

ضمیمه

آشنایی با ماژول Dataset و DataLoader

کار با داده‌های بزرگ نیاز به بارگذاری یک باره کل داده در حافظه را دارد. این باعث قطع حافظه و اجرای بسیار کند برنامه‌ها می‌شود. علاوه بر این حفظ کد پردازش داده‌ها هم دشوار است. پایتورچ دو تا ماژول به اسم Datasets و DataLoader برای موازی سازی بارگذاری داده با رویکرد batching خودکار و برای خوانا و مدوله بودن کدها ارائه می‌دهد. Datasets و DataLoader به کاربران این امکان را می‌دهند که از داده‌های خود و همچنین مجموعه داده‌های از پیش بارگذاری شده استفاده کنند. در حالی که Dataset نمونه‌ها و برجسب‌های مربوطه را در خود جای می‌دهد، DataLoader مجموعه داده و نمونه‌گر را ترکیب می‌کند و یک تکرار در اطراف Dataset پیاده‌سازی می‌کند تا کاربران بتوانند به راحتی به نمونه‌ها دسترسی داشته باشند.

کلاس Dataset شامل سه متد اصلی است:

متد `__init__`: در این متد، آماده‌سازی جهت خواندن داده انجام می‌شود و سری شناسه‌های مرتبط جهت خواندن و تبدیلات داده تعریف می‌شود.

متد `__len__`: این متد تعداد نمونه‌های داده را بر می‌گرداند.

متد `__getitem__`: این متد، اندیس داده را دریافت کرده و داده مشخص را به همراه لیبل داده می‌خواند سپس پیش‌پردازش‌های لازم را انجام می‌دهد و در خروجی قرار می‌دهد.

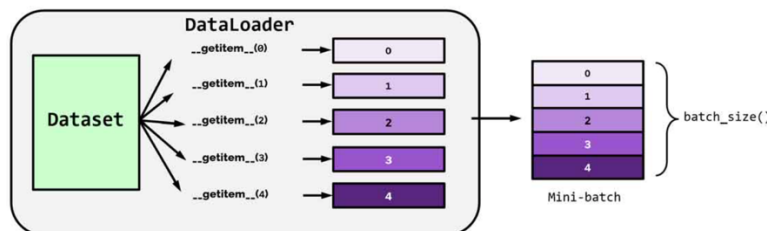
مزایای استفاده از DataLoader

Efficient data loading: به کاربر اجازه می‌دهد تا داده را به صورت موازی لود کند.

Data augmentation: به کاربر اجازه می‌دهد تا قبل اعمال داده به شبکه ی عصبی، تبدیلاتی روی داده اعمال کند.

flexibility: بسیار منعطف است و می‌توان با کمک آن فرمت‌ها و منابع مختلف داده را مدیریت کرد.

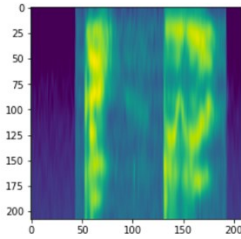
Shuffling: به کاربر اجازه می‌دهد که در هر تکرار آموزش شبکه عصبی داده را shuffle کرد. با این کار از پیش برآزش شبکه جلوگیری می‌کند و خاصیت عمومیت (generalization) شبکه را بهبود می‌دهد.



شکل ۱: نمایی از کارکرد ماژول DataLoader

داده‌ها

داده‌ها شامل پوشه‌ای حاوی MFCC و پوشه دیگر شامل برجسب‌ها است. در این تمرین به تعداد ۲۸۵۳۹ داده‌های صوتی و ۲۸۵۳۹ برجسب متناظر داریم. اطلاعات بیشتر در رابطه با بدست آوردن نمایش melspectrogram در لینک زیر قابل مشاهده می‌باشد. (<https://haythamfayek.com/2016/04/21/speech-processing-for-machine-learning.html>)



شکل ۲ نمایش melspectrogram داده

بردار زمینه

همانطور که در بالا توضیح داده شد، برای اینکه بتوانیم بفهمیم یک گام زمانی با چه واجی مطابقت دارد، به بردار «زمینه» نیاز داریم و برای آن نیاز به برخی اطلاعات قبل از مرحله زمانی و برخی اطلاعات بعد از مرحله زمانی هستیم. برای روشن تر شدن این موضوع، گام زمانی مورد نظر را t و اندازه بافت را c را در نظر بگیرید و فرض کنید واج مرتبط به این مرحله زمانی p می باشد. در این شبکه قصد داریم که DataLoader اطلاعات را از $[t - c, t + c]$ (ورودی) بگیرد و واج (هدف) را برگرداند. به عنوان مثال، فرض کنید داده صوتی زیر را داریم - برای سادگی در تصویرسازی فریم ها را به عنوان متن در نظر گرفتیم -:

$\text{mfcc} = [\text{A B C D E}]$

در تابع `get item`، اگر بخواهیم فریم در اندیس 0 ، بدون هیچ زمینه ای دریافت کنیم، $\text{mfcc}[0] = \text{A}$ را دریافت خواهیم کرد. اکنون می خواهیم این داده را با صفرهای زمینه در ابتدا و انتها اضافه کنیم تا به جای یک فریم در هر زمان، قاب پنجره ای را برای MLP فراهم کنیم تا واج در نمایه i را پیش بینی کند. فرض کنید زمینه را به طول ۳ در نظر بگیریم. بنابراین mfcc به صورت زیر می شود:

$\text{mfcc} = [0\ 0\ 0\ \text{A B C D E}\ 0\ 0\ 0]$

حال اگر بخواهیم به اولین فریم در ایندکس $i = 0$ دسترسی داشته باشیم، $\text{mfcc}[0] = 0$ را دریافت می کنیم که فریم مورد نظر ما نمی باشد.

یادآوری: فریم مورد نظر ما منطبق با واج داده شده شامل فریم A به همراه سه فریم قبل و بعد آن می باشد. بنابراین، ایندکس های شروع و پایان برای برگرداندن ورودی به صورت پنجره ای از هر فریم، از نظر تئوری، به این شکل می باشد:

$\text{start} = i - \text{context}$

$\text{end} = i + \text{context} + 1$

$\text{window_of_interest} = \text{mfcc}[\text{start}:\text{end}] = \text{mfcc}[i - \text{context}:i + \text{context} + 1]$

اما از آنجایی که در اطراف اولین و آخرین فریم، از `padding` استفاده می کنیم، اکنون برای دسترسی به فریم مورد نظر، `padding` مورد نظر قرار بدهیم.

$\text{start} = i - \text{context} + \text{context} = i$

$\text{end} = i + \text{context} + \text{context} + 1 = i + 2 * \text{context} + 1$

$\text{window_of_interest} = \text{mfcc}[\text{start}:\text{end}] = \text{mfcc}[i:i + 2 * \text{context} + 1]$

به این ترتیب، اگر بخواهید به فریم A در mfcc دسترسی داشته باشید، پنجره مورد نظر با $\text{mfcc}[0:7]$ مطابقت دارد، $[0\ 0\ 0\ \text{A B C D}]$ ، که دارای ۳ بردار زمینه صفر قبل و بعد از قاب مورد نظر A می باشد.