

تمرین پنجم

هدف: آشنایی با شبکه‌های عصبی بازگشتی^۱ و واحد حافظه کوتاه‌مدت بلند^۲

کد: کد این فعالیت را با استفاده از چارچوب‌های کاری تنسورفلو^۳ یا تورچ و به زبان پایتون بنویسید.

گزارش: ملاک اصلی انجام فعالیت گزارش آن است و ارسال کد بدون گزارش فاقد ارزش است. برای این فعالیت یک فایل گزارش در قالب pdf تهیه کنید و در آن برای هر سوال، تصاویر ورودی، تصاویر خروجی و توضیحات مربوط به آن را ذکر کنید. سعی کنید توضیحات کامل و جامعی تهیه کنید.

تذکر ۱: مطابق قوانین دانشگاه هر نوع کپی برداری و اشتراک کار دانشجویان غیر مجاز بوده و شدیداً برخورد خواهد شد. استفاده از کدها و توضیحات اینترنت به منظور یادگیری بلامانع است، اما کپی کردن غیرمجاز است.

تذکر ۲: در تمام موارد، مجموعه داده‌های ورودی را به سه قسمت آموزش (۷۰ درصد)، اعتبارسنجی^۴ (۲۰ درصد) و آزمون (۱۰ درصد) تقسیم نمایید. آموزش شبکه را روی مجموعه داده آموزشی انجام دهید و تعداد تکرارها را با استفاده از مجموعه داده اعتبارسنجی بیابید. در نهایت برای ارزیابی عملکرد مدل از مجموعه داده آزمون استفاده نمایید.

تذکر ۳: برای ارزیابی عملکرد مدل‌ها در دسته‌بندی از معیارهای دقت^۴ و ماتریس درهم‌ریختگی^۵ استفاده نمایید. بعلاوه، برای ارزیابی کیفی جملات تولید شده از معیار BLEU که در پکیج nltk پیاده‌سازی شده است استفاده نمایید. توضیحات بیشتر در مورد این معیار در انتهای فایل تعریف پروژه ذکر شده است.

تذکر ۴: برای تسریع در پیدا کردن مقادیر بهینه پارامترها، کفایست شبکه خود را فقط روی یک بخش کوچک از داده‌ها آموزش بدهید. پس از این که تاثیر افزایش یا کاهش مقدار هر پارامتر بر عملکرد مدل و حدود مقادیر بهینه آن را یافتید، ساختار بهینه را یکبار روی تمام داده‌ها آموزش دهید تا بتوانید مقایسه دقیقی از عملکرد شبکه در دو حالت بدست بیاورید.

راهنمایی: در صورت نیاز میتوانید سوالات خود را در خصوص پروژه از تدریس‌یار درس، از طریق ایمیل زیر بپرسید.

E-mail: ann.ceit.aut@gmail.com

ارسال: فایل‌های کد و گزارش خود را در قالب یک فایل فشرده با فرمت StudentID_HW۰۵.zip تا تاریخ ۱۴۰۰/۰۵/۰۱ ارسال نمایید. شایان ذکر است هر روز تاخیر باعث کسر ۱۰٪ نمره خواهد شد.

^۱ Recurrent neural network

^۲ Long Short-Term Memory (LSTM)

^۳ Validation

^۴ Accuracy

^۵ Confusion matrix

در این تمرین قصد داریم عملکرد شبکه‌های بازگشتی را در تولید متن مورد بررسی قرار دهیم. مجموعه داده مورد استفاده در این تمرین، مجموعه اشعار حکیم ابوالقاسم فردوسی است که از سایت گنجور گردآوری شده و به فایل تعریف پروژه پیوست شده است. فایل زیپ ضمیمه شده شامل یک فایل به نام allShahnameWords.txt است که محتوی لیست تمام کلمات مورد استفاده در این اشعار است. هر یک از اشعار این شاعر بزرگ، در یک فایل متنی جداگانه در دایرکتوری shahname در این فایل زیپ ذخیره شده است.

برای انجام این تمرین، مراحل زیر را انجام داده و نتایج را در فایل گزارش خود ذکر نمایید.

۱- تولید متن با استفاده از شبکه‌های عصبی بازگشتی را می‌توان به دو روش تولید متن کاراکتر به کاراکتر و کلمه به کلمه انجام داد. در فایل گزارش خود ساختار کلی مدل در هریک از این روش‌ها را توضیح داده و تفاوت‌ها، مزایا و معایب این روش‌ها نسبت به یکدیگر را بیان نمایید.

۲- مجموعه اشعار را به سه دسته آموزشی، اعتبارسنجی و تست تقسیم نمایید. توجه نمایید این تقسیم‌بندی را می‌توانید روی فایل‌ها انجام دهید به این شکل که تعدادی از اشعار را به طور کامل و تصادفی در مجموعه آموزشی، تعدادی را در مجموعه اعتبارسنجی و مابقی را در مجموعه تست قرار دهید.

۳- یک واحد حافظه کوتاه مدت بلند را برای آموزش کاراکتر به کاراکتر این اشعار ایجاد نمایید و آن را با استفاده از مجموعه داده‌های آموزشی و اعتبارسنجی آموزش دهید. تمام آزمایشات لازم برای رسیدن به مقادیر بهینه پارامترها را انجام داده و نتایج را در فایل گزارش خود ذکر نمایید. عملکرد شبکه را از لحاظ هر سه معیار BLEU، دقت و ماتریس درهم‌ریختگی مورد بررسی قرار دهید. چند نمونه از اشعار تولید شده توسط شبکه در مرحله تست را نیز در فایل گزارش خود ذکر نمایید.

۴- سوال ۳ را این بار برای حالت کلمه به کلمه تکرار کنید.

۵- با توجه به نتایج حاصل، روش‌های تولید متن به صورت کلمه به کلمه و کاراکتر به کاراکتر را با هم مقایسه نمایید و بر اساس نتایج آزمایشات انجام شده توضیح دهید کدام یک از این روش‌ها برای تولید اشعار مشابه با اشعار شاهنامه مناسب‌تر هستند. سعی کنید تحلیل جامعی ارائه دهید.

توضیحات تکمیلی در مورد معیار BLEU

این معیار، یکی از معیارهای پرکاربرد در حوزه ترجمه ماشینی برای ارزیابی کیفیت جملات تولید شده توسط ماشین است. در این کاربرد، یک جمله در یک زبان مبدا (بعنوان مثال زبان انگلیسی) به مدل داده می‌شود و مدل، یک جمله به زبان مقصد (بعنوان مثال زبان فارسی) تولید می‌کند. جمله تولید شده به زبان مقصد باید هم‌معنی با جمله ورودی باشد. در این کاربرد، معیار مناسبی برای ارزیابی خودکار میزان صحت جمله تولیدی وجود ندارد.

یکی از راهکارهای مناسب برای ارزیابی کیفی جمله تولید شده توسط مدل در این شرایط این است که برای هر یک از جملات ورودی موجود در مجموعه داده آموزشی، چند جمله معادل در زبان مقصد توسط عوامل انسانی تولید شود. سپس جمله تولیدی در زبان مبدا که توسط مدل تولید شده است، با جملات مرجع تولید شده توسط عوامل انسانی مقایسه شود. این مقایسه می‌تواند بر اساس تعداد کلمات مشترک بین این جملات باشد. در این شرایط، هر چه تعداد کلمات تولید شده مشترک بین جمله تولیدی مدل و جملات مرجع بیشتر باشد، جمله تولید شده توسط مدل باکیفیت‌تر ارزیابی می‌شود.

برای افزایش صحت عملکرد این معیار ارزیابی می‌توان به میزان اهمیت هر یک از کلمات مشترک، طول بزرگترین زیردنباله مشترک بین جملات و موارد مشابه دیگر امتیازهای مختلفی داد و با ترکیب این عبارات، یک امتیاز به کیفیت جمله تولیدی اختصاص داد. معیار BLEU با ساختاری مشابه، یک امتیاز کیفی به جمله تولید شده توسط مدل در مقایسه با یک یا چند جمله مرجع، اختصاص می‌دهد.

نمونه کد زیر به زبان پایتون نحوه محاسبه این امتیاز را نمایش می‌دهد.

```
from nltk.translate.bleu_score import sentence_bleu  
  
reference = [['this', 'is', 'a', 'test'], ['this', 'is', 'test']]  
  
candidate = ['this', 'is', 'a', 'test']  
  
score = sentence_bleu(reference, candidate)  
  
print(score)
```

در این تمرین در فرایند تست کافیسیت یک کلمه از ابتدای یکی از اشعار را انتخاب کرده و به عنوان ورودی به مدل بدهید. سپس شعر تولید شده توسط مدل را به عنوان جمله **candidate** و ادامه شعر در نسخه اصلی را به عنوان **reference** در نظر بگیرید.

موفق باشید