



پاییز ۱۴۰۳

به نام خدا

تمرین اول درس فهم زبان: تشخیص وزن شعر

دانشکده مهندسی کامپیوتر، دانشگاه صنعتی امیرکبیر

در این تمرین به دنبال یافتن وزن عروضی در شعر فارسی (poem metre) هستیم.^۱ وزن شعر، نظم و چیدمان آهنگین الفاظ مصرع‌های شعر در کنار یکدیگر است. اشعار فارسی در وزن‌های مختلفی که به آن‌ها وزن عروضی نیز گفته می‌شود، سروده می‌شوند. وزن اشعار با توجه به نوع خواندن هر مصراع و تقطیع هجایی کلمات آن به دست می‌آید. لازم به ذکر است که همه‌ی مصراع‌های یک شعر دارای یک وزن یکسان هستند. به عنوان مثال برای مصرع زیر هم تقطیع هجایی انجام شده و هم وزن عروضی آن مشخص شده است. تعیین وزن شعر می‌تواند در سایر وظایف حوزه شعر مانند سرودن شعر، تعیین قالب شعری و ... نیز مفید واقع شود.

شب عاشقان بیدل چه شبی دراز باشد

شَ	پَ	عَا	شَ	قَا	نَ	بِی	دَل	چَ	شَ	بِی	دَ	رَا	زَ	بَا	شَد
—	—	U	—	—	U	—	—	U	—	—	U	—	—	U	—
فعلات				فاعلاتن				فعلات				فاعلاتن			

می‌توانید این مسئله را به دو صورت دسته‌بندی متن و دنباله به دنباله در نظر بگیرید. در حالت دسته‌بندی می‌توانید ورودی دنباله کاراکتر (مصراع شعر) باشد و خروجی یکی از صورت‌های رایج وزن شعر فارسی باشد. در حالت دنباله به دنباله 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). (راهنمایی: در این حالت مسئله‌ی تشخیص وزن هر مصراع به یک مسئله‌ی ترجمه‌ی ماشینی می‌تواند نگاشت شود).

با توجه به این نکته که غالباً همه‌ی مصراع‌های یک شعر دارای یک وزن یکسان هستند، با دادن یک مصراع از شعر شما می‌توانید وزن کل شعر را تعیین کنید. در نتیجه داده این تمرین شامل یک مصرع و وزن آن مصرع (که می‌توان به دو صورت نمایش داد) می‌باشد (جدول ۱). اینکه از کدام صورت وزن شعر استفاده می‌کنید کاملاً اختیاری است.

¹ https://fa.wikipedia.org/wiki/%D9%88%D8%B2%D9%86_%D8%B4%D8%B9%D8%B1_%D9%81%D8%A7%D8%B1%D8%B3%D8%8C

وزن شعر (صورت اول)	وزن شعر (صورت دوم)	مصراع شعر
فاعلاتن فاعلاتن فاعلن	-U---U---U-	شرح بسیاری بگفت از کائنات
مفاعِلن فعلاتن مفاعِلن فعِلن	-U-U-UU--U-U-UU	چه جای رفتن باغ است و گشتن بستان

مجموعه داده‌ای که در اختیار شما قرار می‌گیرد به سه بخش آموزش، ارزیابی و تست تقسیم شده است. بخش آموزش و ارزیابی را برای آموزش و پیدا کردن بهترین هایپرپارامترهای مدل می‌توانید استفاده کنید. قسمت تست که به شما داده می‌شود بدون برچسب وزن شعر است و با مدل نهایی‌تان آن را برچسب زده و هنگام تحویل کد و گزارش آن را هم تحویل می‌دهید.

برای این تمرین باید در هر دو حالت زیر آن را پیاده‌سازی و سپس خروجی آن‌ها را با یکدیگر مقایسه و تفاوت در نتایج را تحلیل کنید.

الف) (اجباری) دیدن این مسئله به عنوان یک مسئله دسته‌بندی و استفاده از ترنسفورمرهایی مثل BERT و XLMRoBERTa برای دسته‌بندی آن

ب) دیدن این مسئله به عنوان یک مسئله دنباله به دنباله و استفاده از مدل کدگذار و کدگشا به همراه attention (Encoder-Decoder with Attention). در این حالت می‌توانید از دو معماری زیر استفاده کنید:

- **ب ۱) (اجباری)** شبکه‌های بازگشتی LSTM یا GRU به همراه Attention و همچنین استفاده از دو جستجوی متفاوت گریدی، بیم سرچ برای دنباله خروجی. استفاده از لایه‌های دوطرفه نیز بلامانع است.

مواردی که می‌تواند در بهبود نتایج و نمره نهایی تمرین شما موثر باشد: استفاده از embedding های از پیش آموزش دیده مانند FastText به عنوان embedding اولیه، استفاده از توکنایز BPE به جای صرفاً index کردن کلمات شعر، استفاده از مکانیزم‌های مختلفی مانند early stopping و ذخیره بهترین گام، استفاده از teacher forcing به صورت schedule شده (مثلاً در ابتدا حداکثر استفاده از برچسب داده برای تولید خروجی مسئله توسط کدگشا و هرچه به انتها می‌رود از خروجی خود مدل استفاده شود)

- **ب ۲) (امتیازی ۱۰ درصد)** استفاده از ترنسفورمر کدگذار و کدگشا به صورت دنباله به دنباله: در این حالت صرفاً مدل شما تغییر کرده و باقی قسمت‌ها مثل حالت ۱ خواهد بود.

همچنین چون تعداد نمونه‌های وزن‌های اشعار متفاوت است، راهکارهایی برای رفع مشکل بالانس نبودن داده و بهبود نتایج بیان و پیاده‌سازی کنید (امتیازی ۵ درصد).

معیار ارزیابی در این تمرین Precision، Recall و F1-score همگی در حالت macro می‌باشد.

نکته مهم: 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. ۲۰ درصد نمره به نتایج بدست آمده اختصاص دارد و ۵۰ درصد تمرین هم به پیاده‌سازی مناسب و کدنویسی تمیز داده خواهد شد. (باتوجه به نمرات امتیازی، نمره کل تمرین تا ۱۲۰ درصد می‌تواند باشد). علاوه‌براین به بهترین روش که بالاترین F1-score macro را روی دادگان تست کسب کرده و از حداقلی بیشتر باشد ۱۰ درصد نمره تشویقی داده خواهد شد.

ددلاین تمرین تا جمعه ۴ آبان 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.

موارد تحویلی:

- کد تمرین به صورت بخش بندی شده
- گزارش شامل نتایج روی بخش ارزیابی دادگان
- داده تست برچسب زده شده (می‌تواند با هر سه مدل یا بهترین مدل برچسب زده شود)

موفق باشید

پاییز ۱۴۰۳