

به نام خدا



دانشگاه تهران



دانشکده مهندسی برق و کامپیوتر

درس شبکه‌های عصبی و یادگیری عمیق

تمرین پنجم

محمدجواد مومنی‌نژاد	نام دستیار طراح	پرسش ۱
momeni.nezhad@ut.ac.ir	رایانامه	
حسام اسداله زاده	نام دستیار طراح	پرسش ۲
asadzadeh.hesam@ut.ac.ir	رایانامه	
۱۴۰۲.۰۳.۱۱	مهلت ارسال پاسخ	

فهرست

۱	قوانین
۳	پرسش ۱. ساخت یک سیستم QA
۳	ساخت یک سیستم Extractive QA
۴	۱-۱ مدل‌سازی مسئله (۱۰ نمره)
۴	۲-۱ پیش پردازش داده ها (۱۰ نمره)
۵	۳-۱ پیاده سازی مدل (۲۰ نمره)
۵	۴-۱ ارزیابی و پس پردازش (Postprocessing) (۱۰ نمره)
۶	منابع
۷	پرسش ۲ - استفاده از Vision Transformer برای طبقه‌بندی تصاویر
۷	۱-۲. آشنایی با تبدیل‌کننده تصاویر
۸	۲-۲. پیاده‌سازی و ارزیابی نتایج

شکل‌ها

شکل ۱. عنوان تصویر نمونه **Error! Bookmark not defined.**

جدول‌ها

جدول ۱. عنوان جدول نمونه **Error! Bookmark not defined.**

قبل از پاسخ دادن به پرسش‌ها، موارد زیر را با دقت مطالعه نمایید:

- از پاسخ‌های خود یک گزارش در قالبی که در صفحه‌ی درس در سامانه‌ی Elearn با نام **REPORTS_TEMPLATE.docx** قرار داده شده تهیه نمایید.
- پیشنهاد می‌شود تمرین‌ها را در قالب گروه‌های دو نفره انجام دهید. (بیش از دو نفر مجاز نیست و تحویل تک نفره نیز نمره‌ی اضافی ندارد) توجه نمایید الزامی در یکسان ماندن اعضای گروه تا انتهای ترم وجود ندارد. (یعنی، می‌توانید تمرین اول را با شخص A و تمرین دوم را با شخص B و ... انجام دهید)
- **کیفیت گزارش شما در فرآیند تصحیح از اهمیت ویژه‌ای برخوردار است؛** بنابراین، لطفاً تمامی نکات و فرض‌هایی را که در پیاده‌سازی‌ها و محاسبات خود در نظر می‌گیرید در گزارش ذکر کنید.
- در گزارش خود مطابق با آنچه در قالب نمونه قرار داده شده، برای شکل‌ها زیرنویس و برای جدول‌ها بالانویس در نظر بگیرید.
- الزامی به ارائه توضیح جزئیات کد در گزارش نیست، اما باید نتایج بدست آمده از آن را گزارش و تحلیل کنید.
- **تحلیل نتایج الزامی می‌باشد، حتی اگر در صورت پرسش اشاره‌ای به آن نشده باشد.**
- **دستیاران آموزشی ملزم به اجرا کردن کدهای شما نیستند؛** بنابراین، هرگونه نتیجه و یا تحلیلی که در صورت پرسش از شما خواسته شده را به طور واضح و کامل در گزارش بیاورید. در صورت عدم رعایت این مورد، بدیهی است که از نمره تمرین کسر می‌شود.
- **کدها حتماً باید در قالب نوت‌بوک با پسوند ipynb تهیه شوند، در پایان کار، تمامی کد اجرا شود و خروجی هر سلول حتماً در این فایل ارسالی شما ذخیره شده باشد.** بنابراین برای مثال اگر خروجی سلولی یک نمودار است که در گزارش آورده‌اید، این نمودار باید هم در گزارش هم در نوت‌بوک کدها وجود داشته باشد.
- **در صورت مشاهده‌ی تقلب امتیاز تمامی افراد شرکت‌کننده در آن، 100- لحاظ می‌شود.**
- تنها زبان برنامه نویسی مجاز **Python** است.
- **استفاده از کدهای آماده برای تمرین‌ها به هیچ وجه مجاز نیست.**

- نحوه محاسبه تاخیر به این شکل است: پس از پایان رسیدن مهلت ارسال گزارش، حداکثر تا یک هفته امکان ارسال با تاخیر (به ازای هر روز 5 درصد کسر نمره) وجود دارد، پس از این یک هفته نمره آن تکلیف برای شما صفر خواهد شد.
- لطفا گزارش، کدها و سایر ضمایم را به در یک پوشه با نام زیر قرار داده و آن را فشرده سازید، سپس در سامانه‌ی Elearn بارگذاری نمایید:

HW[Number]_[Lastname]_[StudentNumber]_[Lastname]_[StudentNumber].zip

(مثال: HW1_Ahmadi_810199101_Bagheri_810199102.zip)

- برای گروه‌های دو نفره، بارگذاری تمرین از جانب یکی از اعضا کافی است ولی پیشنهاد می‌شود هر دو نفر بارگذاری نمایند.

پرسش ۱. سامانه‌ی پرسش-پاسخ^۱

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^۱ 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 ParsBERT² [3] و ALBERT^۳ 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 بدلیل 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).

¹ <https://github.com/AUT-NLP/PQuAD>

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

³ <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>

پرسش ۲ – استفاده از Vision Transformer برای طبقه‌بندی تصاویر

در این پرسش با کاربرد تبدیل‌کننده‌ها در تصویر آشنا خواهید شد و مقاله‌ای را در این رابطه پیاده‌سازی خواهید کرد.

۲-۱. آشنایی با تبدیل‌کننده تصاویر

در توسعه‌های اخیر حوزه بینایی ماشین، افزایش محسوسی در استفاده از ساختارهای مبتنی بر ترنسفورمر مشاهده شده است. این ساختارها عملکردی بهتر از ساختارهای شبکه عصبی کانولوشنی (CNN) ارائه می‌دهند؛ اما از سوی دیگر، هزینه محاسباتی آنها برای آموزش از ابتدا بسیار زیاد است. از آنجایی که این مدل‌ها در حوزه بینایی ماشین به تازگی معرفی شده‌اند، نیاز به مطالعه قابلیت‌های یادگیری انتقالی آنها و مقایسه آن با CNN‌ها وجود دارد تا بتوانیم ساختار مناسب‌تر را پیدا کرده و هنگام استفاده در مسائل واقعی با مجموعه داده‌های کوچک از آنها استفاده کنیم.

این تبدیل‌کننده‌های تصویر با عملکرد بالا با استفاده از صدها میلیون تصویر به عنوان پیش‌آموزش، با یک زیرساخت بزرگ آموزش داده شده‌اند، که به همین دلیل توانایی استفاده مجدد از آنها محدود شده است. مدل DeiT، تبدیل‌کننده‌ی بدون کانولوشن است که فقط با آموزش بر روی ImageNet ایجاد شده. مدل DeiT، از یک استراتژی معلم-شاگرد خاص برای تبدیل‌کننده‌ها استفاده می‌کند. این استراتژی بر یک توکن خلاصه‌سازی^۱ تکیه می‌کند که مطمئن شود که شاگرد از طریق مکانیزم توجه از معلم یاد می‌گیرد. این روش خلاصه‌سازی مبتنی بر توکن، به خصوص زمانی که یک شبکه کانولوشنی به عنوان معلم استفاده می‌شود، بهتر عمل می‌کند. دسترسی به مقاله DeiT از طریق پیوند زیر ممکن است:

<https://arxiv.org/abs/2012.12877>

همانطور که می‌دانید یکی از روش‌های استفاده مجدد از مدل‌ها، fine-tuning است. در مقاله‌ی

Investigating Transfer Learning Capabilities of Vision Transformers and CNNs by Fine-Tuning a Single Trainable Block

روشی پیشنهاد شده که فقط با فاین-تیون کردن وزن‌های آخرین بلاک تبدیل‌کننده و MLP Head مدل، بتوان مدل را فاین-تیون کرد. از طریق پیوند زیر می‌توانید به مقاله ذکر شده دسترسی داشته باشید:

¹ Distillation

۲-۲. پیاده‌سازی و ارزیابی نتایج

در این بخش ابتدا به پیاده‌سازی مقاله، سپس ارزیابی نتایج خود خواهید پرداخت:

۲-۲-۱- لود کردن دیتاست و انجام پیش‌پردازش‌های لازم

(۱۰ نمره)

برای پیاده‌سازی این بخش در محیط گوگل کولب لازم است کتابخانه‌های transformers و datasets را با دستور pip install نصب کنید (این کتابخانه‌ها به صورت پیش‌فرض روی محیط کولب نصب نیستند).

حال دیتاست CIFAR-10 را لود کرده و در صورت لزوم، پیش‌پردازش‌های ذکر شده در مقاله را انجام دهید.

۲-۲-۱- شبکه کانولوشنی

(۳۰ نمره)

پس از مطالعه‌ی مقاله‌ی بالا، یکی از مدل‌های تماماً کانولوشنی را انتخاب کرده و با unfreeze کردن لایه‌های ذکر شده در مقاله، مدل را روی دیتاست CIFAR-10 فاین-تیون کنید. نتایج Validation Loss و Validation Accuracy را گزارش کنید.

۲-۲-۲- شبکه ViT (تبدیل‌کننده تصویر)

(۶۰ نمره)

یکی از مدل‌های تماماً ترنسفورمری ذکر شده در مقاله را انتخاب کرده و با unfreeze کردن لایه‌های ذکر شده در مقاله، مدل را روی دیتاست CIFAR-10 فاین-تیون کنید. می‌توانید از آموزش‌های Hugging Face برای نحوه فاین-تیون کردن مدل تبدیل‌کننده تصویر خود استفاده کنید.

نتایج Validation Accuracy و Validation Loss را گزارش کنید. نتایج خود را با نتایج مقاله مقایسه کنید.