پروژه

یادگیری عمیق - دانشکده مهندسی برق - زمستان ۱۴۰۱

پروژه درس یادگیری عمیق طراحی سیستم مولتی‌مودال برای تحلیل احساسات است. در این پروژه ابتدا با مجموعه دادگان این حوزه آشنا خواهید شد و سپس شروع به آموزش مدل‌هایی بر پایه شبکه عصبی برای تحلیل احساسات داده مولتی‌مودال (شامل متن, تصویر و صوت) خواهید کرد. مجموعه دادگان استفاده در این پروژه همگی به زبان انگلیسی هستند.

قواعد پروژه :

- راه ارتباطی با تیم پروژه تنها از طریق گروه درس در تلگرام و یا بخش پرسش و پاسخ کوئرا بوده و اعضای تیم پروژه به سوالات مستقیم پاسخ نخواهند داد.
- پروژه با احتساب فاز صفر مجموعا چهار فاز خواهد داشت و مجموعا ۶ روز تاخیر مجاز. پس از این مدت به ازای هر روز ۲ درصد از نمره بخش مربوطه از دست خواهد رفت. توجه بفرمایید در فاز چهارم امکان تاخیر وجود نداشته و پس از ددلاین این فاز، تحویل پروژه خواهید داشت.
- پس از ارسال کد هر فاز امکان ایجاد تغییر در کد خود برای فارهای بعدی پروژه را خواهید داشت اما ملاک ارزیابی هر فاز کد آپلود شده برای آن فاز می‌باشد نه کد ارائه شده در انتهای پروژه.
- آپلود پروژه از طریق کوئرا انجام می‌شود. برای راحتی دوستان در پروژه استفاده از GitHub اجباری نمی‌باشد اما توصیه اکید می‌شود به منظور مدیریت بهتر کار گروهی از ابزارهای مربوطه استفاده بفرمایید.

فاز یک

هدف اصلی این فاز تحلیل احساسات تصاویر با استفاده از تصاویر موجود در دیتاست می‌باشد. برای انجام این بخش مشابه فاز صفر پروژه از مجموعه دادگان به شرح زیر استفاده می‌کنید:

- MSCTD: A Multimodal Sentiment Chat Translation Dataset
    - Github: https://github.com/XL2248/MSCTD
    - Paper: https://aclanthology.org/2022.acl-long.186/

خروجی این فاز می‌بایست یک فایل ژوپیتر نوت‌بوک به همراه پارامترها و فایل‌هایی باشد که **امکان اجرای کد به صورت مستقل** را بدهد. در صورتی که حجم فایل‌هایی که باید همراه با نوت‌بوک خود ضمیمه کنید بالا می‌باشد و امکان آپلود در کوئرا نمی‌باشد، می‌توانید موارد را در گوگل درایو خود آپلود کرده و یک لینک دسترسی عمومی از آن دریافت کنید. سپس داخل کد موارد را از آن لینک دانلود بفرمایید تا امکان اجرا مجدد کد فراهم باشد.
همچنین نیازی به نوشتن گزارش مجزا نمی‌باشد و می‌توانید توضیحات به همراه نتایج خود را در فایل ژوپیتر نوت‌بوک خود قرار دهید.

توجه: علی رغم آنکه پیچیدگی مساله به نسبت پایین بوده و خروجی سه کلاسه می‌باشد، اما با توجه به آنکه در این فاز از مدالیته‌ی تصویر استفاده می‌کنیم و حجم اطلاعات کد شده در حوزه‌ی احساسات در متن و یا صحبت بیشتر از تصویر می‌باشد، ممکن است دقت شبکه به اندازه کافی چشم‌گیر نباشد. در نتیجه در این فاز طی کردن درست مسیر پیاده‌سازی اهمیت بالاتری نسبت به دقت نهایی خروجی خواهد داشت، البته که دقت مدل شما نیز ارزش و نمره خود را خواهد داشت.

• بخش اول - تحلیل چهره‌ها

در این بخش می‌خواهیم از چهره‌های موجود در تصاویر برای تشخیص احساسات استفاده کنیم.

- زیر بخش اول - پیاده سازی روش

- گام۱: در این بخش ابتدا با توجه به روشی که در فاز صفر، چهره‌ها را از تصویر استخراج کردید، باید مدلی پیاده سازی کنید تا با گرفتن تصویر اصلی به عنوان ورودی، مرزهای هر چهره را از تصویر استخراج کرده و مجموعه تصاویر جدیدی که هر تصویر شامل یک چهره می‌باشد به عنوان خروجی تحویل دهد.

نکته: صحت عملکرد این بخش از مدل بسیار مهم بوده و در صورتی که روش دقت قابل قبولی در تشخیص چهره‌ها نداشته باشد، در ورودی شبکه در گام‌های بعدی صحت کافی را نخواهد داشت. همچنین اگر شبکه استفاده شده در فاز صفر به اندازه کافی عملکرد خوبی نداشته باشد یا مرز چهره را به عنوان خروجی بر نمی‌گرداند، می‌توانید از شبکه‌ها و پیکیج‌های آماده دیگر مانند Facenet[1] یا هر شبکه‌ای که صحت مناسبی در این کار دارد استفاده نمایید.

- گام۲: حال باید یک شبکه مبتنی بر CNN طراحی کنید تا تصویر چهره را به عنوان ورودی بگیرد و لیبل سه کلاسه احساسات را به عنوان خروجی بدهد. توجه کنید که تصاویر چهره‌های جدا شده از بخش قبل ابعاد مختلفی خواهند داشت، و قبل از استفاده از آن‌ها به عنوان ورودی شبکه، باید ابعاد چهره‌ها را به یک ابعاد مشخص مدنظرتان اسکیل کنید. صحت خروجی شبکه‌ی خود را پس از آموزش بررسی کنید.

نکته: ممکن است شبکه تشخیص چهره شما برای بعضی از تصاویر دیتاست، چهره‌ای تشخیص ندهد، و یا برای بعضی دیگر چند چهره تشخیص دهد. توجه کنید که در این گام، برای آموزش شبکه تنها لازم است چهره‌های هر تصویر را به همراه لیبل تصویر مربوطه از دیتاست جدا کنید و به عنوان دیتاست آموزش شبکه استفاده نمایید.

- گام۳: اکنون دو شبکه تشخیص چهره و تشخیص احساسات از روی چهره را به یکدیگر متصل کنید. برای لیبل زدن تصاویری که دارای چند چهره هستند و در نتیجه ممکن است خروجی شبکه تشخیص احساسات برای چهره‌های مختلف، متفاوت باشد، می‌توانید از ترکیب لیبل‌ها به صورتی خلاقانه استفاده کنید تا نهایتا به یک لیبل برسید. (می‌توانید حتی برای ترکیب لیبل‌ها برای گرفتن لیبل خروجی نیز یک شبکه کوچک طراحی کنید!) همچنین برای حالتی که هیچ چهره‌ای در تصویر تشخیص داده نمی‌شود می‌توانید به صورت تصادفی و یا با توجه به آنالیزهای آماری که در فاز صفر بر روی دیتاست انجام داده اید، یک لیبل به عنوان خروجی بدهید و نهایتا دقت شبکه را در این تست سه کلاسه ارزیابی کنید.

[1] https://github.com/timesler/facenet-pytorch

- زیر بخش دوم - تغییر و افزایش دادگان

- گام۱: هدف از این بخش بررسی عملکرد شبکه زیر بخش قبل، که فقط روی تصاویر صورت دستکاری نشده آموزش دیده است، روی تصاویری از صورت است که تحت سه مورد از معمول‌ترین تبدیل‌های مورد استفاده برای تخریب تصویر، یعنی تبدیل‌های فرکانسی، مکانی و روشنایی، تغییر داده شده اند. لذا ابتدا نیاز است با استفاده از روش مذکور در این مقاله[2]، مجموعه تصاویر جدیدی از روی مجموعه تصاویر چهره ساخته شده در قسمت قبل ایجاد کنید.

- گام۲: حال با اعمال مجموعه تصاویر دستکاری شده به عنوان ورودی به شبکه زیر بخش قبل، مجددا عملکرد این شبکه را بررسی نمایید.

- گام۳: با ترکیب مجموعه تصاویر صورت و مجموعه تصاویر دستکاری شده، مدل جدیدی برای تشخیص احساس از روی چهره آموزش دهید. عملکرد شبکه جدید را بر روی هر یک از مجموعه تصاویر به طور جداگانه بررسی نمایید و تاثیر اضافه کردن تصاویر دستکاری شده در فرآیند آموزش را تحلیل کنید.

- بخش دوم - تحلیل احساسات با استفاده از ویژگی های تمام تصویر

در این بخش می‌خواهیم از کل تصویر برای تشخیص احساسات استفاده کنیم.

- گام۱: ابتدا یک شبکه معروف در حوزه پردازش تصویر را به دلخواه انتخاب کنید. می‌توانید از شبکه‌های موجود در این لیست و یا هر شبکه دلخواه دیگری استفاده نمایید.

- گام۲: اکنون از لایه‌های انتهایی شبکه که ویژگی‌هایی با بالاترین سطح اطلاعات از تصویر هستند، خروجی بگیرید و با استفاده از آن به عنوان دیتاست آموزش، یک شبکه چند لایه با طراحی دلخواه با خروجی سه کلاسه احساسات را آموزش دهید. علت استفاده از یک شبکه‌ی آموزش دیده حوزه تصویر آن است که برای استخراج ویژگی‌های مناسب از تصویر، یک شبکه به آموزش بسیار زیادی احتیاج دارد و ممکن است در نهایت نیز به دقت و پایداری این شبکه‌ها نرسد.

- گام۳: حال شبکه خود را آموزش دهید. توجه کنید در هنگام آموزش وزن‌های شبکه اصلی را که به عنوان شبکه استخراج ویژگی از تصویر استفاده کرده‌اید، آپدیت نکنید. دقت ترکیب حاصل را ارزیابی کنید. آیا این روش برای تسک مورد نظر جوابگو است؟ دقت روش پیاده‌سازی شده در کدام بخش بالاتر است؟

[2] https://arxiv.org/pdf/2112.13547

- بخش سوم - تحلیل احساسات با ترکیب دو روش

اکنون روشی مناسب برای ترکیب دو روش قبل پیاده‌سازی کنید که تعداد چهره موجود در هر تصویر و خروجی دو روش قبلی را به عنوان ورودی بگیرد و لیبل احساس را به عنوان خروجی برگرداند. در این بخش می‌توانید ایده‌های خلاقانه‌ای به منظور افزایش دقت انجام دهید. به طور مثال می‌توانید روی حالتی که تصویری در شبکه بخش اول شناسایی نمی‌شود تنها از نتیجه‌ی بخش دوم استفاده کنید یا در حالتی که چندین تصویر شناسایی می‌شود اهمیت خروجی لیبل‌های بخش چهره را نسبت به خروجی بخش دوم به طریقی در روش خود بالاتر ببرید.