



دانشگاه صنعتی امیرکبیر  
دانشکده مهندسی کامپیوتر  
پروژه درس رایانش عصبی و یادگیری عمیق



## پروژه هفتم

هدف: آشنایی با مکانیزم توجه و معماری ترنسفورمر

کد: پیاده سازی این پروژه را به زبان پایتون انجام دهید؛ در این فعالیت مجاز به استفاده از tensorflow یا pytorch یا jax می‌باشید. فایل‌های کد خود را بر اساس شماره سوال و زیر قسمت خواسته شده‌ی آن نام گذاری کنید (برای مثال می‌توان نام گذاری قسمت اول برای سوال سوم تمرین را بصورت `P3_a_preprocessing.py` در نظر گرفت). فایل‌های ارسالی‌تان بایستی با فرمت `py` یا `ipynb` (با حفظ خروجی هر سلول) باشد.

گزارش: ملاک اصلی انجام فعالیت، گزارش آن است و ارسال کد بدون گزارش فاقد ارزش است. برای این فعالیت یک فایل گزارش در قالب pdf تهیه کنید که دارای فهرست بوده و پاسخ‌ها بترتیب در آن قرار گرفته اند و نام، نام خانوادگی و شماره دانشجویی‌تان در قسمت چپ سربرگ تمامی صفحات تکرار شده است. علاوه بر خواسته‌ی مستقیم هر سوال، مقتضی است که نمودارهای خطا (loss) و صحت (accuracy) را به ازای مجموعه داده‌های آموزش و اعتبارسنجی رسم نمایید. همچنین در صورت امکان ماتریس درهم‌ریختگی را بصورت رنگ‌آمیزی شده به همراه اعداد متناظر برای مجموعه داده‌های آموزش، آزمون و اعتبارسنجی نیز تولید نمایید. لازم به ذکر است که در هر آموزش بایستی موارد مهم تنظیم شده نظیر تابع خطا، بهینه‌ساز (به همراه پارامترهای تنظیم شده‌ی آن مانند نرخ یادگیری)، معماری شبکه‌ی آموزشی (کتابخانه‌ها و ابزارهایی برای بصری‌سازی موجود است)، تعداد گام آموزشی، اندازه دسته (Batch Size)، آمارگان تفکیک مجموعه داده (به آموزش، آزمون و اعتبارسنجی)، پیش‌پردازش‌های اعمالی بروی داده‌گان ورودی و... ذکر گردد.

تذکر: مطابق قوانین دانشگاه هر نوع کپی برداری و اشتراک کار دانشجویان غیر مجاز بوده و با تمامی طرفین برخورد خواهد شد. استفاده از کدها و توضیحات اینترنت به منظور یادگیری صرفاً با ارجاع به آن بلامانع است، اما کپی کردن آن غیرمجاز است.

راهنمایی: در صورت نیاز می‌توانید سوالات خود را در خصوص پروژه از تدریس‌یارهای درس، از طریق ایمیل زیر یا در گروه تلگرامی بپرسید. [لینک گروه تلگرامی](#)

Email: [ann.ceit.aut@gmail.com](mailto:ann.ceit.aut@gmail.com) CC: [m.ebadpour@aut.ac.ir](mailto:m.ebadpour@aut.ac.ir)

توجه: می‌توانید از منابع و بسترهای سخت افزاری برخط رایگان نظیر Google Colab یا Kaggle استفاده نمایید.

تاخیر مجاز: در طول ترم، ده روز زمان مجاز تاخیر برای ارسال پروژه‌ها در اختیار دارید (بدون کسر نمره). این تاخیر را می‌توانید بر حسب نیاز بین پروژه‌ها مختلف تقسیم کنید که مجموع آن نباید بیشتر از ده روز شود. پس از استفاده از این تاخیر مجاز، هر روز تاخیر باعث کسر ۱۰٪ نمره‌ی کسب شده‌ی آن تمرین خواهد شد.

ارسال: فایل‌های کد و گزارش خود را در قالب یک فایل فشرده با فرمت `StudentID_HW07.zip` تا تاریخ ۱۴۰۳/۰۴/۲۰ صرفاً از طریق سایت کورسز ارسال نمایید. ارسال از طریق تلگرام، ایمیل و سایر راه‌های ارتباطی مجاز نبوده و تصحیح صورت نخواهد گرفت.

۱- یکی از دلایل نیاز به مکانیزم توجه، گلوگاهی بود که بین رمزگذار و رمزگشا در مدل‌های seq2seq به وجود می‌آمد. این مشکل را توضیح دهید و نشان دهید چطور مکانیزم توجه این مشکل را حل کرد. یکی دیگر از مشکلات، عدم توجه مدل به گذشته دور بود. به طور مثال در یک متن به کلمات نزدیک‌تر اهمیت بیشتری داده می‌شد تا کلمات دورتر و وزن کلمات دورتر به صورت نمایی کاهش پیدا می‌کرد. آیا استفاده از lstm و یا lstm دوطرفه می‌تواند این مشکل را به طور کامل رفع کند؟ توضیح دهید. (۵ نمره)

۲- در شبکه‌های بازگشتی ما می‌توانستیم با استفاده از خروجی مرحله قبل در ورودی، تاریخچه و گذشته را مدل کنیم. اما با توجه به اینکه مدل‌های ترنسفورمر از شبکه‌های بازگشتی استفاده نمی‌کنند، چطور می‌توانند بهتر از شبکه‌های بازگشتی گذشته را در نظر بگیرند (نشان دهید). مشکلات ترنسفورمر را در مقایسه با شبکه‌های بازگشتی بیان کنید. (۵ نمره)

۳- ترنسفورمرها نسبت به شبکه‌های seq2seq قابلیت موازی‌سازی بیشتری دارند. با ذکر جزئیات توضیح دهید. (۵ نمره)

۴- یکی از مشکلات ترنسفورمرها مرتبه هزینه محاسباتی و هزینه ذخیره‌سازی عملیات self-attention است که از مرتبه  $O(N^2)$  می‌باشد. تلاش‌هایی برای کاهش این مشکل انجام شد. مقالاتی نشان دادند که عملگر softmax باعث می‌شود تا بتوانیم این پیچیدگی را کاهش دهیم. توضیح دهید چرا عملگر softmax باعث وجود این مسئله می‌شود. همچنین یکی از پیشنهادها برای حل این مشکل استفاده از مکانیزم‌های توجه کرنلی است. در مورد این مکانیزم تحقیق کنید و نشان دهید چطور این روش منجر به کاهش پیچیدگی می‌شود. یک کرنل به دلخواه انتخاب کنید و عبارت (۱) را بازنویسی کنید و مرتبه زمانی و حافظه مورد نیاز برای عملگر self-attention را محاسبه کنید. لطفاً به مقاله که برای انتخاب کرنل مراجع کردید، ارجاع دهید. (۱۰ نمره)

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

۵- در دوران ابتدایی برای اینکه درک بهتری از جملات و جایگاه کلمات در جمله داشته باشیم تمرینی تحت عنوان "با کلمات زیر جمله بسازید" داشتیم. در این سوال می‌خواهیم یک مدل ترنسفورمر را از ابتدا برای این وظیفه آموزش دهیم. به این منظور مراحل زیر را دنبال کنید. (۵۰ نمره)

- مجموعه داده‌ای فارسی به انتخاب خودتان از اینترنت دانلود کنید.
- جملات هر متن را جدا کنید. (ممکن است چالش‌هایی داشته باشید. ایده این قسمت را بطور کامل بیان کنید. در صورتی که بتوانید تا حد خوبی جملات هر متن را جدا کنید، نمره اضافه برای شما در نظر گرفته می‌شود).
- مجموعه داده مربوط به این سوال را بسازید. ستون اول جمله‌ای که به صورت تصادفی کلماتش جابجا شدند و ستون دوم مرتب شده آن جمله است.

- مدل ترنسفورمر خود را پیاده‌سازی کنید و مدل را آموزش دهید. دقت کنید برای رسیدن به صحت مناسب به دیتا زیادی نیاز دارید و ممکن است منابع شما محدود باشد. در اینجا با توجه به منابع خودتان این موضوع را مدیریت کنید. یک دقت حداقلی برای این سوال کافی است.
- مدل را با داده‌های آزمون ارزیابی کرده. ۵ نمونه از داده‌های آزمون را به صورت تصادفی انتخاب کرده، کلمات آن را جابجا کنید و به مدل بدهید. قبل و بعد این ۵ نمونه را در گزارش خود بیاورید.
- توضیح دهید در مرحله قبل با چه روشی مدل را ارزیابی کردید و دلایل خود را بیان کنید.

۶- مجموعه داده CoLA (Corpus of Linguistic Acceptability) یک مجموعه داده مهم در زمینه پردازش زبان طبیعی (NLP) است که برای ارزیابی مقبولیت زبانی جملات استفاده می‌شود. مقبولیت زبانی به این معنی است که آیا یک جمله از نظر دستوری و نحوی توسط گویشوران بومی یک زبان درست است یا نه. در این سوال قصد داریم تا با تنظیم-دقیق مدل BERT، یک طبقه‌بند دو کلاسه برای این مجموعه داده پیاده‌سازی کنیم. موارد زیر را دنبال کنید: (۳۰ نمره امتیازی)

- دو فایل "in\_domain\_train.tsv" و "out\_of\_domain\_dev.tsv" در اختیار شما قرار گرفته است. این فایل‌ها را در محیط برنامه‌نویسی خود بارگذاری کنید. پیش پردازش‌های لازم مانند اضافه کردن کارکترهای خاص ([SEP] و ...) به جملات، توکنایز کردن و ...
- ۱۰ درصد از داده‌های "in\_domain\_train.tsv" را به برای اعتبارسنجی در نظر بگیرید.
- مدل BERT را بارگذاری و پیکره‌بندی کنید. (پیشنهاد می‌شود از کتابخانه transformers استفاده کنید).
- مدل را آموزش دهید. در هر epoch، خطا و صحت را برای داده‌های اعتبارسنجی چاپ کنید. همچنین بعد از اتمام آموزش نمودار خطا را به ازای هر دسته (batch) آموزش رسم کنید. (هر epoch می‌تواند شامل چندین دسته باشید).
- از داده‌های out\_of\_domain\_dev.tsv برای ارزیابی مدل تنظیم-دقیق شده خود استفاده کنید. برای این قسمت از معیار F1 و  $MCC^1$  استفاده کنید. این معیار را توضیح دهید و بگویید چرا استفاده از این معیار در اینجا نسبت به F1 بهتر است.
- معیار MCC شما برای داده‌های out\_of\_domain\_dev.tsv نباید کوچکتر از ۰.۵ باشد.

---

موفق باشید

---

<sup>1</sup> Matthews Correlation Coefficient