

فاز دوم پروژه هوش محاسباتی پاییز ۱۴۰۲

مقدمه

در این فاز، هدف آشنایی و پیاده سازی با روش های تشخیص اهمیت ویژگی ها، و انتخاب ویژگی های برتر، در یادگیری ماشین است.

برای این منظور، شما در ادامه فاز پیشین، ابتدا باید در هر تصویر، مهم ترین نواحی را تشخیص دهید. ناحیه مهم، بخشی از تصویر است که شما در فاز پیشین جدا کرده اید، و بیشترین اطلاعات را در خصوص تشخیص کلاس تصویر به ما میدهد.

در ادامه، قصد داریم به نحوی، مهم ترین ویژگی های استخراج شده از تصاویر را پیدا کنیم. در صورتی که در فاز پیشین موفق به استخراج نواحی از تصویر نشده اید، و یا به کد فاز پیشین خود شک دارید، می توانید از این کد نیز استفاده کنید:

از slic از لایبری skimage استفاده کنید، که نمونه استفاده آن در توضیح زیر قرار گرفته است:
<https://www.geeksforgeeks.org/image-segmentation-using-pythons-scikit-image-module/>

بخش اول

در بخش اول پروژه، قصد داریم نواحی ای را پیدا کنیم که بیشترین اطلاعات را در خصوص طبقه بندی تصویر بهما می دهند. برای انجام این مهم، ابتدا شما باید الگوریتم کد پیشین را به نحوی تغییر دهید که:

- برای طبقه بندی، بجای استفاده از روش گفته شده، پس از استخراج ویژگی از ناحیه ها، میانگین ناحیه ها را بدست آورده، و با استفاده از آن عمل کند.

- سپس به ازای هر تصویر، قصد دارید ناحیه های مهم را به ترتیب پیدا کنید. با توجه به زمان بر بودن این مرحله، پیشنهاد می شود که به صورت رندم، بخش کوچکی از پایگاه داده را انتخاب کرده، و برروی آن پیاده سازی کنید. در هر مرحله، قصد داریم در تمامی تصاویر پایگاه داده، ناحیه با کمترین اهمیت را پیدا کنیم، و آن را حذف کنیم (به عبارت دیگر، در میانگین اعمال نشود). برای انجام این مهم، به ازای هر تصویر، به ازای ناحیه های باقی مانده، هر کدام را یک بار حذف کرده و ببینید کدام ناحیه، پس از حذف، کمترین تاثیر را در خراب کردن تشخیص درست مدل (یا بیشترین تاثیر در بهبود تشخیص درست مدل) را دارد. آن ناحیه را حذف کنید و رتبه آن را، از اخر اختصاص دهید. برای تمامی تصاویر دیتاست این کار را انجام دهید. پس از حذف یک ناحیه از تمامی تصاویر، طبقه بند را با تصاویر بعد از حذف ناحیه (تصاویری که حداقل یک ناحیه دارند) آموزش داده، و این کار را تکرار کنید. تا به ازای تمامی تصاویر، به رتبه بندی نواحی آنها برسید.

به صورت کیفی، بررسی کنید در تعدادی تصویر که آیا ناحیه بندی به شکل مناسب انجام شده است یا خیر.

بخش دوم

بخش دوم پروژه، به انتخاب ویژگی اختصاص دارد. در این بخش، شما باید پس از بدست آوردن رتبه اهمیت ناحیه های مختلف تصویر، به سراغ مهم ترین ناحیه هر تصویر بروید.

حال فرض کنید که به ازای کل تصاویر دیتاست که نواحی آنها مشخص شده، قصد دارید که صرفاً فقط با استفاده از مهم ترین ناحیه تصویر آنها را طبقه بندی کنید. ابتدا با استفاده از کل ویژگی ها، یک طبقه بند بر روی داده ها فیت کنید. با توجه به اینکه صرفاً از یک ناحیه استفاده کرده اید، ۲۰ درصد داده ها که طبقه بند در آنها بیشترین اشتباه را داشته است را حذف کنید (به عبارتی، طبقه بند به هر کلاس، یک درصد اطمینان اختصاص میدهد. ۲۰ درصدی که کمترین درصد را برای کلاس اصلی دارند را حذف کنید)

حال با داده های باقی مانده، الگوریتم انتخاب ویژگی را شروع کنید:

۱- در هر گام، قصد داریم ویژگی ای را انتخاب کنیم که دقت مناسبی برای طبقه بندی داده به ما ارائه دهد (به تنهایی اگر بخواهیم دیتاست را با آن طبقه بندی کنیم)

۲- کمترین کورلیشن را با ویژگی های انتخاب شده در مرحله قبلی داشته باشد.

دو معیار بالا را، به شیوه گفته شده در کلاس حل تمرین محاسبه کرده، نرمالایز کنید، و به شکل محاسبه f1-score ترکیب کنید.

برای شرط توقف، می توانید از دقت بدست آمده با مجموعه ویژگی های انتخاب شده، و یا میزان افزایش دقت با افزودن ویژگی استفاده کنید.

ویژگی های بدست آمده، و شرط توقف خود را توضیح دهید.