



پروژه درس رایانش عصبی و یادگیری عمیق

پروژه هفتم

هدف: آشنایی با شبکه ترنسفورمر

کد: پیاده سازی این پروژه را به زبان پایتون انجام دهید؛ در این فعالیت مجاز به استفاده از tensorflow یا pytorch می‌باشید. همچنین برای استفاده از مدل‌های آماده ترنسفورمر می‌توانید از کتابخانه [transformers](#) استفاده کنید.

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 و در آن برای هر سوال، تصاویر ورودی، تصاویر خروجی و توضیحات کامل و جامعی تهیه کنید. تذکر: مطابق قوانین دانشگاه هر نوع کپی برداری و اشتراک کار دانشجویان غیر مجاز بوده و شدیداً برخورد خواهد شد. استفاده از کدها و توضیحات اینترنت به منظور یادگیری بلامانع است، اما کپی کردن 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 در سایت کورسز در دسترس بوده و قبلاً به همه‌ی دانشجویان ایمیل شده است)

Email: ann.ceit.aut@gmail.com

توجه: برای آموزش شبکه های عمیق می‌توانید از منابع و بسترهای سخت افزاری برخط رایگان نظیر Google Colab یا Kaggle استفاده نمایید.

تاخیر مجاز: در طول ترم، مجموعاً مجاز به حداکثر ده روز تاخیر برای ارسال تمرینات هستید (بدون کسر نمره). این تاخیر را می‌توانید بر حسب نیاز بین تمرینات مختلف تقسیم کنید؛ اما مجموع تاخیرات تمام تمرینات شما نباید بیشتر از ده روز شود. پس از استفاده از این تاخیر مجاز، هر روز تاخیر باعث کسر ۱۰٪ نمره‌ی آن تمرین خواهد شد.

ارسال: فایل های کد و گزارش خود را در قالب یک فایل فشرده با فرمت StudentID_HW07.zip تا تاریخ ۱۳/۰۴/۰۲ ارسال نمایید.

مدل‌های زبانی بزرگ^۱ در سال‌های اخیر جریان قالب در حوزه پردازش زبان‌های طبیعی بوده‌اند. مدل‌های زبانی بزرگ ساختارهای عصبی عمیق (معمولا مبتنی بر ترنسفورمر) هستند که بر روی داده‌های متنی فراوان (معمولا به صورت بدون نظارت) آموزش دیده‌اند. به لطف داده‌های فراوان، آموزش منعطف و البته قدرت پردازشی صدها عدد پردازنده گرافیکی، این مدل‌ها با کمترین تنظیم دقیق^۲ و یا حتی بدون آن، در بسیاری از وظایف سطح بالای حوزه پردازش زبان به بهترین نتایج رسیده‌اند^۳.

الف) ایده (ها) و کاربرد مدل زبانی BERT را به اختصار توضیح دهید. چرا این مدل تنها از بخش کدگذار ترنسفورمر استفاده می‌کند؟

ب) ایده (ها) و کاربرد مدل زبانی GPT-2 را به اختصار توضیح دهید. چرا این مدل تنها از بخش کدگشای ترنسفورمر استفاده می‌کند؟

در ادامه با چند روش مختلف استفاده از شبکه BERT در حل مسئله دسته‌بندی متون آشنا خواهید شد. دیتاست مسئله را از این [لینک](#) دریافت کنید. این دیتاست شامل نظرات کاربران توئیتر درباره خطوط هوایی ایالات متحده آمریکا در سال ۲۰۱۶ است. هدف از این مسئله، دسته‌بندی نظرات کاربران با توجه به متن توئیت آن‌ها، به سه دسته مثبت، منفی و خنثی است^۴.

توجه: استفاده از کتابخانه transformers در حل این پروژه مجاز است و پیشنهاد می‌شود. از این کتابخانه صرفا برای ساخت مدل مناسب در ترنسورفلو استفاده شود. آموزش شبکه، ساخت دیتاست و یا پس پردازش خروجی‌ها (برای قسمت آخر تمرین) به وسیله این کتابخانه مجاز نیست.

ج) پیش پردازش‌های لازم را با استفاده از کتابخانه‌های آماده یا به صورت دستی بر روی داده‌ها انجام دهید و در گزارش خود ارائه کنید. ورودی‌های مسئله تنها شامل متن توئیت و خروجی یکی از سه کلاس دیتاست است.

د) هیستوگرام کلاس‌های دیتاست را رسم کنید. با توجه به این هیستوگرام، دیتاست از نوع نامتوازن^۵ است. برای رفع این مشکل از یک روش مناسب استفاده کنید.

ه) یک مدل BERT را بر روی داده‌های آموزشی تنظیم دقیق کنید. نتایج شامل نمودارهای تابع هزینه آموزش و ارزیابی و دقت نهایی بر روی داده‌های تست را گزارش کنید.

و) تنظیم دقیق BERT اگرچه در مسائلی با داده‌های زیاد معمولا مفید است، اما در مسائل با داده‌های کم ممکن است منجر به بیش برآزش مدل شود. همچنین تنظیم دقیق مدل BERT معمولا نیاز به منابع پردازشی زیادی دارد. برای رفع این دو مشکل، معمولا از مدل BERT به عنوان استخراج کننده ویژگی با وزن‌های ثابت استفاده می‌شود. شما نیز با یک مدل BERT با وزن‌های ثابت به علاوه یک لایه تماما متصل (FC) قسمت قبل را تکرار و مقایسه کنید.

توجه: از آنجایی که طول ورودی و خروجی BERT معمولا متغیر و تعداد کلمات ممکن است بسیار زیاد باشد، می‌توانید از یک لایه pooling برای تجمیع ویژگی‌های استخراج شده برای هر کلمه، پیش از لایه FC استفاده کنید.

¹ large language model (LLM)

² Fine-tune

³ Kenton, Jacob Devlin Ming-Wei Chang, and Lee Kristina Toutanova. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." Proceedings of NAACL-HLT. 2019. Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).

^۴ در ادبیات پردازش زبان‌های طبیعی به این مسئله تحلیل احساسات (sentiment analysis) گفته می‌شود.

⁵ imbalance

ز-امتیازی) یکی از مزایای بسیار مهم مدل‌های زبانی بزرگ، تعمیم‌پذیری آن‌ها به بسیاری از وظایف سطح بالا حتی بدون تنظیم دقیق است. چگونگی انجام این کار را با روش Prompt-Tuning توضیح دهید. یک راهکار برای استفاده از این روش در شبکه BERT برای مسئله این پروژه ارائه و پیاده‌سازی کنید. دقت نهایی روش خود را بر روی داده‌های تست ارائه کنید.

توجه: همانطور که گفته شد این روش دارای هیچ مرحله تنظیم دقیقی نخواهد بود و از وزن‌های BERT مستقیماً برای حل مسئله استفاده می‌شود.