

به نام خدا



دانشگاه تهران



دانشکده مهندسی برق و کامپیوتر

درس شبکه‌های عصبی و یادگیری عمیق

تمرین پنجم

نام دستیار طراح	مریم دادخواه	پرسش ۱
رایانامه	m.dadkhah99@gmail.com	
نام دستیار طراح	محمدجواد مومنی‌نژاد	پرسش ۲
رایانامه	Momeni.nezhad@ut.ac.ir	
مهلت ارسال پاسخ	۱۴۰۲.۱۰.۱۵	

۱	قوانین
۳	پرسش ۱. تشخیص احساسات گفتار (SER)
۳	۱-۱. معرفی مدل HuBERT
۵	۲-۱. سوالات تشریحی
۵	۱-۲-۱. چالش های داده های صوتی در یادگیری
۵	۲-۲-۱. رویکرد HuBERT
۵	۳-۱. معرفی مجموعه دادگان
۵	۱-۳-۱. پیش پردازش داده ها
۶	۲-۳-۱. ساخت دیتالودر
۶	۴-۱. تعریف مسئله
۷	۱-۴-۱. تولید بازنمایی مناسب از کل دنباله ورودی
۷	۲-۴-۱. آموزش مدل
۷	نکات پیاده سازی
۹	پرسش ۲. تنظیم دقیق مدل BERT
۹	۱-۲. آموزش و تحلیل مدل BERT
۹	۱-۱-۲. پیش پردازش داده ها
۹	۲-۱-۲. تنظیم دقیق مدل (fine tune)
۱۰	۳-۱-۲. فریز کردن لایه های مدل
۱۰	۴-۱-۲. تنظیم دقیق مدل بر روی لایه های میانی
۱۰	۵-۱-۲. حذف head های attention در مدل

شکل‌ها

شکل ۱. شمایی کلی از فرایند پیش آموزش مدل HuBERT ۴

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 ... هم‌گروه باشید)
- **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 یک درصد تجمعی) 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. تشخیص احساسات گفتار (SER)^۱

در این پرسش با کاربرد تبدیل کننده‌ها^۲ در صوت آشنا خواهید شد و مقاله ای را در این رابطه پیاده سازی خواهید کرد.

۱-۱. معرفی مدل HuBERT

در سال های اخیر، مدل های مبتنی بر ترنسفورمرها تحول چشمگیری در حوزه پردازش زبان های طبیعی و گفتار ایجاد کرده اند و به مدل های مورد استفاده برای طیف گسترده ای از وظایف، از جمله تجزیه و تحلیل احساسات، تشخیص گفتار و غیره تبدیل شده اند. این مدلها به دلیل درک وابستگی های بلندبرد^۳ و استفاده از مکانیزم های توجه^۴، تبدیل به ابزارهای قدرتمندی در پردازش داده های دنباله ای^۵ گردیده اند. مدل Hubert یک مدل ترنسفورمری است که به منظور تولید بازنمایی صوتی و بر پایه مدل BERT طراحی شده است. ساختار این مدل مانند BERT تنها یک بخش رمزگذار دارد و به طور مشابه، با روش یادگیری خود نظارتی^۶ آموزش داده می شود. به صورت خلاصه ایده اصلی این مدل کشف واحدهای گسسته و مخفی^۷ برای تبدیل داده های گفتاری به ساختاری "زبان مانند" است و بدین ترتیب امکان استفاده از مدل های قدرتمند موجود در حوزه پردازش زبان های طبیعی^۸ مانند BERT را بر روی داده های صوتی فراهم می کند. می توانید مقاله اصلی مدل HuBERT را از [این لینک](#) دریافت کنید.

¹ Speech Emotion Recognition

² Transformers

³ Long-range dependencies

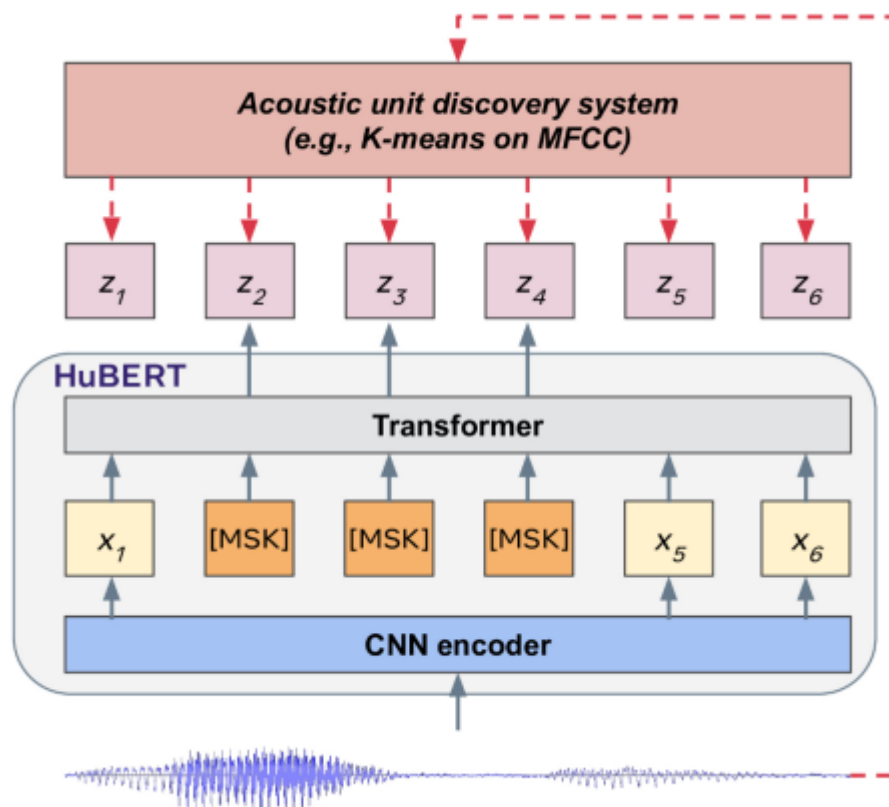
⁴ Attention mechanisms

⁵ Sequential data

⁶ Self-supervised learning

⁷ Hidden units

⁸ Natural language processing (NLP)



شکل ۱. شمایی کلی از فرایند پیش آموزش مدل **HuBERT**

۲-۱. سوالات تشریحی

۱-۲-۱. چالش های داده های صوتی در یادگیری

(۵ نمره)

داده های صوتی و متنی با اینکه هر دو در دسته داده های دنباله ای قرار می گیرند، اما با توجه به ویژگی های سیگنال صوتی (گفتاری)، چالش های منحصر به فردی برای یادگیری خود نظارتی^۱ در حوزه گفتار وجود دارد. دو تا از مهم ترین این چالش ها را بیان کنید.

۲-۲-۱. رویکرد HuBERT

(۵ نمره)

رویکردی که نویسندگان مقاله HuBERT برای مقابله با چالش های مطرح شده در سوال قبل در پیش گرفتند را بطور مختصر شرح دهید. همچنین توضیح دهید این مدل چگونه توانسته است بر پایه BERT که اصولاً در حوزه پردازش زبان های طبیعی به کار گرفته می شود، طراحی شود.

۳-۱. معرفی مجموعه دادگان

مجموعه داده مورد استفاده برای این بخش ShEMO (دادگان گفتار احساسی شریف) نام دارد که توسط مقاله ای با همین عنوان منتشر شده است. این پایگاه داده شامل ۳۰۰۰ فایل صوتی از داده های گفتاری است که از نمایشنامه های رادیویی آنلاین استخراج شده است. این مجموعه متشکل از نمونه های گفتاری ۸۷ فارسی زبان بومی و پنج احساس شامل خشم، ترس، شادی، غم و تعجب و حالت خنثی است که از طریق Kaggle^۲ قابل دسترس است.

۱-۳-۱. پیش پردازش داده ها

(۲۰ نمره)

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. همچنین توزیع دادگان در هر کلاس را نشان داده و در صورت بالانس

^۱ Supervised learning

^۲ <https://www.kaggle.com/datasets/mansourehk/shemo-persian-speech-emotion-detection-database>

نبودن آن‌ها، راهکاری برای حل آن ارائه کرده و آن را پیاده‌سازی کنید. برای آموزش، داده‌ها را با یک بذر تصادفی^۱ مشخص به سه بخش دادگان آموزش، اعتبارسنجی و ارزیابی (با نسبت 0.8، 0.1، 0.1) تقسیم نمایید. دقت کنید که توزیع مجموعه داده اصلی باید هنگام تقسیم بندی حفظ شود.

۱-۳-۲. ساخت دیتالودر

(۲۰ نمره)

همانطور که از قبل می‌دانید طول همه دنباله‌هایی که در یک Batch قرار می‌گیرند باید باهم برابر باشد. بدین منظور دیتالودر را طوری پیاده‌سازی نمایید که طول همه داده‌های **درون یک Batch** به اندازه طول بلندترین دنباله موجود در آن Batch باشد. این روش را با روشی که طول همه دنباله‌های درون دیتاست با پدینگ^۲ برابر با یک مقدار مشخص (max_length یا طول بلندترین دنباله درون دیتاست) قرار می‌دهد، مقایسه کرده و مزایا و معایب آنها را در گزارش خود ذکر کنید.

۱-۴. تعریف مسئله

در این سوال قرار است تا به مسئله تشخیص احساسات گفتار بپردازید و با استفاده از مدل مبتنی بر ترنسفورمر HuBERT (جهت تولید بازنمایی مناسب از سیگنال گفتار) شبکه عمیقی طراحی کرده و به کمک آن داده‌های گفتاری را بر حسب نوع احساسات به کاررفته در بیان، طبقه‌بندی نمایید. برای اینکار لازم است تا از مدل پایه HuBERT استفاده نمایید که در ^۳ Huggingface وجود دارد. به طور کلی برای توصیف احساسات دو رویکرد وجود دارد: روش اول بازنمایی ابعادی است که در آن احساسات در ابعاد مختلف و با ظرفیت‌های متفاوت نمایش داده می‌شوند. روش دوم نیز طبقه‌بندی گسسته است که در آن برای توصیف احساسات از یک برچسب واحد، مانند خشم، شادی، غم و... استفاده می‌شود. بدیهی است که بازنمایی ابعادی دقیق‌تر بوده و اطلاعات بیشتری را ارائه می‌دهد. اما در این تمرین از رویکرد طبقه‌بندی گسسته استفاده خواهیم کرد.

¹ Random seed

² Padding

³ <https://huggingface.co/facebook/hubert-base-ls960>

۱-۴-۱. تولید بازنمایی مناسب از کل دنباله ورودی

(۲۰ نمره)

بر خلاف مدل های مبتنی بر ترنسفورمر متنی مثل BERT ، HuBERT یک بردار تعبیه که بازنمایی کل دنباله ورودی باشد ایجاد نمی کند و لازم است تا به نحوی این بردار تولید شود. در این بخش از شما می خواهیم که روشی مناسب برای ساخت چنین برداری ارائه داده و آن را توجیه و پیاده سازی نمایید.

۱-۴-۲. آموزش مدل

(۳۰ نمره)

در این بخش باید با استفاده از روشی که در بخش قبل جهت ساخت بازنمایی مناسب از کل ورودی ارائه کردید یک مدل کامل برای طبقه بندی داده ها پیاده سازی کنید و فرایند آموزش را بطور کامل انجام داده و نتایج را گزارش نمایید . (بدین منظور لازم است تا جدول حاوی مقادیر انتخاب شده برای پارامترهای آموزش ، گزارش طبقه بندی، نمودار های دقت و loss و ماتریس درهم ریختگی^۱ را گزارش کنید). توجه داشته باشید که تحلیل نتایج حاصل و ارائه نمودار های دقیق در این بخش، اهمیت ویژه ای دارد.

نکات پیاده سازی

در نظر داشته باشید که تنها مجاز به استفاده از دو ابزار ارائه شده توسط Huggingface که در زیر اعلام می کنیم، خواهید بود و به جز آنها سایر موارد را باید خودتان پیاده سازی کنید.

- بارگذاری مدل ترنسفورمری: بدین منظور تنها مجاز به استفاده از کلاس های AutoModel و یا HubertModel متعلق به کتابخانه transformers می باشید.
- ساخت و یا بارگذاری استخراج گر ویژگی: برای اینکار می توانید از کلاس FeatureExtractor2Vec2Wav متعلق به کتابخانه transformers استفاده نمایید.

در نهایت به این نکته توجه کنید که مدل HuBERT شامل پشته ای از لایه های کانولوشنی است که به منظور استخراج ویژگی از سیگنال خام ورودی استفاده می شود. این بخش طی فرایند پیش آموزش تا حد مناسبی آموزش دیده و fine-tune کردن آن تفاوت چشمگیری در عملکرد مدل ایجاد نمی کند. لذا برای

¹ Confusion matrix

تسريع فرايند آموزش مي توانيد محاسبه گراديان براي اين بخش را غير فعال كنيد تا پارامتر هاي آن در طول آموزش به روز نشوند.

پرسش ۲. تنظیم دقیق مدل BERT

۲-۱. آموزش و تحلیل مدل BERT

در این تمرین با مدل ترنسفورمری BERT آشنا می‌شویم و به تنظیم دقیق^۱ و تحلیل و بررسی آن می‌پردازیم. در این راستا آزمایشات مختلفی بر روی وظیفه استنتاج زبان طبیعی^۲ طراحی شده‌اند که در ادامه به آن‌ها می‌پردازیم. در تمامی مراحل از مدل ParsBERT^۳ برای انجام وظایف بهره ببرید.

۲-۱-۱. پیش پردازش داده ها

(۱۰ نمره)

برای تمامی وظایف بیان شده از مجموعه داده FarsTail^۴ استفاده نمایید. اطلاعات آماری شامل توزیع کلاس ها و طول جملات (بر اساس تعداد توکن‌ها) این مجموعه داده را ارائه دهید و روش‌های پیش‌پردازش خود را شرح دهید.

۲-۱-۲. تنظیم دقیق مدل (fine tune)

(۲۰ نمره)

در ابتدا لازم است مدل را بصورت کامل بر روی مجموعه داده مورد نظر تنظیم دقیق نمایید. نتایج حاصل از تنظیم دقیق مدل را بروی روی دادگان تست را در قالب دقت^۵، امتیاز F1^۶ و ماتریس آشفتگی^۷ گزارش دهید. همچنین نمودارهای تغییرات خطا و دقت در هنگام آموزش را بروی مجموعه دادگان ارزشیابی رسم نمایید.

^۱ Fine tune

^۲ natural language inference

^۳ <https://huggingface.co/HooshvareLab/bert-base-parsbert-uncased>

^۴ <https://github.com/dml-qom/FarsTail>

^۵ accuracy

^۶ F1 score

^۷ Confusion matrix

۳-۱-۲. فریز کردن لایه‌های مدل

(۴۰ نمره)

با الهام از نتایج ارائه شده در مقاله [1] و رویکردهای مبتنی بر یادگیری انتقالی، در این قسمت به فریز کردن لایه‌های مختلف مدل BERT می‌پردازیم. بدین منظور لازم است مدل BERT را از طریق دو رویکرد زیر تنظیم دقیق نمایید.

- فریز کردن ۹ لایه ابتدایی مدل همانند آنچه در مقاله ذکر شده است.
- فریز کردن تمامی لایه‌ها بجز لایه آخر و لایه embedding در مدل

مدل BERT را از طریق رویکردهای بیان شده آموزش دهید و نتایج حاصل از تنظیم دقیق مدل‌ها را بروی روی دادگان تست را در قالب دقت، امتیاز F1 و ماتریس آشفتگی گزارش دهید. همچنین نمودارهای تغییرات خطا و دقت را بروی مجموعه دادگان ارزشیابی رسم نمایید.

۴-۱-۲. تنظیم دقیق مدل بر روی لایه‌های میانی

(۱۵ نمره)

علاوه بر استراتژی‌های فریز کردن لایه‌های مدل، در این بخش از تمرین به تأثیر تنظیم دقیق مدل تنها بروی زیرمجموعه‌ای از لایه‌های BERT می‌پردازیم و آن را بررسی می‌کنیم. این رویکرد می‌تواند بینش‌هایی را در مورد سهم لایه‌های مختلف در عملکرد مدل در وظایف خاص ارائه دهد و نشان دهد برای آموزش مدل برای وظایف مختلف آیا ارزش استفاده از تمام مدل برای تنظیم دقیق و آموزش وجود دارد و یا قسمتی کوچکتری از مدل قادر به حل مسئله است. بدین جهت در این قسمت به آموزش و تنظیم دقیق مدل تنها با استفاده از ۹ لایه ابتدایی آن می‌پردازیم. (ما بقی لایه‌ها حذف گردد) پس از آموزش مدل اصلاح شده بر روی دادگان آموزشی، نتایج حاصله را بروی روی دادگان تست همانند قسمت‌های پیشین گزارش نمایید. توضیح دهید آیا توانایی مدل در حل مسئله تغییری داشته است، برداشت خود را بنویسید.

۵-۱-۲. حذف head های attention در مدل

(۱۵ نمره)

در مقاله [2] نشان داده شده است که حذف head های attention پس از پیش آموزش مدل برای تنظیم دقیق مدل بروی وظیفه خاص نهایی می‌تواند بدون افت چشمگیر عملکرد مدل صورت پذیرد. بدین جهت

در این قسمت از تمرین لازم است ۵۰ درصد از head های attention در هرلایه را بصورت رندم حذف نموده و سپس مدل را بروی وظیفه مربوطه آموزش دهید. نتایج حاصل از آموزش اصلاح شده را همانند گذشته ارائه دهید.

(برای حذف head های attention در پایتورچ، مدل متدی با نام `prune_heads` دارد که می توانید از آن استفاده نمایید.)

دقت کنید در تمامی مراحل، نتایج ارزیابی بصورت کامل بیان شده باشد و در نهایت لازم است نتایج مراحل مختلف را در یک جدول در انتهای گزارش خود گردآوری نمایید و استدلال خود را نتایج حاصله بصورت کامل بیان کنید.