

۱. (۱۰٪) [مطالعه و بررسی]

۱-۱ توضیح دهید الگوریتم t-SNE چیست و مثالی از کاربردهای آن بیاورید.

۲-۱ فرض کنید مجموعه داده‌ای از توپیت‌ها که شامل عکس‌هایی

به همراه متن می‌باشد، در اختیار داریم. قصد بر این است که

مفهوم Hate Detection را مدل کنیم. برای این کار نیاز

داریم تا شبکه‌ای طراحی کنیم که رابطه میان عکس و متن را

یافته، به نحوی که تشخیص دهد هر داده به خصوص (شامل

عکس و متن) دارای مضامین نژادپرستی و یا جنیست زده

می‌باشد یا خیر. عنوانی از عکس را که مفهومی نژادپرستانه

دارد در زیر مشاهده می‌کنید. الف) نخست توضیح دهید چه

چالش‌هایی برای حل و مدل کردن این مسئله وجود دارد. ب)

یک معماری پیشنهاد دهید که کار تشخیص این الگو را انجام

دهد، با جزئیات توضیح دهید.



۲. (۴۵٪) [پایاده‌سازی: شبکه عصبی بازگشتی] در این پرسش قصد داریم با استفاده از شبکه عصبی

بازگشتی، زیرنویس مناسبی برای تصاویر مورد نظر تولید کنیم. عنوان نویسی تصویر که از آن با عناوینی

همچون شرح نویسی و یا زیرنویس تصویر هم یاد می‌شود، توصیف محتوای تصویر با استفاده از نمایش

متنی است. در واقع با توجه به موجودیت‌های مشاهده شده در تصویر و همچنین درک صحنه و ارتباط

بین بخش‌های مختلف آن، به توصیف خودکار تصویر پرداخته و این توصیف را در قالب جمله‌ای متناسب

با معیارهای زبان برمی‌گرداند. این موضوع که ترکیبی از دانش بصری کامپیوتری و پردازش زبان طبیعی

می‌باشد، اخیراً به یکی از مهمترین مباحث در بینایی ماشین تبدیل شده است. رویکردهای مختلفی برای

این کار پیشنهاد شده است که در بین آنها مدل‌ها مبتنی بر یادگیری عمیق به عنوان پیشرفته‌ترین مدل‌های

ثابت شده‌اند. اصولاً عنوان نویسی مبتنی بر یادگیری عمیق به مدل‌های رمزگشا-رمزگذار متکی است.

روش پیشنهادی برای این سوال نیز، متکی بر همین روش می‌باشد.

توضیحات کد و داده: برای این تمرین از مجموعه داده COCO Caption^۱ که در سال ۲۰۱۴ منتشر

شده است استفاده می‌کنیم، که شامل ۸۰۰۰۰ تصویر آموزشی و ۴۰۰۰۰ تصویر اعتبارسنجی است. داده‌ها

¹ <https://cocodataset.org>



از قبل برای شما پیش‌پردازش و در یک فایل ذخیره شده‌اند که شامل ۱۰۰۰۰ جفت تصویر-متن برای آموزش و ۵۰۰ جفت برای آزمایش می‌باشد. برای پردازش متن‌ها نیز به هر کلمه یک شناسه عدد صحیح اختصاص داده شده است که این امکان را می‌دهد که یک عنوان را با دنباله‌ای از اعداد صحیح نشان دهیم. نگاشت بین شناسه‌های اعداد صحیح و کلمات در ورودی به نام «vocab» (هر دو «idx_to_token» و «token_to_idx») ذخیره می‌شود و از تابع «decode_captions» از «a5_helper.py» برای تبدیل تنسورهای شناسه‌های صحیح به رشته‌ها استفاده می‌کنیم.

به طور کلی، ابتدا این شبکه ویژگی‌های تصویر را استخراج کرده و به عنوان ورودی لایه مخفی به شبکه RNN می‌دهد و کلمه به کلمه جمله‌ای برای توصیف تصویر ساخته می‌شود. مراحل این عملیات در غالب توابع سازنده شبکه RNN توسط شما پیاده سازی خواهند شد. توابعی که توسط شما پیاده سازی خواهند شد شامل موارد زیر می‌شوند:

- ۱- تابع rnn_step_forward که گامی از عملیات پیشخور و تابع rnn_forward که تمامی دنباله پیشخور به شبکه را مدیریت می‌کنند.
- ۲- تابع rnn_step_backward که گامی از عملیات پس انتشار خطا و تابع rnn_backward که تمامی دنباله پس انتشار خطا را در شبکه را مدیریت می‌کنند.
- ۳- تابع WordEmbedding که وظیفه آن تعبیه سازی هر کلمه در یک بردار عددی در فضای چندبعدی (Embedding Space) است. این نمایش برداری به مدل کمک می‌کند تا مفاهیم معنایی یا ارتباطات بین کلمات را بهتر درک کند.
- ۴- تابع temporal_softmax_loss که وظیفه آن محاسبه‌ی خطای پیش‌بینی مدل در طول توالی‌های زمانی است، جایی که هر عنصر توالی یک پیش‌بینی برای یک کلمه از واژگان است.
- ۵- کلاس CaptioningRNN که در برگیرنده تمامی توابع قبلی است و در آن caption برای تصاویر تولید می‌شود.

برای پیش‌برد تمرین و تکمیل کدهای مورد نظر به شکل زیر عمل کنید:

- گام اول: هنگامی که محتوای دانلود شده را از حالت فشرده خارج کردید، لطفاً پوشه را در Google Drive خود آپلود کنید. سپس، با کلیک راست روی فایل *.ipynb، آن را با Google Colab باز کنید. توصیه می‌کنیم فایل *.py خود را در Google Colab ویرایش کنید. توجه شود، نوت بوک rnn_captioning.ipynb و فایل rnn_captioning.py را در کنار هم بگذارید (به طور کلی به ساختار قرار گیری فایل‌هایی که در اختیار شما قرار داده شده است، دست نزنید!)
- گام دوم: سلول‌های نوت بوک وابستگی اجرایی به تکمیل کدهای داخل فایل *.py دارند و برای اجرای هر سلول، توابعی داخل فایل *.py نیازمند تکمیل می‌باشند. در انتها نتیجه کار خود را در هر دو فایل *.ipynb و *.py، در Google Drive ذخیره کنید (روی «فایل» - «ذخیره» کلیک کنید) و کل فولدري که در اختیار شما قرار داده شده است حاوی نوت بوک و فایل *.py.



ای که توسط شما به روزرسانی شده اند را برای ما ارسال کنید.

هنگام کار موارد زیر را در نظر داشته باشید:

- به کدهای داخل فایل ipynb. نباید دست بزنید و فقط کدهای فایل py. باید توسط شما تکمیل شوند.
- در فایل py. بلوک‌هایی برای نوشتن کد توسط شما مشخص شده است. هیچ کدی را خارج از این بلوک‌ها ننویسید یا تغییر ندهید.
- سلول‌های نوت بوک را نیز اضافه یا حذف نکنید.

نکته: قبل از ارسال، همه سلول‌های فایل ipynb. را اجرا کنید و خروجی‌ها را پاک نکنید.

۳. (۴۵٪) [پیاده‌سازی: ساز و کار توجه و مبدل‌ها]

تحلیل احساس یکی از زیرشاخه‌های پردازش زبان طبیعی است که هدف آن طبقه‌بندی متون با توجه به احساس، عقیده و نگرش بیان شده در آنها است. در دهه‌های گذشته روی استفاده از رویکردهای یادگیری ماشین جهت حل مسائل تحلیل احساس کارهای زیادی انجام گرفت، اما تمرکز اصلی آنها بر روی ساخت استخراج کننده‌های ویژگی قویتر بوده است، زیرا که عملکرد این یادگیرنده‌ها بیشتر به انتخاب نحوه بازنمایی داده وابسته است.

مدل‌های شبکه عصبی بازگشتی مطرح شده، در مدل کردن داده‌های متوالی برتری دارند. اما این مدل‌ها دو اشکال عمده دارند: نخست، مشکل ناپدید شدن گرادپان برای دنباله‌های طولانی و دوم زمان زیاد آموزش مدل.

مسئله اول توسط معماری‌های شبکه عصبی بازگشتی (مشابه LSTMها، GRUها) حل می‌شود، اما حل مشکل دوم با این راهکار امکان‌پذیر نمی‌باشد. مبدل‌ها این مشکلات را تا بخشی با امکان پردازش ورودی به صورت موازی حل می‌کنند. اگرچه محاسبات با توجه به طول توالی ورودی درجه دوم است، اما با GPUهای مدرن قابل مدیریت است.

در این بخش مدل مبدل را به صورت گام به گام با ارجاع به مقاله اصلی^۱ پیاده‌سازی می‌کنیم. مبدل‌ها از یک رمزگذار و یک رمزگشا تشکیل شده‌اند. در این بخش یاد می‌گیرید که چگونه یک مبدل رمزگذار را به صورت گام به گام پیاده‌سازی کنید. با توجه به کد، بخش‌های زیر را پیاده‌سازی خواهید کرد و در نهایت بر روی مجموعه داده مشخص شده آموزش می‌دهید.

مجموعه داده استفاده شده در این قسمت شامل داده‌های دادگان sst2 از معیار سنجش glue می‌باشد. این دادگان شامل داده‌های تحلیلی احساسات معیار glue بوده و در این تمرین با هدف تست مدل مبدل استفاده خواهد شد. این دادگان شامل ۶۷۳۴۹ داده آموزش، ۸۷۲ داده اعتبار سنجی و ۱۸۲۱ داده آزمون است.



در این تمرین شما ماژول‌های مورد نیاز مدل مبدل را در فایل `py` تکمیل کرده و بلوک‌های سازنده اصلی این مدل را خواهید ساخت. این بلوک‌ها شامل موارد زیر می‌شوند:

۱. توابع `scaled_dot_product_two_loop_single`، `scaled_dot_product_no_loop_batch` و `caled_dot_product_two_loop_batch` که عملیات ضرب داخلی را در مکانیزم توجه پشتیبانی می‌کنند و به ترتیب گفته شده بهبودی بر یکدیگر هستند.
۲. کلاس `MultiHeadAttention` که با دربر گرفتن شی‌هایی از کلاس `SelfAttention` و با استفاده از تابع `scaled_dot_product_no_loop_batch`، ساختار توجه را در مدل مدیریت می‌کند.
۳. کلاس `LayerNormalization` که در آن فرایند نرمال‌سازی را برای هر عنصر و در فضای ویژگی آن عنصر انجام می‌دهد.
۴. کلاس `FeedForwardBlock` که عملیات پیشخور در بلوک‌های مبدل را انجام می‌دهد.
۵. کلاس `EncoderBlock` که تمامی کلاس‌های تعریف شده در گذشته را در راستای ساختن یک بلوک رمزگذار در مبدل در بر می‌گیرد.
۶. تابع `position_encoding_simple` که رمزگذاری مکانی را برای دنباله‌های متنی انجام می‌دهد.
۷. کلاس `Transformer_encoder` که هسته سازنده مدل مبدل است و با دربرگرفتن کلاس‌های سازنده رمزگذارها و همچنین مدیریت فرآیند عبور داده‌ها از مدل مبدل، پیش‌بینی احساسات را برای داده‌های دادگان `sst2` انجام می‌دهد.

برای پیش‌برد تمرین و تکمیل کدهای مورد نظر به شکل توضیح داده شده در سوال قبل عمل کنید.

۴ (۳۰٪ نمره اضافی: اختیاری) [پیاده‌سازی شبکه GAN]

در این پرسش به پیاده‌سازی یک شبکه مولد تخصصی `GAN` می‌پردازیم. هدف ما در این تمرین بررسی بخش‌های مختلف شبکه‌های مولد و یادگیری مفهوم و درک توانایی این شبکه می‌باشد. مجموعه دادگان این تمرین، تصاویر صورت گربه‌ها هستند.

- **نکته:** در این تمرین مجاز به استفاده از کتابخانه‌های آماده هستید.
- **نکته:** در این تمرین تنها بخش‌هایی که ناقص هستند باید توسط دانشجویان تکمیل شود. این بخش‌ها در کد که در پیوست ارسال شده مشخص شده‌اند.



(۱) نمونه ای از مجموعه دادگان تصاویر صورت گربه ها^۲

نکته: در این تمرین برای بررسی هر بخش، قسمت هایی به عنوان تست قرار داده شده است که اگر کد نوشته شده مشکل نداشته باشد، خروجی تست 'OK' را چاپ می کند.

۴-۱ [۵٪] [تولید توزیع گوسی] در بسیاری از مسائل واقعی، داده های ما توزیع های پیچیده ای دارند و GANها ابزاری هستند که می توانند این توزیع ها را مدل سازی کرده و داده های جدید تولید کنند. در این بخش قصد داریم یک توزیع گوسی (نرمال) یک بعدی را از یک توزیع یکنواخت تولید کنیم. هدف اصلی این بخش آشنا کردن شما با ساختار اولیه یک شبکه مولد متخاصم (GAN) و گام هایی است که برای پیاده سازی این مدل باید طی شود. این تمرین به شما کمک می کند تا مفاهیم اساسی شبکه های مولد را به صورتی بنیادین درک کنید.

در این بخش شما باید قسمت های مختلف شبکه GAN شامل بخش های:

۱- متمایز کننده (Discriminator)

۲- مولد (Generator)

۳- تابع آموزش شبکه و توابع مکمل آن

را تکمیل کنید.

۴-۲ [۲۰٪] [تولید تصویر صورت گربه] در این بخش با برداشتن گامی فراتر، هدف ما تولید تصاویر

^۲ <https://www.kaggle.com/datasets/borhanitrash/cat-dataset>



چهره گربه از بردار نویز می‌باشد. با کامل کردن بخش‌های مشخص شده که شامل توابع مولد، متمایزکننده، ضرر برای هر دو قسمت شبکه و نهایتاً تابع آموزش، این شبکه خواهید ساخت. پارامترهای شبکه به صورت زیر هستند:

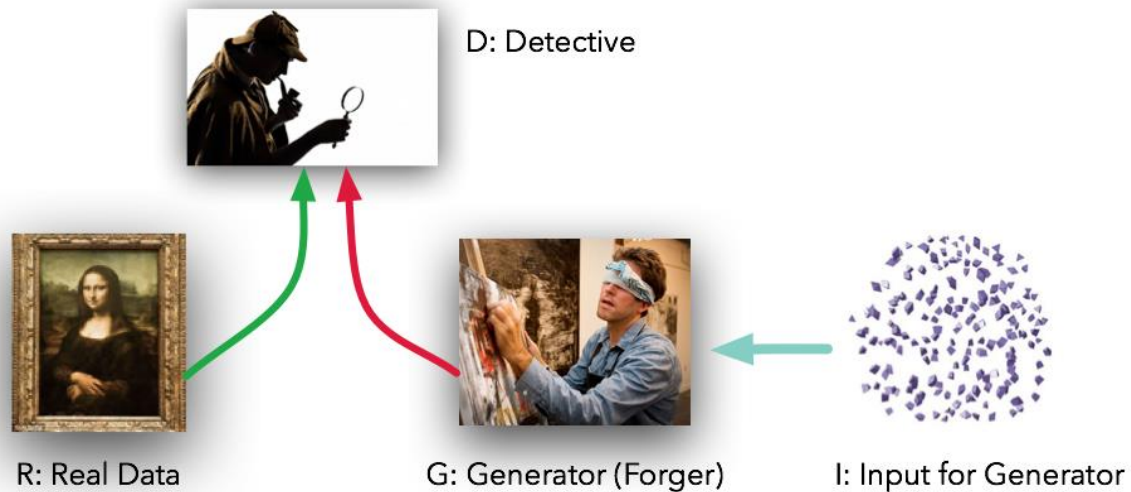
- ۱- اندازه بردار نویز برابر با ۱۰۰
- ۲- نرخ یادگیری برابر با ۰.۰۰۰۲
- ۳- ابعاد تصاویر ۲۵۶*۲۵۶ در فرمت RGB
- ۴- اندازه batch برابر با ۶۴
- ۵- تعداد epoch حداقل برابر با ۲۰ دور

۳-۴ [۵٪] [سوالات تکمیلی] به سوالات زیر پاسخ دهید.

- ۱-۳-۴ با توجه به پیاده سازی انجام شده، به نظر شما مشکلات این شبکه چیست؟
- ۲-۳-۴ چه راه کاری برای بهبود پیشنهاد می‌دهید؟ آیا تغییر پارامترها به تنهایی کفایت؟ در صورت مخالفت، راهکار پیشنهادی خود را بیان کنید.

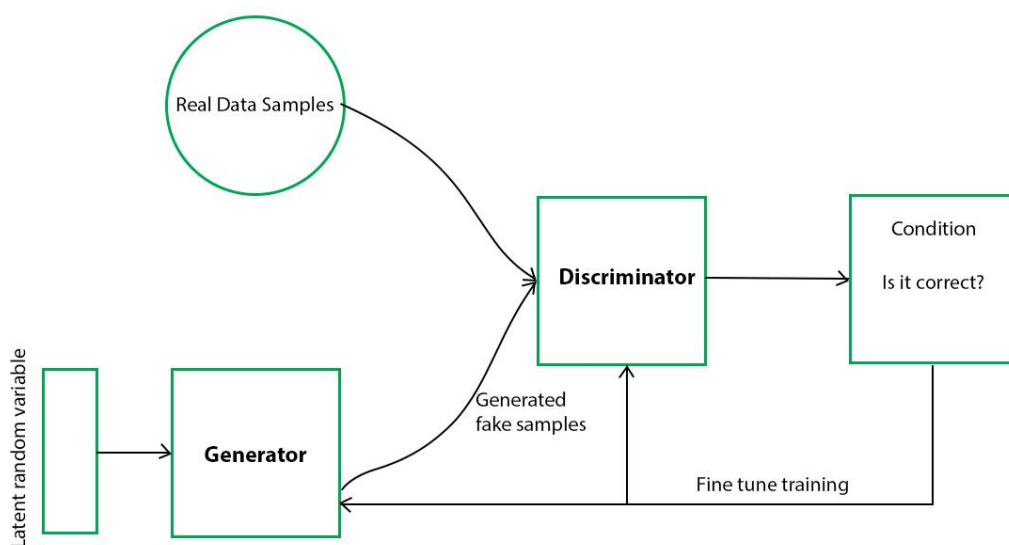
پیوست: مروری بر شبکه های مولد تخصصی

شبکه های مولد متخصص یا GAN دسته ای از شبکه های عصبی عمیق هستند که از دو جزء اصلی تشکیل شده اند: یک مولد و یک متمایزکننده. وظیفه مولد ایجاد داده های مصنوعی است، درحالی که نقش متمایزکننده تمایز میان داده های واقعی و تولیدشده است. این دو مؤلفه با هم در یک فرآیند رقابتی و تکراری شرکت می کنند و دائماً عملکرد خود را بهبود می بخشند و به تولید خروجی های واقعی تر می انجامند.



(۲) شما کلی شبکه مولد متخصص

همان طور که در تصویر زیر مشاهده می کنید، شبکه مولد (Generator) داده های مصنوعی را از یک نویز تصادفی تولید کرده و به شبکه تمایزدهنده (Discriminator) ارسال می کند. سپس، شبکه تمایزدهنده با ترکیبی از داده های واقعی و داده های مصنوعی تولیدشده توسط مولد آموزش داده می شود تا بتواند این دو را از هم تمیز دهد. برای واقعی تر شدن داده های تولیدی، شبکه مولد براساس بازخورد دریافتی از شبکه تمایزدهنده به روزرسانی می شود. این روند تکرار می شود تا زمانی که معیار همگرایی مورد نظر برآورده گردد و شبکه مولد بتواند داده هایی نزدیک به واقعیت تولید کند.



۳) معماری شبکه مولد متخاصم