به نام خدا



دانشگاه صنعتی امیرکبیر

(پلی‌تکنیک تهران)

درس شبکه‌های عصبی

استاد صفابخش

تمرین هشتم

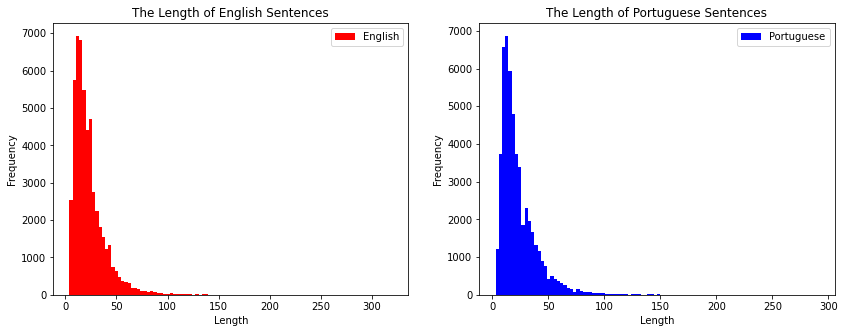
علیرضا مازوچی

۴۰۰۱۳۱۰۷۵

توجه

برای حل این تمرین از مطلب موجود در سایت تسورفلو[[1]](#footnote-1) استفاده شده است. از آنجایی که پیاده‌سازی موجود در این سایت کاملا متناسب با این تمرین بوده است، شباهت زیادی بین پیاده‌سازی من با این پیاده‌سازی وجود دارد. عملا منطق تمام مدل‌ها، پارامتر‌های مدل‌ها و بخشی از توابع پیچیده نظیر آموزش مدل عینا و بدون تغییر مورد استفاده قرار گرفته است.

سوال ۱

برای پیش‌پردازش تصمیم گرفتم تا جملات بلند را حذف کنم. به این ترتیب امکان کاهش حافظه مصرفی وجود خواهد داشت. برای این کار باید طول حداکثری یک جمله را تعیین کنیم. این مقدار باید به گونه‌ای تعیین شود که اکثر جملات کوتاه‌تر از حد بالا باشند چراکه در غیر این صورت مجبور به حذف بخش مهمی از داده آموزشی خواهیم شد. در تصویر زیر مشخص است که برای زبان مبدا (پرتغالی) و مقصد (انگلیسی) هر جمله چه تعداد توکن دارد:

همانطو رکه از تصویر بر می‌آید برای هر دو زبان شرایط تقریبا مشابه است. بیشتر جملات هر دو زبان در حدود ۲۵ توکن دارند و به ندرت جمله‌ای را می‌توان یافت که بالای ۱۲۸ توکن داشته باشد. بنابراین حدبالای تعداد توکن در یک جمله را برابر با 128 درنظر گرفته شده است.

سوال ۳ – بهینه‌ساز

مطابق با آموزش تسورفلو از بهینه‌ساز آدام با یک تابع برای تعیین مقدار نرخ یادگیری در هر لحظه استفاده کردم. این بهینه‌ساز در مقاله معروف Attention Is All You Need[[2]](#footnote-2) ارائه شده است. مطابق این مقاله پارامتر‌های آدام به این شکل کامل می‌شوند: ، و . تابع انتخاب‌شده برای نرخ یادگیری به این شکل عمل می‌کند که ابتدا نرخ یادگیری تا یک یک تعداد گام که اصطلاحا warm up steps‌ نامیده می‌شود به صورت خطی زیاد می‌شود و سپس متناسب با معکوس مربعات ریشه شماره گام کاهش می‌یابد. تابع مقدار نرخ یادگیری به شکل دقیق عبارت است از:

سوال ۳ – نقشه حرارتی توجه

سوال ۳ – نمونه خروجی

1. <https://www.tensorflow.org/text/tutorials/transformer> [↑](#footnote-ref-1)
2. <https://arxiv.org/pdf/1706.03762.pdf> [↑](#footnote-ref-2)