ایران.
۲- دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران، ایران.
*۳- دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران، ایران.
۴- دانشکده مهندسی کامپیوتر، دانشگاه علم و صنعت ایران، تهران، ایران.
۵- دانشکده و پژوهشکده هوش مصنوعی و علوم شناختی، دانشگاه جامع امام حسین (ع)، تهران، ایران.
[1]aghafuri@comp.iust.ac.ir, [2]mobina_taji@comp.iust.ac.ir, [3]* naderi@iust.ac.ir, [4]b_minaei@iust.ac.ir, [5]mrhasani@ihu.ac.ir

* نشانی نویسنده مسئول: حسن نادری، تهران، خیابان هنگام، دانشگاه علم و صنعت ایران، دانشکده مهندسی کامپیوتر.

چکیده- امروزه یکی از محبوب‌ترین و چالش‌برانگیزترین وظایف در پردازش زبان طبیعی پاسخ‌دهی به پرسش‌های پیچیده کاربران است. سامانه‌های پرسش و پاسخ به عنوان نسل جدید موتورهای جستجو پرسش‌های کاربران را به زبان طبیعی و بدون محدودیت معنایی دریافت می‌کنند و پاسخ را به صورت دقیق بر می‌گردانند. در سال‌های اخیر اکثر پژوهش‌های انجام شده در حوزه سامانه‌های پرسش و پاسخ بر روی زبان انگلیسی متمرکز بوده و در زبان‌های با منابع محدود از جمله فارسی تلاش چندانی صورت نگرفته است. این امر سبب می‌شود این سامانه‌ها در پشتیبانی از زبان‌های با منابع محدود همچون فارسی نتوانند کارآمدی خوبی را در مواجه با یک پرسش از خود ارائه دهند. در همین راستا در این مقاله جهت افزایش کارآمدی سامانه‌های پرسش و پاسخ در زبان فارسی نسبت به تولید و توسعه یک مجموعه‌داده برای پاسخ‌دهی به پرسش های پیچیده چندگامی یا به اختصار چندگامی اقدام گردید. پرسش‌های چندگامی، نیازمند حداقل دو گام استدلال برای دستیابی به پاسخ هستند. این مجموعه‌داده یا PersianMHQA[1] به عنوان اولین مجموعه‌داده پرسش و پاسخ دامنه‌باز شامل ۷۰۰۰ پرسش چندگامی بوده و در سازوکاری مشخص با استفاده از متن دانش‌نامه ویکی‌پدیا فارسی تولید شده است. به منظور ارزیابی و محک‌زنی این مجموعه‌داده روی جدیدترین مدل های زبانی پیش آموزش دیده که از زبان فارسی پشیبانی می‌کنند تنظیم دقیق شده است. بهترین نتایج دست آمده مبتنی بر دو معیار اف‌وان و تطابق‌دقیق روی این مجموعه‌داده به ترتیب ۷۵/۹۲ و ۷۱/۷۳ است. نتایج بدست آمده نشانگر این حقیقت است که این مجموعه‌داده شروعی قدرتمند برای بهبود پرسش و پاسخ پیچیده چندگامی برای سامانه‌های پرسش و پاسخ فارسی است.

واژه‌های کلیدی: پرسش و پاسخ دامنه‌باز، پرسش چندگامی، مجموعه‌داده، زبان فارسی، دانش‌نامه ویکی‌پدیا فارسی.

[1] https://huggingface.co/datasets/Arg1990/PersianMHQA

۱- مقدمه

در عصر حاضر، با توجه به افزایش سرعت تولید اطلاعات و دسترسی کاربران به اینترنت سیستم‌هایی که بتوانند به نیازهای اطلاعاتی کاربران که در قالب اغلب پرسش‌های پیچیده[۱] مطرح می‌شوند پاسخ دقیق، کوتاه و جامع بدهند بسیار مورد توجه قرار گرفته است. سامانه‌های پرسش و پاسخ ترکیبی از فنون پردازش زبان طبیعی و بازیابی اطلاعات را برای پاسخ‌دهی به پرسش‌های پیچیده کاربران به کار می‌گیرند. سامانه‌های پرسش و پاسخ به منظور جلوگیری از اتلاف وقت کاربران و افزایش رضایت‌مندی آن‌ها پرسش‌ها را به زبان طبیعی و بدون محدودیت معنایی دریافت می‌کنند و پاسخ آن‌ها را به نزدیک‌ترین و قابل فهم‌ترین شکل ممکن برای انسان یعنی در زبان طبیعی در اختیار کاربر قرار می‌دهند. در واقع سامانه‌های پرسش و پاسخ به عنوان نسل جدید موتورهای جستجو این امکان را برای کاربر فراهم می‌کنند تا بتواند پرسش خود را به زبان طبیعی مطرح کند. اکثر پژوهش‌های انجام شده در حوزه سامانه‌های پرسش و پاسخ بر روی زبان انگلیسی بوده و در زبان‌ها با منابع محدود از جمله زبان فارسی پژوهش چندانی انجام نشده است. بنابراین یکی از دلایلی که این سامانه‌ها نمی‌توانند کارآمدی[۲] خوبی را در پاسخ‌دهی به یک پرسش در زبان‌ها با منابع محدود نسبت به زبان انگلیسی ارائه دهند، عدم پشتیبانی از زبان پرسش است. یکی از راه‌حل‌های بهبود کارآمدی این سامانه‌ها در زبان‌ها با منابع محدود تولید و توسعه مجموعه دادگان‌های معتبر است.

در سال‌های اخیر، اندک پژوهش انجام شده در زبان فارسی نیز متمرکز بر روی پرسش‌های ساده بوده است. امروزه دیگر پرسش‌های ساده نیاز اطلاعاتی کاربران را بر طرف نمی‌سازد. امروزه پرسش‌های کاربران به سمت پرسش‌های پیچیده سوق پیدا کرده است بنابراین پاسخ‌دهی به پرسش پیچیده دارای اهمیت بالایی است. تاکنون در زبان فارسی برای سامانه‌های پرسش و پاسخ، هیچ مجموعه‌داده دامنه‌باز[۳] برای پرسش‌های پیچیده مبتنی بر منابع غیرساخت‌یافته وجود ندارد. همین امر سبب شده است تا سامانه‌های پرسش و پاسخ در زبان فارسی کارآمدی کمی در پاسخ‌دهی به یک پرسش پیچیده داشته باشند. در این مقاله، ما با معرفی اولین مجموعه‌داده پرسش و پاسخ دامنه‌باز برای پرسش‌های پیچیده زبان فارسی مبتنی بر دانش‌نامه ویکی‌پدیا، گام اول را برای ارتقا کارآمدی این سامانه‌ها در زبان فارسی در مواجه با پرسش پیچیده برداشته‌ایم.

سامانه‌های پرسش و پاسخ بر مبنای دامنه پاسخگویی به دو دسته دامنه‌باز و دامنه‌بسته تقسیم می‌شوند [۱]. سامانه‌های دامنه‌باز، سامانه‌هایی هستند که حیطه کاری آن‌ها به دامنه و دانش خاصی محدود نمی‌شود. به بیان دیگر، این سامانه‌ها هیچ‌گونه محدودیتی بر روی دامنه‌ی پرسشی که به عنوان ورودی به سیستم داده می‌شود، ندارند. از طرف دیگر، سامانه‌های دامنه‌بسته می‌توانند به پرسش‌ها در یک دامنه مشخص همچون پزشکی، سیاسی، فرهنگی و غیره پاسخ دهند. پرسش‌های که سامانه‌های پرسش و پاسخ با آن‌ها مواجه هستند براساس میزان پیچیدگی در دو دسته ساده و پیچیده قرار دارند [۲]. پرسش‌های ساده یا حقیقت‌محور[۴]، پرسش‌هایی هستند که پاسخ آن‌ها براساس یک پاراگراف و یا حتی یک جمله به‌دست می‌آید. پاسخ اینگونه پرسش‌ها اغلب یک حقیقت است که به آن‌ها پرسش‌های تک‌گامی نیز گفته می‌شود. در مقابل پرسش‌های پیچیده یا چندگامی[۵]، پرسش‌هایی هستند که پاسخ‌دهی به آن‌ها نیازمند استدلال بر روی بیش از یک متن یا پاراگراف است. انواع مختلفی از پرسش‌های پیچیده وجود دارد که در این مقاله ما دو نوع پرسش پیچیده پل[۶] و مقایسه‌ای[۷] را موردنظر قرار داده‌ایم. پرسش‌های مقایسه‌ای به مقایسه چند موجودیت یا موضوع می‌پردازند که پاسخ این نوع پرسش‌ها می‌تواند گستره‌ای از متن یا بله/خیر باشد. پرسش‌های پل شامل بیش از دو پرسش ساده تودرتو هستند که استدلال در آن‌ها به صورت ترتیبی صورت می‌گیرد. در این مقاله، پرسش‌های پیچیده دوگامی در نظر گرفته شده است. بنابراین یک پرسش پل دوگامی به دو پرسش ساده شکسته می‌شود، با پاسخ‌دهی به پرسش ساده اول و در نظر گرفتن پاسخ آن به عنوان «موجودیت پل» می‌توان به پرسش ساده دوم پاسخ داد که این پاسخ نهایی می‌باشد. برای مثال در شکل ۱ در پرسش «رفیق کلارا زتکین از مهم‌ترین چهره‌های سوسیالیستی در آغاز کدام قرن بود؟»، ابتدا باید پاسخ داد «رفیق کلارا زتکین؟»، «روزا لوکزامبورگ» است و سپس با در نظر گرفتن این پاسخ به عنوان موجودیت پل به بخش دوم پرسش، پاسخ داد.

سامانه‌های پرسش و پاسخ براساس نوع داده‌ای که در منابع اطلاعاتی متنی وجود دارد در دو دسته منابع اطلاعاتی ساخت‌یافته و منابع اطلاعاتی غیرساخت‌یافته قرار دارند. جدول‌ها و گراف‌های دانش از جمله منابع اطلاعاتی ساخت‌یافته و متن موجود در صفحه‌های وب یا سندها از جمله منابع اطلاعاتی غیرساخت‌یافته هستند.


پاراگراف اول: کتابشناسی کلارا زتکین

[۱] کتابشناسی کلارا زتکین فهرستی از آثار کلارا زتکین نظریه‌پرداز مارکسیست، کمونیست، سوسیالیست و عضو هیئت رئیسه حزب سوسیال دموکرات آلمان بود. [۲] زتکین مدافع حقوق زنان و از رهبران جنبش بین‌المللی زنان بود و بر همین اساس در سال ۱۹۱۱، یکی از بنیانگذاران هشت مارس به عنوان روز جهانی زن شداو فعالیت خود را از سال ۱۳۵۸ آغاز کرد. [۳] کلارا زتکین رفیق روزا لوکزامبورگ و جزو فعالان تشکیل فراکسیون اسپارتاکیست‌ها و حزب کمونیست آلمان ۱۹۱۸ بود.

پاراگراف دوم: رزا لوکزامبورگ

[۱] روزا لوکزامبورگ انقلابی یهودی و رهبر سوسیالیست و کمونیست آلمانی لهستانی‌تبار بود. [۲] او از مهم‌ترین چهره‌های سوسیالیستی در آغاز قرن بیستم بود. [۳] او از اصلی‌ترین بنیان‌گذاران حزب سوسیال دموکرات پادشاهی لهستان و لیتوانی و اتحادیه اسپارتاکیست بود.

پرسش: رفیق کلارا زتکین از مهم‌ترین چهره‌های سوسیالیستی در آغاز کدام قرن بود؟

پاسخ: قرن بیستم

شکل ۱: مثالی از یک پرسش پل در مجموعه‌داده PersianMHQA. که موجودیت پل در پاراگراف اول آن هایلایت شده است.

هدف ما در این مقاله بهبود کارآمدی سامانه‌های پرسش و پاسخ در زبان فارسی با جمع‌آوری یک مجموعه‌داده پرسش و پاسخ متنوع است که نیاز به استدلال چندگامی دارد. یکی از راه‌های انجام این کار، تعریف زنجیره‌های استدلال براساس پایگاه‌دانش است. با این حال، مجموعه‌داده‌های حاصل به دلیل ناقص‌بودن روابط موجودیت‌ها و عدم تنوع در نوع پرسش‌ها محدود شده‌اند. به همین دلیل، ما به منظور تنوع بخشیدن به پرسش و پاسخ‌ها، بر پاسخگویی به پرسش‌های مبتنی بر متن تمرکز می‌کنیم. منبع اطلاعاتی ما در جمع‌آوری این مجموعه‌داده دانش‌نامه ویکی‌پدیا فارسی است.

سازماندهی مقاله در ادامه به این شرح است. در بخش دوم کارهای مرتبط با پژوهش انجام شده معرفی می‌گردد. سپس در بخش سوم چگونگی جمع‌آوری داده‌ها و تولید مجموعه‌داده تشریح می‌گردد. در ادامه و در بخش چهارم مجموعه‌داده مورد تجزیه و تحلیل قرار می‌گیرد. در بخش پنجم برای محک‌زنی[۸] مجموعه‌داده و کسب اطمینان از کارآمدی آن، با انجام آزمایش‌های متعدد بهترین مدل‌های زبانی موجود که از زبان فارسی پشتیبانی می‌کنند با استفاده از مجموعه‌داده تنظیم‌دقیق میگردند و نتایج آن‌ها گزارش می‌گردد. در بخش آخر ضمن نتیجه‌گیری از پژوهش انجام شده پیشنهاداتی به عنوان کارهای آینده ارائه می‌گردد.

۲- کارهای مرتبط

در سال‌های اخیر، مجموعه‌داده‌های متعددی در حوزه سامانه‌های پرسش و پاسخ ارائه شده است. برای حوزه سامانه‌های پرسش و پاسخ به زبان فارسی نیز تعدادی مجموعه‌داده ارائه گردیده است که در ادامه به بررسی آن‌ها پرداخته شده است.

PeCoQ [۳] یک مجموعه‌داده برای سامانه‌های پرسش و پاسخ فارسی است. این مجموعه‌داده شامل ۱۰۰۰۰ پرسش و پاسخ پیچیده مبتنی بر گراف دانش فارس‌بیس[۹] [۴] است که متناسب برای سیستم‌های پرسش و پاسخ مبتنی بر گراف است. این مجموعه‌داده براساس تعدادی قاعده و الگو تولید شده است. برای هر پرسش، فرم منطقی اسپارکل و دو بازنویسی که توسط زبان‌شناسان نوشته شده نیز ارائه شده است. انواع مختلف پیچیدگی از جمله چند رابطه‌ای، چند موجودیتی، زمانی، عملیاتی در این مجموعه وجود دارد.

ParsSimpleQA [۵] یک مجموعه‌داده پرسش و پاسخ فارسی شامل ۳۶۱۲۲ پرسش ساده و اولین مجموعه‌داده توسعه داده شده برای پرسش‌های تک رابطه‌ای زبان فارسی بر مبنای گراف دانش فارس‌بیس است. این مجموعه‌داده به‌طور نیمه‌خودکار در دو مرحله ساخته شده است. ابتدا، الگوهای پرسش تک رابطه‌ای ساخته شده است. سپس، پرسش‌های ساده با استفاده از الگوها، موجودیت‌ها و روابط بر مبنای فارس‌بیس تولید شده‌اند.

PQuAD [۶] یک پیکره بزرگ پرسش و پاسخ دامنه‌باز برای پرسش‌های ساده بر مبنای صفحه‌های ویکی‌پدیا فارسی است. این مجموعه‌داده شامل ۸۰۰۰۰ پرسش ساده به همراه پاسخ‌هایشان است که ۲۵٪ آن‌ها دارای پاسخ نمی‌باشند. هدف از توسعه این مجموعه‌داده ساده کردن درک مطلب فارسی و نیز کمک به توسعه سامانه‌های پرسش و پاسخ فارسی بوده است. این پرسش‌ها توسط نیروی انسانی تولید شده و مورد حاشیه‌نویسی قرار گرفته است.

PerCQA [۷] اولین مجموعه‌داده فارسی برای انجمن پرسش و پاسخ است. این مجموعه‌داده شامل پرسش و پاسخ‌هایی است که از شناخته شده‌ترین انجمن فارسی خزش شده است. پس از جمع‌آوری داده‌ها، دستورالعمل‌های حاشیه‌نویسی دقیق را در یک فرآیند تکراری و سپس حاشیه‌نویسی جفت‌های پرسش و پاسخ ارائه شده است. این مجموعه‌داده شامل ۹۸۹ پرسش و ۲۱۹۱۵ پاسخ حاشیه‌نویسی شده است.

PersianQA [۸] یک مجموعه‌داده پرسش و پاسخ دامنه‌باز بر مبنای صفحات ویکی‌پدیا فارسی است. این مجموعه‌داده شامل ۹۰۰۰ پرسش ساده است که توسط نیروی انسانی طرح شده‌اند. هر پرسش، غیرقابل پاسخ‌دهی باشد یا پرسشی دارای یک یا چند

پاسخ باشد. در این مجموعه‌داده هر متن، دارای ۷ جفت پرسش با یک پاسخ و ۳ پرسش بدون پاسخ است.

IslamicPCQA [۹] یک مجموعه‌داده پرسش و پاسخ دامنه‌بسته برای پرسش‌های چندگامی زبان فارسی بر مبنای منابع غیرساخت‌یافته است. این مجموعه‌داده شامل ۱۲۲۸۲ پرسش و پاسخ پیچیده تولید شده براساس ۹ دانش‌نامه اسلامی است. پرسش و پاسخ در این مجموعه‌داده نیازمند استدلال بر روی بیش از یک پاراگراف است.

PersianQuAD [۱۰] شامل تقریبا ۲۰۰۰۰ سه‌تایی پرسش-متن-پاسخ است و بر مبنای مجموعه‌ای از صفحه‌های ویکی‌پدیا فارسی ساخته شده است. پاسخ هر پرسش بخشی از متن صفحه مربوطه به آن پرسش است. این مجموعه‌داده دارای پرسش‌های ساده است. همچنین این مجموعه‌داده با استفاده از سه مدل زبانی از پیش آموزش دیده MBERT [۱۱]، ALBERT-FA [۱۲] و پارس‌برت [۱۳] مورد ارزیابی قرار گرفته است.

ParSQuAD [۱۴] یک مجموعه‌داده پرسش و پاسخ به زبان فارسی است که از مجموعه‌داده معروف SQuAD 2.0 ترجمه شده است. این مجموعه در دو نسخه ارائه می‌شود: نسخه‌ای که به صورت دستی و دیگری که به صورت خودکار ویرایش شده است. ParSQuAD شامل بیش از ۱۹۰۰۰ سوال در نسخه دستی و بیش از ۷۲۰۰۰ سوال در نسخه خودکار است و با هدف ایجاد یک استاندارد برای سیستم‌های پرسش و پاسخ فارسی توسعه یافته است.

جدول ۱ مجموعه‌داده‌های موجود در زبان فارسی در حوزه سامانه‌های پرسش و پاسخ را براساس ویژگی‌های آن‌ها نشان می‌دهد.

۳- مجموعه‌داده

در این بخش راجع به چگونگی ایجاد مجموعه‌داده توضیح داده شده است. پرسش‌های موجود در مجموعه‌داده ما پرسش‌های دوگامی هستند، که براساس محتوای دو پاراگراف تولید شده‌اند. بنابراین برای رسیدن به پاسخ آن‌ها می‌بایست استدلال بر روی محتوای هر دو پاراگراف انجام شود. از آن‌جا که ممکن است طرح پرسش دوگامی از هر دو پاراگراف دلخواه دشوار باشد، بنابراین ما یک فرآیند خط‌لوله‌ای را به منظور تسهیل طرح پرسش‌های دوگامی معنادار و تولید جفت پاراگراف‌های مناسب طراحی کرده‌ایم که در شکل ۲ به نمایش درآمده است.

۳-۱- جمع‌آوری و پیش‌پردازش داده‌ها

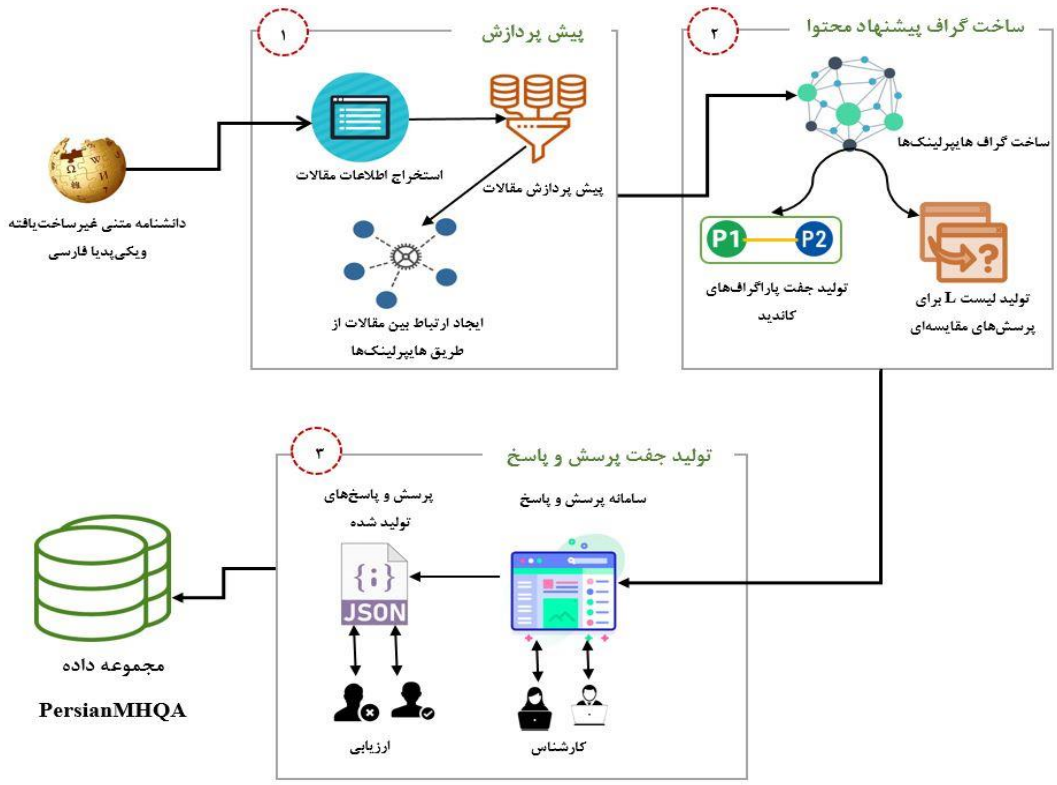
ما در این مقاله از دانشنامه متنی ویکی‌پدیا فارسی به عنوان منبع اطلاعاتی خود برای طرح پرسش چندگامی در زبان فارسی استفاده کرده‌ایم. به این منظور از آخرین نسخه دامپ ویکی‌پدیا فارسی در ۱ می ۲۰۲۲ استفاده شده است. دو فرض مهم در جمع‌آوری داده در نظر گرفته شده است که به شرح زیر است.

فرض اول) هر هایپرلینک در مقاله‌های ویکی‌پدیا مستلزم یک رابطه بین دو مقاله یا موجودیت است که بالقوه می‌توان برای طرح پرسش پیچیده از آن استفاده کرد.

فرض دوم) پاراگراف اول هر مقاله اغلب اطلاعات کافی برای طرح پرسش را دارد [۱۵]. با در نظر گرفتن فرضیه‌های بالا و براساس آن که پرسش‌های موجود در مجموعه‌داده PersianMHQA دوگامی در نظر گرفته شده‌اند، نیازمند به یک مجموعه مقاله آغازین برای تولید جفت پاراگراف‌های کاندید بودیم. بنابراین از تمامی مقاله‌های موجود در دامپ ویکی‌پدیای فارسی استفاده شده است.

جدول ۱: مقایسه مجموعه‌داده‌های موجود برای سامانه‌های پرسش و پاسخ در زبان فارسی

عنوان	سال انتشار	زبان	سطح پیچیدگی	اندازه	نوع منبع اطلاعاتی
PeCoQ	۲۰۲۰	فارسی	پیچیده	۱۰۰۰۰	گراف دانش فارس‌بیس
PerCQA	۲۰۲۱	فارسی	ساده	۹۸۹	انجمن‌های پرسش و پاسخ
ParSQuAD	۲۰۲۱	فارسی	ساده	۱۹۰۰۰، ۷۲۰۰۰	ویکی‌پدیا
PersianQA	۲۰۲۱	فارسی	ساده	۹۰۰۰	ویکی‌پدیا
ParsSimpleQA	۲۰۲۲	فارسی	ساده	۳۶۱۲۲	گراف دانش فارس‌بیس
PQuAD	۲۰۲۲	فارسی	ساده	۸۰۰۰۰	ویکی‌پدیا
PersianQuAD	۲۰۲۲	فارسی	ساده	۲۰۰۰۰	ویکی‌پدیا
IslamicPCQA	۲۰۲۳	فارسی	پیچیده	۱۲۲۸۲	ویکی‌شیعه، ویکی‌فقه، ویکی‌اهل‌بیت، امامت‌پدیا، اسلام‌پدیا، ویکی‌حج، ویکی‌نور، ویکی‌پاسخ و ویکی‌حسین
PersianMHQA	۲۰۲۴	فارسی	پیچیده	۷۰۰۰	ویکی‌پدیا



شکل ۲: نمایی کلی از فرایند ساخت مجموعه‌داده PersianMHQA.

مقاله‌های موجود با هدف در کنار هم قرار دادن مقاله‌هایی که از نظر معنایی و محتوایی با هم مرتبط هستند، پیش پردازش شدند. بسیاری از مقاله‌های موجود در ویکی‌پدیا دارای محتوای متنی نمی‌باشند بلکه شامل تصاویر، جداول، لینک به دیگر مقاله‌ها بوده و یا تنها دارای عنوان هستند. مقاله‌هایی نیز در ویکی‌پدیا وجود دارند که پاراگراف اول آن‌ها محتوای لازم به‌منظور طرح یک پرسش پیچیده معنادار را ندارد. همچنین تعدادی از مقاله‌ها در دارای هایپرلینک به دیگر مقاله‌ها نمی‌باشند. بنابراین، این مقاله‌ها شناسایی شده و کنار گذاشته شدند. در نهایت پس از انجام پیش پردازش، تعداد ۳۹۹۹۹ مقاله ویکی‌پدیا به عنوان مقاله‌های آغازین در نظر گرفته شده است.

۳-۲- ساخت گراف پیشنهاد محتوا

ماهیت مقاله‌های موجود در ویکی‌پدیا به دلیل وجود هایپرلینک بین مقاله‌ها مبتنی بر گراف است. بنابراین، به‌منظور تولید جفت پاراگراف‌های کاندید مناسب برای طرح پرسش پیچیده چندگامی و با داشتن مقاله‌های آغازین و هایپرلینک‌های موجود در پاراگراف اول آن‌ها ما نیز زیرساختی مبتنی بر گراف را بین این مقاله‌ها تشکیل داده‌ایم. گراف هایپرلینک بین مقاله‌ها یک گراف جهت‌دار G است که در آن هر گره نمایانگر یک مقاله و هر یال نمایانگر وجود یک هایپرلینک از مقاله اول به مقاله دوم است برای مثال یال (a, b) نشان‌دهنده وجود یک هایپرلینک در پاراگراف اول مقاله a به مقاله b است. در حقیقت مقاله‌های b زمینه مشترک بین دو مقاله را مشخص می‌کنند. در ادامه و در گام اول، با داشتن ۳۹۹۹۹ مقاله آغازین در مرحله قبلی و با استفاده از کتابخانه نتورک ایکس گراف هایپرلینک بین مقاله‌ها ایجاد شده است؛ به طوری که هر یک از ۳۹۹۹۹ مقاله آغازین موجود به عنوان یک گره در گراف در نظر گرفته شده است. در گام دوم، برای ایجاد ارتباط بین مقاله‌ها و تولید زنجیره چندگامی بین آن‌ها، هایپرلینک‌های موجود در پاراگراف اول این مقاله‌ها به عنوان مقاله‌های ثانویه در نظر گرفته شده و به عنوان گره به گراف افزوده می‌شوند. سپس یال‌های گراف بر مبنای ارتباط بین گره‌های آغازین و ثانویه به وجود می‌آیند. در پایان این گام، گراف ما دارای ۸۵۱۵۰ گره می‌باشد. در گام سوم، الگوریتم پیج‌رنک[۱۰] بر روی این گراف اعمال شده است تا مقاله‌ها با اهمیت بیشتر را در اختیار داشته باشیم. بعد از اعمال الگوریتم پیج‌رنک بر روی گراف جهت‌دار G، ۸۰ درصد از مقاله‌ها با بالاترین امتیاز و رتبه در گراف حفظ می‌شوند. در نهایت گراف جهت‌دار G دارای ۶۸۱۲۰ گره و ۲۰۰۵۵۷ یال است.

بنابراین تعداد ۲۰۰۵۵۷ جفت پاراگراف کاندید برای طرح پرسش‌های پل دوگامی تولید شده است.

همانطور که بیان گردید هدف در پرسش‌های مقایسه‌ای، مقایسه بین دو یا چند موجودیت از یک دسته است که منجر به نتایج جالب می‌شود. به‌منظور تهیه جفت پاراگراف‌های کاندید مناسب جهت طرح پرسش مقایسه‌ای، گره‌های موجود در گراف جهت‌دار G بر مبنای موضوعشان دسته‌بندی شده و در دسته‌های مختلف قرار داده شدند. بدین صورت هر مقاله با یک مقاله در دسته خود که دارای موضوع مشابه است، مقایسه خواهد شد و یک جفت پاراگراف کاندید را تولید خواهد کرد. به‌عنوان مثال مقایسه مقاله «دانشگاه علم و صنعت ایران» با مقاله «تهران» مقایسه معناداری نخواهد بود. اما براساس دسته‌بندی انجام شده مقایسه مقاله «دانشگاه علم و صنعت ایران» با مقاله «دانشگاه صنعتی خواجه نصیرالدین طوسی» منجر به یک مقایسه معنادار خواهد شد. به این ترتیب، تعداد ۲۴۲۵۱۲ جفت پاراگراف کاندید برای طرح پرسش مقایسه‌ای تولید شده است.

۳-۳- حاشیه‌نویسی پرسش و پاسخ

به منظور تسهیل فرآیند تولید پرسش و پاسخ از طریق حاشیه‌نگارها[11] براساس جفت پاراگراف‌هایی از طریق گراف جهت‌دار G تولید شدند، یک سامانه برچسب‌زنی توسعه داده شد. در این سامانه هر بار یک جفت پاراگراف کاندید نمایش داده می‌شود. برای طرح هر پرسش پیچیده به‌طور میانگین بین ۳ تا ۵ دقیقه زمان لازم است. برای حاشیه‌نگاری و ایجاد رکوردهای مجموعه‌داده بیش از ۲۲ دانشجوی کارشناسی ارشد مهندسی کامپیوتر گرایش نرم افزار و هوش مصنوعی به خدمت گرفته شدند. این افراد به صورت حضوری و غیرحضوری (با استفاده از سند دستورالعمل آموزشی مصور) مورد آموزش قرار گرفتند. این امر سبب شد تا طیف گسترده‌ای از تنوع موضوعی و پیچیدگی را در پرسش و پاسخ‌ها داشته باشیم. تعداد ۱۱۳۷۷ پرسش و پاسخ در نهایت از طریق حاشیه‌نگارها تولید گردید. علی‌رغم نظارت بر کیفیت تولید پرسش و پاسخ در حین انجام کار برای هر نفر از حاشیه‌نگارها توسط تیم خبره توسعه‌دهنده سامانه (شامل دانشجویان ارشد و دکتری مهندسی کامپیوتر)، با هدف افزایش کیفیت پرسش و پاسخ‌های طرح شده، این پرسش و پاسخ‌ها در انتها نیز توسط تیم خبره مورد بررسی مجدد قرار گرفتند. در نهایت تعداد ۷۰۰۰ پرسش نهایی شده و مجموعه‌داده PersianMHQA تشکیل گردید.

برای هر پرسش در این مجموعه‌داده علاوه بر پاسخ، حقیقت‌های پشتیبان در سطح جمله برای هر پاراگراف، نوع پرسش و نوع پاسخ نیز توسط حاشیه‌نگارها مشخص شده و حاشیه‌نویسی شده است. حقیقت‌های پشتیبان در هر پاراگراف در حقیقت همان جمله‌هایی هستند که پرسش براساس آن‌ها طرح شده و پاسخ نیز در یکی از آن‌ها قرار دارد. مشخص‌کردن حقیقت‌های پشتیبان برای هر پاراگراف به سامانه‌های پرسش و پاسخ کمک می‌کند تا مسیر استدلال برای رسیدن به پاسخ نهایی را توضیح دهد. شکل ۳ مثالی از انواع پرسش‌های دوگامی را در مجموعه‌داده PersianMHQA نشان می‌دهد.

۴- تجزیه و تحلیل مجموعه‌داده

در این بخش پرسش‌های موجود در مجموعه‌داده PersianMHQA را مورد تحلیل و بررسی قرار می‌دهیم.

۴-۱- نوع پرسش‌ها

پرسش‌های موجود در مجموعه‌داده ما همان‌طور که بیان شد در دو دسته پرسش‌های پل و مقایسه‌ای قرار دارند. از میان ۷۰۰۰ داده موجود در این مجموعه‌داده تعداد ۴۹۲۸ پرسش پل و تعداد ۲۰۷۲ پرسش مقایسه‌ای وجود دارد. جدول ۲ نرخ توزیع پرسش پل و مقایسه‌ای را نمایش می‌دهد.

جدول ۲: توزیع نوع پرسش در مجموعه‌داده PersianMHQA.

نوع پرسش	تعداد	توزیع
پل	4928	70.4
مقایسه‌ای	2072	29.6

۴-۲- تنوع موضوعی پاسخ‌ها

تنوع بالایی از انواع پاسخ در مجموعه‌داده PersianMHQA به کار رفته است. پاسخ‌های موجود در مجموعه‌داده ما از نظر تنوع موضوعی در ۱۱ دسته کلی دسته‌بندی می‌شوند؛ مانند: اسامی عام، مکان، تاریخ، گروه یا سازمان، شخص و غیره. شکل ۴ تعداد تکرار هر یک از ۱۱ نوع پاسخ را در مجموعه‌داده ما نشان می‌دهد. جدول ۳ دسته‌بندی نوع پاسخ را در مجموعه‌داده ما نشان می‌دهد.

(مثال اول) نمونه رکورد مجموعه داده برای پرسش چندگامی پل

پاراگراف اول: ویلیام پیت جوان‌تر

[۱] ویلیام پیت جوان‌تر سیاستمدار بریتانیایی و نخست‌وزیر بریتانیا به مدت ۱۷ سال بود. [۲] از او به عنوان جوان‌ترین نخست‌وزیر بریتانیا یاد می‌شود.[۳] وی هنگامی که به کاخ نخست‌وزیری رفت، تنها ۲۴ سال داشت. [۴] و در تاریخ ۱۷۸۳ جوان‌ترین نخست‌وزیر تاریخ بریتانیا شد و تا این تاریخ نیز در همین رتبه قرار دارد.

پاراگراف دوم: چارلز استانهوپ

[۱] چارلز استانهوپ مخترع بریتانیایی و برادرخوانده ویلیام پیت جوان‌تر بوده است. [۲] وی درس ریاضیات را در دانشگاه ژنو تحصیل کرد و مطالعاتی نیز در زمینه الکتریسیته انجام داده است. [۳] وی موفق شد به پیشرفت‌هایی در زمینه افزایش سرعت چاپ کاغذ دست یابد.

پرسش: برادرخوانده سیاستمدار بریتانیایی و نخست‌وزیر بریتانیا چه کسی بود؟

پاسخ: چارلز استانهوپ

نوع پاسخ: شخص

نوع پرسش: پل

حقیقت پشتیبان: ویلیام پیت جوان‌تر [۱]، چارلز استانهوپ [۱]

(مثال دوم) نمونه رکورد مجموعه داده برای پرسش چندگامی مقایسه‌ای

پاراگراف اول: موعود بنیادی‌فرد

[۱] موعود بنیادی‌فرد داور ایرانی فوتبال است. [۲] وی در سال ۱۳۸۱ داوری را شروع کرده و از سال ۱۳۹۰ در لیگ برتر فوتبال ایران قضاوت می‌کند. [۳] وی در سال ۲۰۱۳ وارد لیست بین‌المللی شد. [۴] بنیادی‌فرد در تیم داوری دربی شماره ۸۶ قرار داشت و داور چهارم این بازی مهم بود.

پاراگراف دوم: خداداد افشاریان

[۱] خداداد افشاریان داور ایرانی فوتبال است. [۲] وی در سال ۱۳۶۹ داوری را شروع کرده و در سال ۲۰۰۳ وارد لیست فیفا شده است. [۳] وی در اسفند ۱۳۹۹ به عنوان عضو هیات رئیسه فدراسیون فوتبال ایران انتخاب شد.

پرسش: آیا موعود بنیادی‌فرد زودتر از خداداد افشاریان داوری را شروع کرده است؟

پاسخ: خیر

نوع پاسخ: بله/خیر

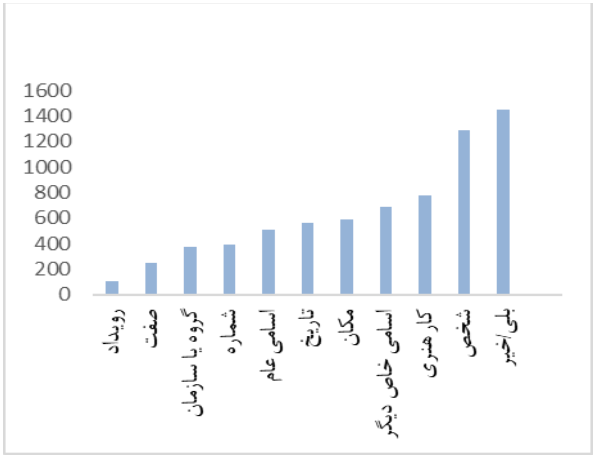
نوع پرسش: مقایسه‌ای

حقیقت پشتیبان: موعود بنیادی‌فرد [۰و۱]، خداداد افشاریان [۰و۱]

شکل ۳: مثال‌هایی از انواع پرسش‌های چندگامی در مجموعه‌داده PersianMHQA.

۵- محک‌زنی مجموعه‌داده

به منظور ارزیابی و محک‌زنی مجموعه‌داده PersianMHQA، داده‌ها به سه دسته آموزش (۸۰ درصد)، اعتبارسنجی (۱۰ درصد) و آزمون (۱۰ درصد) تقسیم شده است. سپس از مدل‌های زبانی از پیش آموزش داده شده مختلف برای ارزیابی و آزمون این مجموعه‌داده استفاده شده است.



شکل ۴: تعداد تکرار دسته‌بندی‌های موضوعی پاسخ در مجموعه‌داده PersianMHQA.

جدول ۳: دسته‌بندی نوع پاسخ‌های موجود در مجموعه‌داده PersianMHQA به همراه مثال و نرخ توزیع آن‌ها.

نوع پاسخ	مثال	تعداد	نرخ توزیع
رویداد	جنگ جهانی دوم، جنگ ایران و عراق	107	1.52
صفت	ملکه پاپ، صاحبقران	245	3.5
گروه یا سازمان	علم و صنعت، شرکت ماکروسافت	374	5.34
شماره	۱۷، ۷ میلیون	392	5.6
اسامی عام	صدر اعظم آلمان، ایدئولوژی حزب نازی	513	7.32
تاریخ	سال ۲۰۱۸، قرن بیستم	564	8.05
مکان	ایران، همدان	592	8.45
اسامی خاص دیگر	منظومه خورشیدی	692	9.88
کار هنری	تابلو لبخند مونالیزا، رمان درو	781	11.15
شخص	جرج اورول	1286	18.37
بلی/خیر	-	1454	20.77

۵-۱- جزییات آزمون

برای آزمون مجموعه داده از یک پردازنده گرافیکی انویدیا ای۱۰۰[۱۲] در مجموع با ۸۰ گیگابایت حافظه موقت استفاده شده است. برای انتخاب هایپرپارامترها در این پژ.هش، از بهینه‌ساز آدام استفاده شد که یکی از قدرتمندترین و کارآمدترین روش‌های بهینه‌سازی در حوزه یادگیری ماشین است. هدف از تنظیم هایپرپارامترها، دستیابی به بهترین عملکرد مدل با حداقل خطا و بیشترین دقت بود. جزئیات تنظیم هایپرپارامترها به شرح جدول ۴ است.

جدول ۴: هایپر پارامترهای استفاده شده برای تنظیم دقیق مدل‌های زبانی بر روی مجموعه‌داده PersianMHQA.

نام هایپرپارامتر	مقدار
تعداد دسته[۱۳]	16
تعداد تکرار[۱۴]	3
ماکزیمم طول[۱۵]	512
نرخ یادگیری[۱۶]	3e-5
اپسیلون آدام[۱۷]	1e-8
دراپ‌اوت[۱۸]	0.1
بذر[۱۹]	1
کاهندگی وزن[۲۰]	0.85

۵-۲- نتایج آزمون

در ادامه، ضمن معرفی مدل‌های زبانی تنظیم‌دقیق شده روی مجموعه‌داده، نتایج به‌دست آمده بر معیارهای اف‌وان[۲۱] و تطابق‌دقیق[۲۲] گزارش می‌گردد.

۵-۲-۱- مدل زبانی اول: آریابرت

مدل زبانی آریا برت[۲۳] [۱۶] به عنوان مدلی پیشرفته در زمینه درک زبان فارسی توسعه داده شده است. به منظور در اختیار داشتن داده‌های متنی متنوع برای آموزش این مدل زبانی، از انواع داده‌های متنی فارسی استفاده شده است؛ که در سه گروه متون محاوره‌ای، رسمی و ترکیبی شامل توییت‌ها، اخبار، شعرها، متون پزشکی، متون دایره‌المعارفی و نیز نظرات کاربران در وب‌سایت‌ها دسته‌بندی شده‌اند. حجم کل دادگان آموزشی استفاده شده به منظور آموزش این مدل بیش از ۳۲ گیگابایت است. این مدل، همچنین بر خلاف بسیاری از مدل‌های زبانی تک‌زبانه و چندزبانه که از مدل پایه برت استفاده می‌کنند، از معماری روبرتا [۱۷] و رمزگذاری بایت-جفت[۲۴] استفاده می‌کند. آریابرت در مقایسه با پارس‌برت و دیگر مدل‌های چندزبانه مبتنی بر برت، برای برخی از کارهای پردازش زبان طبیعی در فارسی به نتایج بهتری دست یافته است. مجموعه‌داده PersianMHQA بر روی مدل آریابرت براساس هایپرپارامترهای ذکر شده در جدول ۴ آموزش داده شده است. نتایج حاصل از این ارزیابی مقدار ۷۳.۱۱ برای معیار اف‌وان و مقدار ۷۰.۱۱ برای معیار تطابق‌دقیق است.

۵-۲-۲- مدل زبانی دوم: پارس‌برت

پارس‌برت [۱۳] یک مدل زبانی تک‌زبانه مبتنی بر مبدل است که بر مبنای معماری برت پایه توسعه یافته است. پارس‌برت از مدل زبانی برت چندزبانه سبک‌تر است و نتایج پیشرفته‌ای را در حوزه تحلیل احساسات، طبقه‌بندی متن، پرسش و پاسخ و شناسایی موجودیت نامدار به دست آورده است. برای آموزش این مدل زبانی از منابع مختلفی در زبان فارسی استفاده شده است. مجموعه‌داده PersianMHQA بر روی مدل پارس‌برت براساس هایپرپارامترهای ذکر شده در جدول ۴ آموزش داده شده است. نتایج حاصل از این ارزیابی مقدار ۷۵.۹۲ برای معیار اف‌وان و مقدار ۷۱.۷۳ برای معیار تطابق‌دقیق است.

۵-۲-۳- مدل زبانی سوم: ایکس‌ال‌ام روبرتا

ایکس‌ال‌ام روبرتا[۲۵] [۱۸] یک مدل زبانی مبتنی بر مبدل است که بر روی ۲.۵ ترابایت دادگان آموزشی از ۱۰۰ زبان مختلف آموزش داده شده است. ایکس‌ال‌ام روبرتا، یک مدل چندزبانه قوی است که در برخی از کارهای پردازش زبان طبیعی توانسته است نتایج بسیار بهتری را نسبت به مدل چندزبانه برت (MBERT) به دست آورد. مجموعه‌داده PersianMHQA بر روی مدل ایکس‌ال‌ام روبرتا براساس هایپرپارامترهای ذکر شده در جدول ۴ آموزش داده شده است. نتایج حاصل از این ارزیابی مقدار ۷۵.۳۸ برای معیار اف‌وان و مقدار ۶۹.۴۱ برای معیار تطابق‌دقیق است.

۵-۲-۴- مدل زبانی چهارم: لبسه

مدل مبدل جمله لبسه[۲۶] [۱۹]، یک مدل زبانی است که از یک معماری ترکیبی تشکیل شده است: یک شبکه عصبی بازگشتی برای پردازش جملات و یک شبکه عصبی پیچشی برای تولید بردارهایی با ابعاد بالا. لبسه بر روی دادگان وسیع و متنوع آموزش دیده است که از ۱۰۹ زبان مختلف پشتیبانی می‌کند. PersianMHQA بر روی مدل لبسه براساس هایپرپارامترهای ذکر شده در جدول ۴ آموزش داده شده است. نتایج حاصل از این ارزیابی مقدار ۷۰.۴۸ برای معیار اف‌وان و مقدار ۶۳.۷۶ برای معیار تطابق‌دقیق است.

۵-۲-۵- مدل زبانی پنجم: آلبرت

مدل زبانی آلبرت [۲۰] نسخه سبک‌تری از مدل زبانی برت است؛ که نسبت به برت تعداد پارامترهای کمتری دارد. آلبرت پیکره‌بندی بسیار بزرگتری نسبت به مدل زبانی برت با تعداد پارامترهای کمتر و عملکرد بهتر دارد. برای زبان فارسی نسخه تک‌زبانه آلبرت فارسی توسعه داده شده است که بر روی دادگانی شامل ۳.۹ میلیون سند، ۷۳ میلیون جمله و ۱.۳ میلیارد کلمه آموزش داده شده است. مجموعه‌داده PersianMHQA با مدل آلبرت براساس هایپرپارامترهای ذکرشده در جدول ۴ آموزش داده شده است. نتایج حاصل از این ارزیابی مقدار ۶۹.۳۹ برای معیار اف‌وان و مقدار ۶۳.۱۷ برای معیار تطابق‌دقیق است.

جدول ۵ نتایج حاصل از مدل‌های زبانی از پیش آموزش داده شده را بر روی مجموعه آزمون نشان داده است. طبق آزمایش‌های انجام شده مدل زبانی پارس‌برت توانسته است بهترین نتیجه را به دست آورد و مدل زبانی آلبرت بدترین نتایج را داشته است.

جدول ۵: نتایج حاصل از ارزیابی مجموعه‌داده PersianMHQA بر روی برخی از مدل‌های زبانی پشتیبانی کننده از زبان فارسی.

نام مدل	اف وان	تطابق دقیق
آریا برت	73.11	70.11
پارس برت	75.92	71.73
ایکس ال ام روبرتا (نسخه پایه)	75.38	69.41
لبسه (مبدل جمله)	70.48	63.76
آلبرت	69.39	63.17

۶- نتیجه‌گیری و کارهای آینده

در این مقاله با هدف بهبود کارآمدی سامانه‌های پرسش و پاسخ فارسی در پاسخ‌دهی به پرسش‌های پیچیده، اولین مجموعه‌داده از پرسش‌های چندگامی دامنه‌باز فارسی با نام PersianMHQA شامل ۷۰۰۰ پرسش پیچیده، متنوع و توضیح‌پذیر ارائه گردید. این مجموعه‌داده براساس محتوای دانشنامه ویکی‌پدیای فارسی و تلاش تعداد زیادی از حاشیه‌نگاران ایجاد شده است. مجموعه‌داده PersianMHQA طیف گسترده‌ای از پرسش‌های چندگامی شامل پرسش‌های پل و مقایسه‌ای را در بر می‌گیرد که برای هر پرسش، حقایق پشتیبان نیز مشخص شده‌اند. برای ارزیابی کارآمدی این مجموعه‌داده، از آخرین مدل‌های زبانی موجود برای زبان فارسی استفاده شد و نتایج براساس دو معیار اف‌وان و تطابق دقیق گزارش گردید.

به عنوان پیشنهادی برای کارهای آینده در وهله‌ی اول می‌توان با ارائه روش‌های جدید، گام‌های بلندتری برای عبور از خط‌پایه[۲۷] ایجاد شده در راستای بهبود پاسخگویی به پرسش‌های چندگامی فارسی برداشت و به نتایج بهتر روی معیارهای ارزیابی مختلف دست یافت.

همچنین می‌توان با استفاده از سازوکار پیشنهادی برای توسعه مجموعه‌داده، علاوه بر افزایش تعداد رکوردها، پیچیدگی تولید داده‌ها را از دوگامی به سطوح بالاتر ارتقا داد. بسیاری از پرسش‌ها وجود دارند که برای پاسخگویی به آن‌ها نیاز به پرش بیش از دو گام در اسناد و پاراگراف‌های متعدد است.

مراجع

[1] M. A. Calijorne Soares and F. S. Parreiras, "A Literature Review on Question Answering Techniques, Paradigms and Systems," J. King Saud Univ. - Comput. Inf. Sci., vol. 32, no. 6, pp. 635–646, 2020, doi: 10.1016/j.jksuci.2018.08.005.

[2] Y. Feldman and R. El-Yaniv, "Multi-hop Paragraph Retrieval for Open-domain Question Answering," ACL 2019 - 57th Annu. Meet. Assoc. Comput. Linguist. Proc. Conf., pp. 2296–2309, 2020, doi: 10.18653/v1/p19-1222.M.B.A. Haghighat, "Biometrics for Cybersecurity and Unconstrained Environments", Ph.D. Thesis, University of Miami, USA, 2016.

[3] R. Etezadi and M. Shamsfard, "PeCoQ: A Dataset for Persian Complex Question Answering over Knowledge Graph," in 2020 11th International Conference on Information and Knowledge Technology (IKT), Dec. 2020, pp. 102–106. doi: 10.1109/IKT51791.2020.9345610.

[4] M. B. Sajadi, B. Minaei, and A. Hadian, "Farsbase: A cross-domain farsi knowledge graph." SEMANTICS&Posters&Demos, 2018.

[5] H. Babaei Giglou, N. Beyranvand, R. Moradi, A. M. Salehoof, and S. Bibak, "ParsSimpleQA: The Persian Simple Question Answering Dataset and System over Knowledge Graph," in Proceedings of the 2nd International Workshop on Natural Language Processing for Digital Humanities, Nov. 2022, pp. 59–68. [Online]. Available: https://aclanthology.org/2022.nlp4dh-1.9

[6] K. Darvishi, N. Shahbodagh, Z. Abbasiantaeb, and S. Momtazi, "PQuAD: A Persian Question Answering Dataset," Feb. 2022.

[7] N. Jamali, Y. Yaghoobzadeh, and H. Faili, "PerCQA: Persian Community Question Answering Dataset." 2021.

[8] S. Ayoubi, https://github.com/sajjjadayobi/PersianQA.

[9] A. Ghafouri, H. Naderi, M. Aghajani, M. Firouzmandi, "IslamicPCQA: A Dataset for Persian Multi-hop Complex Question Answering in Islamic Text Resources." 2023.

[10] Kazemi, J. Mozafari, and M. A. Nematbakhsh, "PersianQuAD: The Native Question Answering Dataset for the Persian Language," IEEE Access, vol. 10, pp. 26045–26057, 2022, doi: 10.1109/ACCESS.2022.3157289.

[11] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pretraining of deep bidirectional transformers for language understanding," in NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, 2019, vol. 1, pp. 4171–4186. [Online]. Available: https://github.com/tensorflow/tensor2tensor

[12] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "ALBERT: A Lite BERT for Self-supervised Learning of Language Representations." 2020.

[13] M. Farahani, M. Gharachorloo, M. Farahani, and M. Manthouri, "ParsBERT: Transformer-based Model for Persian Language Understanding," Neural Process. Lett., vol. 53, no. 6, pp. 3831–3847, 2021, doi: 10.1007/s11063-021-10528-4.

[14] N. Abadani, J. Mozafari, A. Fatemi, M. A. Nematbakhsh, and A. Kazemi, "ParSQuAD: Machine Translated SQuAD dataset for Persian Question Answering," in 2021 7th International Conference on Web Research (ICWR), 2021, pp. 163–168. doi: 10.1109/ICWR51868.2021.9443126.

[15] Z. Yang et al., "Hotpotqa: A dataset for diverse, explainable multi-hop question answering," in Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018, 2020, pp. 2369–2380. doi: 10.18653/v1/d18-1259.

[16] A. Ghafouri, M. A. Abbasi, and H. Naderi. "AriaBERT: A Pre-trained Persian BERT Model for Natural Language Understanding." 2023.

[17] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach." arXiv preprint arXiv:1907.11692, 2019.

[18] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer and V. Stoyanov, "Unsupervised cross-lingual representation learning at scale," arXiv preprint arXiv:1911.02116, 2019.

[19] F. Feng, Y. Yang, D. Cer, N. Arivazhagan and W. Wang, "Language-agnostic BERT sentence embedding," arXiv preprint arXiv:2007.01852, 2020.

[20] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "ALBERT: A Lite BERT for Self-supervised Learning of Language Representations." 2020.

پاورقی‌ها:

[1] Complex Questions
[2] Effectiveness
[3] Open Domain
[4] Factoid
[5] Multi-Hop
[6] Bridge
[7] Comparison
[8] Benchmarking
[9] Farsbase
[10] Page Rank
[11] Taggers
[12] Nvidia A100
[13] Batch_Size
[14] N_Epochs
[15] Max_length
[16] Learning_Rate
[17] Adam_Epsilon
[18] Dropout
[19] Seed
[20] Lr_Decay
[21] F1 Measure
[22] Exact Match
[23] AriaBert
[24] Byte-Pair
[25] XLM-RoBERTa
[26] LaBSE
[27] Baseline