

به نام خدا

پروژه پایانی درس مدل‌های زبانی بزرگ ۱۴۰۳-۱۴۰۴

تاریخ های مهم:

۵ خرداد: شروع پروژه

۱۰ تیر: ارائه مدل و گزارش متنی پایانی پروژه

نمای کلی

هدف از این پروژه آشنایی و کار با مدل‌های بزرگ چند مدالیت (Large Multimodal Models) می باشد. در این پروژه با یک مدل زبانی-تصویری کار کرده و و روی پایگاه داده خاصی این مدل را آموزش و بهبود می دهید و تست می کنید.

گروه بندی

این پروژه را می توانید به صورت انفرادی یا در گروه های ۲ یا ۳ نفره انجام دهید. لازم است همه اعضای گروه در پروژه به صورت فعال مشارکت داشته باشند اما انتخاب اعضای گروه و چگونگی تقسیم وظایف بر عهده خودتان می باشد.

داده ها

در این پروژه روی پایگاه داده IP102 که مربوط به طبقه بندی انواع آفات گیاهی است کار خواهید کرد. این پایگاه داده از ۱۰۲ کلاس مختلف تشکیل شده و پژوهش های زیادی برای طبقه بندی این پایگاه داده انجام شده است. چند نمونه از مقالات اخیر منتشر شده در این زمینه به صورت زیر است که توانسته اند به دقت ۷۰-۸۰ درصد روی این پایگاه داده دست یابند:

ROI-Aware Multiscale Cross-Attention Vision Transformer for Pest Image Identification

Pest-ConFormer: A hybrid CNN-Transformer architecture for large-scale multi-class crop pest recognition

Deep-Wide Learning Assistance for Insect Pest Classification

داده های کلاس های مختلف از نظر تعداد تعادل ندارند و برای نتیجه گیری بهتر می توان از روش های data augmentation استفاده کرد.

کسب اطلاعات بیشتر و دانلود این پایگاه داده از [اینجا](#) امکان پذیر است.

مدل

یک مدل زبانی-تصویری دلخواه انتخاب کنید. می توانید از مدلهایی که در درس دیده ایم (CLIP, BLIP, OpenFlamingo, miniGPT4,...) استفاده کنید یا از مدلهای موجود دیگر (مثلا ... , ALBEF, GIT, LAVIS). با توجه به امکانات محدود پردازشی توجه کنید که مدل انتخابی بیش از حد بزرگ نباشد و حتما در همان ابتدا نیازمندی های پردازشی مدل را بررسی کنید. مدلهایی که بولد شده اند احتمالا از نظر محاسباتی مناسب تر باشند. می خواهیم از این مدل برای انجام عمل طبقه بندی عکس های پایگاه داده بالا استفاده کنیم. برای کار کردن با مدل انتخابی از صفحه [github](#) هر مدل استفاده کنید.

دستورالعمل

- ۱- با استفاده از مدل انتخابی خود و در صورت نیاز با تغییر جزئی ساختار مدل، عمل طبقه بندی پایگاه داده را انجام دهید و نتایج را گزارش کنید. برای این کار نام کلاس ها را به صورت یک پرامپت درآورده و شباهت بردارهای **embedding** تصاویر و این پرامپتها را محاسبه کرده و عمل طبقه بندی را انجام دهید.
- ۲- مدل را با این پایگاه داده **finetune** کنید. اگر این امکان وجود ندارد بخشی از مدل را ریزتنظیم کنید (مثلا **encoder** ها را فریز کنید) و یا از روشهایی مثل **LoRA** استفاده کنید یا می توانید فقط چند لایه به انتهای **encoder** تصویر اضافه کنید. همچنین می توانید از ویژگی های به دست آمده از مدل برای عمل طبقه بندی استفاده کنید، یعنی یک مدل طبقه بندی آموزش دهید که بر اساس ویژگی های بدست آمده از مدل، عمل طبقه بندی را انجام دهد. هم چنین می توانید رزولوشن تصاویر را کاهش دهید. نتایج طبقه بندی روی مدل جدید (روی داده های تست که در ریز تنظیم استفاده نشده اند) را محاسبه و گزارش کنید.
- ۳- بررسی کنید تغییر پرامپت چه تاثیری روی نتایج دارد؟ آیا می توانید پرامپتی پیشنهاد کنید که نتایج را بهبود بخشد؟
- ۴- تصاویری که برای آنها خطای مدل زیاد است را بررسی کرده و دلیل این خطای زیاد را توجیه کنید. چه ایده ای دارید که این خطا کاهش یابد؟ در صورت امکان آن را امتحان کنید و نتایج جدید را ارائه کنید (**error analysis**).
- ۵- برای بهبود عملکرد مدل غیر از موارد بالا، ایده دیگری دارید؟ آیا امکان پیاده سازی آن وجود دارد؟ چه تغییراتی لازم است روی ساختار مدل یا داده اعمال کنیم؟

نتایج

تا جای امکان معیار های مختلف مربوط به طبقه بندی را روی داده های خود محاسبه و ارائه کنید. بررسی کنید برای چه گروه هایی نتایج بهتر و برای چه گروه هایی نتایج ضعیف تر است و این مشاهده را توجیه کنید.

چه چیزی را تحویل دهیم؟

هم کد و هم گزارشی از راه حل و نتایج خود را تحویل دهید.

گزارش باید حاوی موارد زیر باشد:

- ۱- توضیح مختصری از راه حل پیاده سازی شده شما، با تمرکز ویژه بر بخش های "غیر پیش پا افتاده" یا جالب تر راه حل. برخی از مشکلات و محدودیت های روش شما چیست و دلایل احتمالی آنها چیست؟ از توضیح مطالب واضح و غیرضروری خودداری کنید. طول گزارش ارتباطی با نمره دریافتی ندارد!
- ۲- نتایج درخواست شده در همه بخش ها را به صورت واضح و مرتب با ذکر شماره بخش ارائه کنید.
- ۳- اگر ایده بیشتری برای این پروژه اجرا کرده اید در گزارش خود به طور واضح و مشخص بیان کنید.
- ۴- گزارش به عنوان معیار همه کارهای انجام شده در نظر گرفته می شود. چنانچه ایده ای را پیاده سازی کرده اید حتما باید در گزارش خود به آن اشاره کنید (کد ارسالی فقط برای چک کلی استفاده می شود و نه اجرای ایده های دیگر)

دستورالعمل ارسال

فایل های زیر را روی سامانه یکتا آپلود کنید:

- ۱- کد: نام فایل باید lastname_firstname_p (کدهای لازم برای اجرا) باشد.
- ۲- گزارش مختصر در یک فایل PDF با تمام نتایج: نام فایل باید lastname_firstname_p.pdf باشد.