

## پرسش ۱. سامانه‌ی پرسش-پاسخ<sup>۱</sup>

در این پرسش با یک سامانه‌ی پرسش-پاسخ که از زیرشاخه‌های بسیار پرطرفدار پردازش زبان طبیعی است آشنا خواهید شد و ساخت یک نمونه از آن را تمرین خواهید کرد.

### ۱-۰. ساخت یک سیستم Extractive QA

سامانه‌های پاسخ‌گویی به پرسش‌ها (یا QA) در سال‌های اخیر به دلیل توانایی آن‌ها در استخراج خودکار پاسخ‌های یک پرسش، از یک قطعه معین در متن، توجه سرشاری را به خود جلب کرده‌اند. این سامانه‌ها در زمینه‌های مختلف از جمله بازیابی اطلاعات، پشتیبانی مشتری و دستیاران مجازی نقشی حیاتی دارند. ظهور مدل‌های یادگیری عمیق، مانند BERT [1] (Bidirectional Encoder Representations from Transformers)، موجب پیشرفت قابل توجهی در پردازش زبان طبیعی (NLP) شده‌اند. این مدل‌ها عملکرد قابل توجهی را در انجام وظایف مختلف NLP و دستیابی به نتایج state-of-the-art از خود نشان داده‌اند. توانایی BERT برای گرفتن اطلاعات متنی از هر دو زمینه چپ و راست با استفاده از مکانیسم‌های خودتوجهی، انقلابی در این زمینه ایجاد کرده است، و امکان درک دقیق‌تر زبان را فراهم نموده و راه را برای پیشرفت در تسک‌هایی همچون پاسخ‌گویی به سؤال، تحلیل احساسات، طبقه‌بندی متن و دیگر موارد هموارتر کرد.

هدف از این تمرین طراحی و پیاده‌سازی یک سیستم QA استخراجی (extractive question answering) می‌بنتی بر ترانسفورمرها می‌باشد که با دریافت یک متن و سؤال مربوط به آن، بهترین پاسخ مناسب را از متن استخراج می‌نماید. برخلاف سیستم‌های QA مولد که پاسخ‌ها را از ابتدا تولید می‌کنند، سیستم‌های QA استخراجی برای تولید پاسخ‌ها به اطلاعات موجود در متن زمینه متکی هستند. (مانند تصویر ۱) با استفاده از قدرت یادگیری عمیق و مدل‌های زبانی از پیش آموزش‌دیده‌شده مانند BERT، این سیستم‌ها می‌توانند به طور موثر ساختارهای پیچیده زبان، روابط متنی و تفاوت‌های معنایی لازم برای استخراج دقیق پاسخ را درک کرده و بهترین پاسخ را استخراج نمایند.

**Question:** The New York Giants and the New York Jets play at which stadium in NYC ?

**Context:** The city is represented in the National Football League by the New York Giants and the New York Jets , although both teams play their home games at MetLife Stadium in nearby East Rutherford , New Jersey , which hosted Super Bowl XLVIII in 2014 .

(Training example 29,883)

تصویر ۱ - نمونه pair سوال و متن و جواب مشخص شده در متن ، برگرفته از دیتاست squad

در این تمرین، شما با طراحی و پیاده‌سازی یک مدل مبتنی بر BERT بر روی مجموعه داده PQquad، وارد دنیای سیستم‌های QA extractive در زبان فارسی خواهید شد. در این تمرین شما ابتدا مجموعه داده‌ها را پیش پردازش می‌کنید، یک مدل QA مبتنی بر BERT طراحی نموده و پیاده می‌نمایید، استثناها را مدیریت می‌کنید و عملکرد سیستم را با استفاده از معیارهای ارزیابی مانند امتیاز تطابق دقیق (EM) و امتیاز F1 ارزیابی می‌کنید. از طریق این تمرین، شما تجربه عملی در استفاده از قابلیت‌های BERT برای استخراج پاسخ سؤالات، از متن‌های زمینه داده شده را به دست می‌آورید.

## ۱-۱ مدل سازی مسئله

(۲۰ نمره)

برای آشنایی و فهم عملکرد مدل ترنسفورمری BERT [1]، مقاله BERT اصلی (Devlin et al., 2018) را بخوانید تا معماری و تسک‌ها و توابع هدفی که BERT بر روی آنها پیش آموزش شده است (pretrain) را درک کنید. اجزای کلیدی مدل BERT، از جمله معماری ترنسفورمری، نمایش ورودی و اهداف pretraining آن را شرح دهید. سپس ساختار کلی مدل مورد نظر را خود برای حل این مسئله طراحی کرده و آنرا بصورت کامل بیان کنید. (برای این منظور بطور کامل توضیح دهید ورودی، خروجی، ساختار مدل، توابع خطای مورد استفاده چیست و مدل قرار است چه چیزی را آموزش ببیند).

## ۲-۱ پیش پردازش داده ها

(۲۰ نمره)

دیتاست مورد استفاده در این تمرین دیتاست PQuAD [2] می باشد که از طریق گیت هاب<sup>۱</sup> قابل دسترسی است. مقاله مرتبط با دیتاست ، ضمیمه گشته است. در ابتدا لازم است اطلاعات آماری دیتاست مورد نظر را نمایش داده و سپس پیش پردازش های لازم برای انجام تسک مورد نظر را بیان کنید . برای train ، validation و test از داده های با همین اسامی در دیتاست استفاده کنید.

## ۳-۱ پیاده سازی مدل

(۴۰ نمره)

برای پیاده سازی، از دو مدل مبتنی بر ParsBERT<sup>2</sup> [3] و ALBERT<sup>۳</sup> استفاده نمایید. مدل های از پیش آموزش دیده شده آن ها در huggingface قابل دسترسی می باشند. ParsBERT همان ساختار مدل BERT را داشته که بروی متون فارسی آموزش دیده شده است. ALBERT (A Lite BERT) یک مدل مبتنی بر ترانسفورمر ها است که بر اساس معماری BERT ساخته شده است، اما از تکنیک های اشتراک گذاری پارامتر را برای کاهش اندازه و نیازهای محاسباتی مدل استفاده کرده است ، مدل آموزش دیده شده ی آن بر روی دیتای فارسی قابل دسترس میباشد. شبکه طراحی شده در قسمت قبل را با استفاده از این دو مدل پیاده سازی نمایید. (توجه فرمایید استفاده از کلاس AutoModelForQuestionAnswering مجاز نمی باشد).

## ۴-۱. ارزیابی و پس پردازش (Postprocessing)

(۲۰ نمره)

در طول انجام تسک توسط مدل استثنائاتی در حین بارگیری داده ها، پیش پردازش و پس پردازش به دلیل طولانی بودن متون زمینه و محدودیت ورودی مدل های ترانسفورمری بوجود می آید، که نیاز است آنها را مدیریت کنید، اطمینان حاصل کنید که مدل شما می تواند اینگونه استثنائات را مدیریت کند ، ان ها را گزارش دهید. در نهایت پس از مدیریت استثنائات ، دو مدل آموزش دیده خود را بر روی مجموعه داده تست با استفاده از دو معیار EM و F1-score ارزیابی کنید و نتایج خود را با نتایج ذکر شده در مقاله مقایسه نمایید. (برای ارزیابی می توانید از ابزار های آماده استفاده نمایید).

<sup>1</sup> <https://github.com/AUT-NLP/PQuAD>

<sup>2</sup> <https://huggingface.co/HooshvareLab/bert-base-parsbert-uncased>

<sup>3</sup> <https://huggingface.co/m3hrdafi/albert-fa-base-v2>

- [1] Devlin et al (2019) “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding” NAACL 2019 . Available at: <https://doi.org/10.18653/v1/N19-1423>
- [2] Darvishi, K. et al. (2023) “PQuAD: A Persian question answering dataset,” Computer Speech & Language, 80, p. 101486. Available at: <https://doi.org/10.1016/j.csl.2023.101486>
- [3] Farahani, M., Gharachorloo, M., Farahani, M. et al. (2021). “ParsBERT: Transformer-based Model for Persian Language Understanding.” Neural Process Lett 53, 3831–3847 <https://doi.org/10.1007/s11063-021-10528-4>