

## به نام خدا

تمرین اول درس فهم زبان، پاییز ۱۴۰۱

دانشکده مهندسی کامپیوتر، دانشگاه صنعتی امیرکبیر

هدف از این تمرین بررسی دو شبکه مشهور برای تبدیل کلمه فارسی به دنباله واج است. همان‌طور که می‌دانید در زبان فارسی ما حرکت‌ها را روی حروف نمی‌گذاریم. این باعث می‌شود که خواندن بعضی از کلمات مشکل شود و همچنین افرادی که تازه فارسی یادگرفته‌اند در تلفظ کلمات مشکل داشته باشند. همین امر برای کامپیوتر هم اتفاق می‌افتد و آن جایی است که می‌خواهیم از هوش مصنوعی برای تبدیل متن به گفتار استفاده کنیم. از این رو یکی از وظایفی که باید در این سامانه به درستی انجام شود تبدیل دنباله حرف (Grapheme) به دنباله واج (Phoneme) است که به این وظیفه Grapheme-to-Phoneme یا G2P گفته می‌شود. همان‌طور که در کلاس نیز شرح داده شد این یکی از مسائل دنباله به دنباله است که ورودی و خروجی از نوع کاراکتری هستند (البته با مجموعه کاراکترهای متفاوت) که قاعده‌ای روی طول آنها نیست (ورودی و خروجی می‌توانند هم طول و یا اینکه یکی از دیگری بزرگتر باشد).

برای این تمرین باید دو شبکه Encoder-decoder with Attention و همچنین Transformers را پیاده‌سازی و با هم مقایسه کنید. برای آموزش و ارزیابی مدل‌ها ۶۵ هزار نمونه آموزشی به شما داده شده است که باید از آن برای آموزش و اعتبارسنجی استفاده کنید (به هر نسبتی که دوست داشتید استفاده کنید). همچنین حدود ۱۰ هزار نمونه جهت ارزیابی مدل‌ها داده شده که باید از آن صرفاً جهت بررسی دقت استفاده کنید (به هر نحوی استفاده از آن در آموزش مجاز نیست). توجه کنید فقط باید از این دو مجموعه در نظر گرفته استفاده کنید که نتایج قابل مقایسه باشند. فایل‌های داده به صورت دو ستونی بوده که با یک کاراکتر tab از هم جدا شده‌اند. ستون اول نشان دهنده کلمه و ستون دوم نشان‌دهنده دنباله واج آن کلمه است. حروف کوچک و بزرگ در ستون دوم معنای متفاوتی دارد و نشان‌دهنده واج‌های متفاوتی است.

خروجی این مدل‌ها دنباله‌ای از واج‌ها است که باید با دنباله اصلی مقایسه شده و معیار ارزیابی بدست آید. برای این وظیفه دو معیار گزارش می‌شود. معیار اول دقت روی کلمه است به این شکل که اگر کل واج‌های یک کلمه به درستی تولید شده باشد درست و در غیر این صورت اشتباه در نظر گرفته می‌شود. معیار دوم نرخ خطای واج (PER) است که در آن واج‌ها با هم مقایسه شده و خطای‌های زیر ابتدا محاسبه می‌شوند.

- خطای درج: یک واج اضافه در خروجی درج شده باشد (I)
- خطای حذف: واجی از خروجی حذف شده باشد (D)
- خطای جانشانی: واجی با واج دیگر جایگزین شده باشد (S)

و در نهایت نرخ خطای واج نسبت مجموع سه خطای بالا به تعداد کل واج‌ها است. برای محاسبه این خطا به راحتی می‌توانید از پکیج jiwer و محاسبه نرخ خطای کاراکتر استفاده کنید (چون هر واج یک کاراکتر است نرخ یکسانی تولید می‌کند).

بعد از انجام تمرین باید یک گزارش مناسب (شامل بخش‌های مورد نیاز و با توضیحات و نتیجه‌گیری مناسب) برای تمرین بنویسید. تمرین فاقد گزارش بدون بررسی نمره صفر خواهد گرفت. گزارش ۳۰ درصد نمره تمرین را دارد. همچنین تمرین باید به طور کامل کد پایتون به صورت ماژولار باشد. یعنی اینکه نوتبوک و یا تبدیل نوتبوک به پایتون قابل قبول نیست و بدون بررسی صفر داده می‌شود. ۲۰ درصد نمره به نتایج بدست آمده اختصاص دارد و ۵۰ درصد تمرین هم به پیاده‌سازی مناسب و کدنویسی تمیز و ماژولار داده خواهد شد. ددلاین تمرین دو هفته است و تا دو روز با کسر نمره یک درصد به ازای هر ساعت می‌توانید با تاخیر ارسال کنید. بعد از این زمان دیگر امکان ارسال پاسخ تمرین وجود ندارد.

برای اینکه ترغیب شوید به انجام هرچه بهتر تمرین و انتخاب بهترین استراتژی‌ها، به نفرات اول تا سوم بر اساس نرخ خطای واج به ترتیب ۱۵، ۱۰ و ۵ درصد نمره اضافه داده خواهد شد. البته باید نرخ خطای واج از حدی کمتر باشد (یعنی اینکه اگر همه پیاده‌سازی‌ها مناسب نباشد، دلیلی برای دادن نمره اضافه نیست).

سوالات خود در رابطه با تمرین را می‌توانید در گروه تلگرام بپرسید.

موفق باشید

زینلی