

به نام خدا

دانشگاه صنعتی امیرکبیر
دانشکده مهندسی کامپیوتر

پاسخ تمرین سری پنجم شبکه‌های عصبی

استاد:

دکتر صفابخش

دانشجو:

حلیمه رحیمی

شماره دانشجویی:

۹۹۱۳۱۰۴۳

بهار ۱۴۰۰

۱- شبکه‌های عصبی بازگشتی برای نوع خاص داده‌ی دنباله زمانی بکار می‌رود. این نوع از داده به علت آنکه مقدار کنونی آن به مقادیر گذشته‌اش بستگی دارد، از دیگر انواع داده متفاوت است. شبکه‌های بازگشتی که دارای حلقه‌ی پس‌خور هستند، باعث می‌شوند هر حالت شبکه نماینده‌ای از حالات پیشین آن باشد.

در حین آموزش، کاراکتر/کلمه اول از داده را به شبکه می‌دهیم، شبکه از این ورودی یک حالت و یک خروجی تولید می‌کند، ورودی بعدی، کاراکتر/کلمه‌ی دوم خواهد بود که با خروجی قبلی مقایسه می‌شود تا وزن‌های آن اصلاح گردد. از طرفی این ورودی با ضربی (وزن) از حالت قبلی، پیشبینی کلمه‌ی بعدی را حاصل می‌دهند. به همین ترتیب این آموزش ادامه می‌یابد. می‌توان حالت را به عنوان یک بردار ویژگی دید.

در حین آزمون، می‌توان کاراکتر/کلمه اول را به شبکه داد، و پس از تولید خروجی اول، این خروجی را به عنوان ورودی بعدی در کنار حالت تولید شده، به شبکه برگرداند. به همین ترتیب پیش می‌رویم تا به تعدادی که می‌خواهیم متن تولید کنیم.

کاراکترها/کلمات را برای دادن به شبکه به شکل one-hot درمی‌آوریم. به عبارتی هر کاراکتر/کلمه‌ی خاص یک مقدار عددی خاص داشته باشد که به شکل one-hot درآید. در مواقعی که تعداد کاراکترها/کلمات خاص زیاد است، می‌توان از embedding استفاده کرد که ورودی one-hot را به بازنمایی‌ای با حجم بخصوص نگاشت می‌کند.

در شبکه‌های بازگشتی تاثیر تمام مقادیر پیشین حفظ نمی‌شود و وابستگی‌های طولانی مدت را نمی‌تواند تشخیص دهد. تصور سلطه‌ی کامل به وابستگی‌های میان مقادیر دنباله، فقط حالت ایده‌آل بوده و در واقعیت چنین اتفاقی نمی‌افتد. به عبارتی در صورتی که عبارت کنونی به وجود عبارتی بستگی داشته باشد که در دو یا چند جمله پیش (به طور مثال) آمده باشد، نمی‌توان انتظار داشت شبکه این ارتباط را بیابد. در LSTM سعی می‌شود با فراموشی برخی مقادیر دنباله زمانی که اهمیت زیادی در پیشبینی ندارند، بتوان وابستگی‌های طولانی مدت را نگاه داشت.

از تفاوت‌های استفاده از کاراکتر و کلمه، ضعف سطح کاراکتر در حفظ وابستگی‌های طولانی مدت است. برای مثال، اگر بخواهیم در یک متن، پنجمین کلمه را پیشبینی کنیم، در سطح کلمه کافیسیت تاثیر چهار کلمه‌ی پیشین را داشته باشیم اما در سطح کاراکتر، شاید تاثیر بیش از بیست کاراکتر لازم باشد.

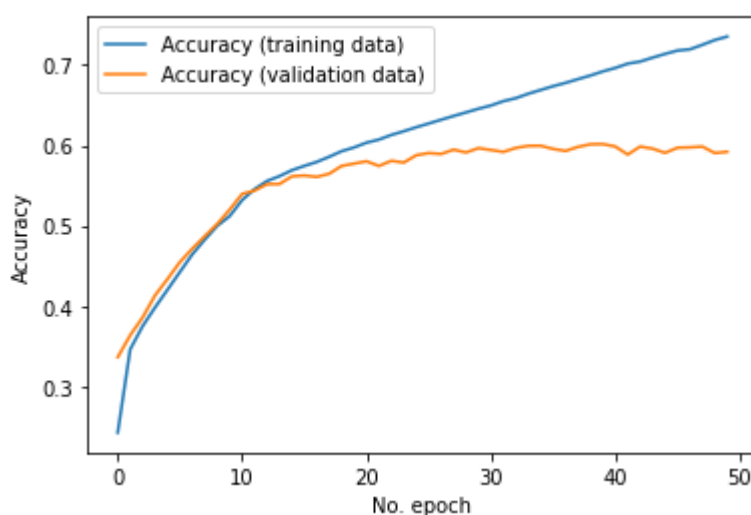
همچنین در حالت "ایده‌آل" وقتی لازم است ساختار جملات آموخته شود، محل قرارگیری کلمات مهم خواهد بود، اما از نگاه وزن، آهنگ و قافیه در شعر، ساختار کلمات مهم است. در صورت استفاده از روش کلمه به کلمه، با آنکه ممکن است محل قرارگیری برخی کلمات را بیاموزد اما با نگاه از سوی وزن، آهنگ و قافیه، این ساختار خود کلمات است که در تعیین کلمه‌ی بعدی شعر تاثیرگذار است. به عبارت بهتر، با آنکه می‌توان گفت همان کلمات را که در سطح کلمه می‌آموخت را فرا می‌گیرد ولی از نظر شهودی، ساختار کلمات را یاد می‌گیرد. هر چند آموزش آنکه هر کلمه در کجا می‌تواند قرار بگیرد نیز می‌تواند راهی برای یادگیری وزن و آهنگ باشد. از سوی دیگر ارتباط معنایی کلمات در سطح کلمه به کلمه قابل درک‌تر است.

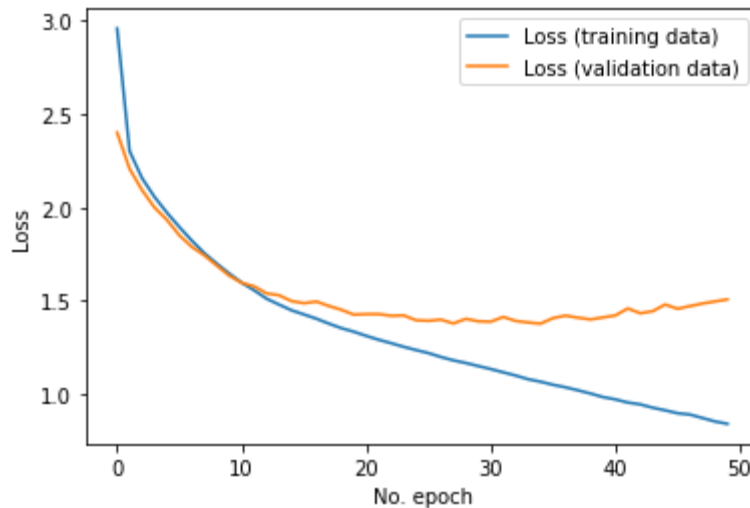
۲- با استفاده از train_test_split، و انتخاب یک random_state خاص، داده‌ها را به طور یکسان در هر دو کد مرتبط با کلمه و کاراکتر تقسیم کرده‌ام.

۳- برای هر دو روش کاراکتر به کاراکتر و کلمه به کلمه از یک معماری بهره بردم: شبکه‌ای با امبدینگ ۵۱۲ و تعداد ۵۱۲ نورن LSTM (با اینکه کاراکترها تعدادشان کمتر از ۵۱۲ است بنظر مشکلی ایجاد نمی‌شود). نتایج برای روش کاراکتر به کاراکتر به ترتیب زیر است:

با وجود آنکه پس از مدتی شبکه دچار اورفیت شد اما نتایج بهتر برای مجموع آموزش موجب افت چندانی برای مجموعه ولیدیشن در ۵۰ تکرار نشد (علاوه بر آن نتایج بهتری برای آزمون داشت که جلوتر به آن می‌پردازم) بنابراین تکرار را تا ۵۰ ادامه دادم.

```
Epoch 37/50
27/27 [=====] - 1s 43ms/step - loss: 1.0376 - accuracy: 0.6778 - val_loss: 1.4210 - val_accuracy: 0.5931
Epoch 38/50
27/27 [=====] - 1s 43ms/step - loss: 1.0216 - accuracy: 0.6824 - val_loss: 1.4102 - val_accuracy: 0.5981
Epoch 39/50
27/27 [=====] - 1s 43ms/step - loss: 1.0045 - accuracy: 0.6870 - val_loss: 1.4006 - val_accuracy: 0.6014
Epoch 40/50
27/27 [=====] - 1s 43ms/step - loss: 0.9855 - accuracy: 0.6921 - val_loss: 1.4114 - val_accuracy: 0.6016
Epoch 41/50
27/27 [=====] - 1s 42ms/step - loss: 0.9730 - accuracy: 0.6964 - val_loss: 1.4220 - val_accuracy: 0.5986
Epoch 42/50
27/27 [=====] - 1s 43ms/step - loss: 0.9562 - accuracy: 0.7017 - val_loss: 1.4595 - val_accuracy: 0.5886
Epoch 43/50
27/27 [=====] - 1s 42ms/step - loss: 0.9463 - accuracy: 0.7043 - val_loss: 1.4339 - val_accuracy: 0.5985
Epoch 44/50
27/27 [=====] - 1s 42ms/step - loss: 0.9285 - accuracy: 0.7092 - val_loss: 1.4448 - val_accuracy: 0.5959
Epoch 45/50
27/27 [=====] - 1s 42ms/step - loss: 0.9144 - accuracy: 0.7138 - val_loss: 1.4807 - val_accuracy: 0.5907
Epoch 46/50
27/27 [=====] - 1s 42ms/step - loss: 0.8989 - accuracy: 0.7180 - val_loss: 1.4577 - val_accuracy: 0.5970
Epoch 47/50
27/27 [=====] - 1s 42ms/step - loss: 0.8917 - accuracy: 0.7195 - val_loss: 1.4723 - val_accuracy: 0.5976
Epoch 48/50
27/27 [=====] - 1s 42ms/step - loss: 0.8733 - accuracy: 0.7251 - val_loss: 1.4860 - val_accuracy: 0.5985
Epoch 49/50
27/27 [=====] - 1s 42ms/step - loss: 0.8549 - accuracy: 0.7311 - val_loss: 1.4971 - val_accuracy: 0.5905
Epoch 50/50
27/27 [=====] - 1s 42ms/step - loss: 0.8422 - accuracy: 0.7354 - val_loss: 1.5078 - val_accuracy: 0.5920
```





نتیجه برای مجموعه آزمون:

8/8 [=====] - 0s 17ms/step - loss: 1.5244 - accuracy: 0.5930
[1.524377465248108, 0.5929631590843201]

متأسفانه مسئله‌ای که در اینجا پیش می‌آید، به دلیل قطع کردن شعر در اندازه‌ی کوچکترین شعر، برخی کلمات نصف و نیمه ذخیره شده‌اند و موجب می‌شود شبکه کلمات اشتباهی را دریابد. علاوه بر این به علت ضعف در حفظ وابستگی‌های طولانی‌انطور که انتظار می‌رود قافیه‌ها آموخته نمی‌شوند، هر چند در برخی اشعار و یا برخی بیت‌ها قافیه، وزن و آهنگ کاملاً رعایت شده بود (این مسئله ممکن است به علت اورفیت باشد، چرا که با رسیدن دقت آموزش به حدود ۹۸ درصد این اتفاق بیشتر حاصل می‌شد).

لازم به ذکر است که احتمال می‌دهم با داشتن شبکه‌ی یکسان، در روش کاراکتر به کاراکتر اورفیت کمتر و یا دیرتر رخ بدهد، به دلیل ضعف در حفظ وابستگی طولانی مدت. به عبارتی این مسئله موجب می‌شود شبکه نتواند خود اشعار آموزش را نیز حفظ کند.

دقت شود که در هر شعر چاپ شده، اولی متن اصلی و دومی متن تولید شده توسط شبکه است. همچنین مصرع سمت راستی در واقع مصرع دوم و مصرع سمت چپی، مصرع اول هر بیت است.

ایا آلت و لشکر و رای پاک	یکی نامور بود نامش سیاک
جهانديده با داد و فرمانروا	که در شهر چهارم بد او يادشا
چو آگه شد از پيش بهمن برفت	مر او را خجسته يسر بود هفت
ایا لشکر و کوس و با دار و گیر	ز چهارم بیامد سوی اردشیر
ز باره درآمد چنانچون سزی	چو چشمش به روی سپید رسید

که با من نیاتند پر از خون کنیم	یکی نامه فرزند هرکس که بود
جهاندار با من به بالان باد	به در به کار و به هر مهتری
به هر دو به مردی به دانش همال	از ایران بدو گفت با موبدان
که خواهران را بر میان اندر افراسن	بدو گفت با موبدان و ردان
به پ	بران باره ای برکشیدند شاه

بییوندم از گفته باستان	کنون یرشگفتی یکی داستان
چه بازی نمود ای يسر گوش دار	نگه کن که مر سام را روزگار
دلش بود جوینده کام را	نیود ایچ فرزند مرسام را
ز گلبرگ رخ داشت و ز مشک موی	نگاری بد اندر شبستان اوی
که خورشید چهر و برومند بود	از آن ماهش امید فرزند بود
	ز م

بسی در بران دین به دست	کنون ای خردمند ماه
برو تیره شد تاج و تخت و سرای	بران دخت زان سان به سر بر نهان
به مردی به روز اندر آمد به دست	به شاه جهان این شگفتی رسید
به هر جای عاج و به بالا بلند	به پيش اندر آمد به کار سپاه
بران دختر آید ز	به پيش اندر آمد به دببای ده

شگفتی بماند اندران نامدار	چو بشنید شد نامه را خواستار
رخ تاجور گشت همچون زریر	چو آن نامه برخواند مرد دبیر
به گفتار مشتاب و تندى مکن	بدو گفت کای مرد چیره سخن
چنان هم نماید همی راه تو	بزرگی نماید همی شاه تو
نیاتم ز گوینده همدستان	کسی باز خواهد ز هندوستان
	به

به دل را به بازی به دست کیان	چو برخاست از روز بگذشت شاه
سخن گفت با من ز جای نشست	بدو گفت کای نامجوی
به هر سو نگویم به زن روزگار	یکی نامه بنوشت با موبد شاه
بران دخت شاهیش بنشست باز	به بالا بلند و زیان دانستی
هم از راه گردان به دست	به نخچیرگه با سپاه انجمن
	چنین

نتیجه‌ی Bleu در سطح حرف برابر با مقدار زیر است:

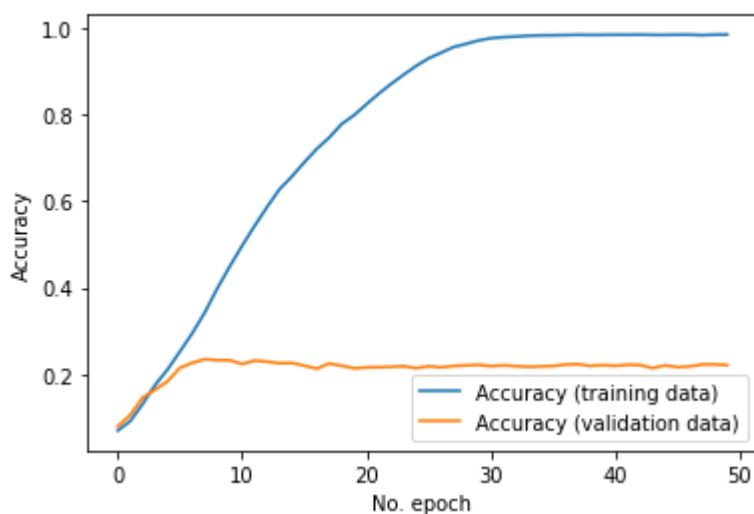
0.8529540004360148

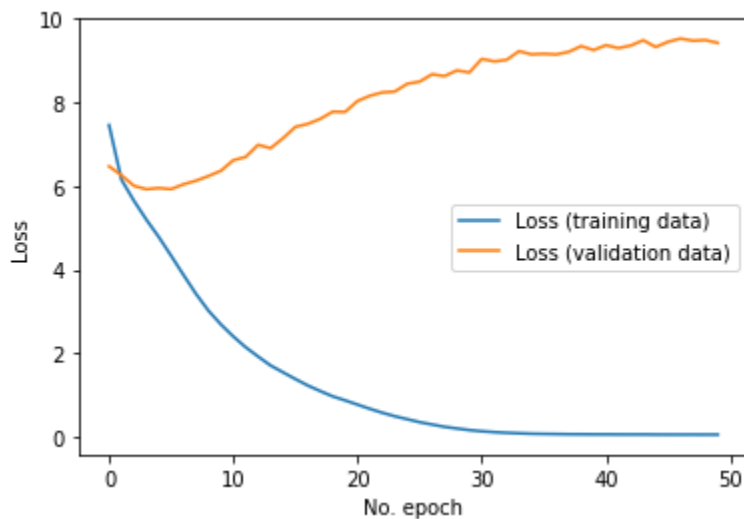
و در سطح کلمه برابر با:

به ترتیب تعداد تشخیص صحیح از آن حرف، و خود حرف، از راست به چپ.

۴- نتایج برای روش کلمه به کلمه:

```
Epoch 37/50
27/27 [=====] - 3s 127ms/step - loss: 0.0667 - accuracy: 0.9834 - val_loss: 9.1475 - val_accuracy: 0.2224
Epoch 38/50
27/27 [=====] - 3s 129ms/step - loss: 0.0627 - accuracy: 0.9840 - val_loss: 9.2096 - val_accuracy: 0.2233
Epoch 39/50
27/27 [=====] - 3s 126ms/step - loss: 0.0613 - accuracy: 0.9837 - val_loss: 9.3471 - val_accuracy: 0.2194
Epoch 40/50
27/27 [=====] - 3s 129ms/step - loss: 0.0595 - accuracy: 0.9839 - val_loss: 9.2517 - val_accuracy: 0.2212
Epoch 41/50
27/27 [=====] - 3s 128ms/step - loss: 0.0587 - accuracy: 0.9841 - val_loss: 9.3692 - val_accuracy: 0.2194
Epoch 42/50
27/27 [=====] - 4s 130ms/step - loss: 0.0574 - accuracy: 0.9840 - val_loss: 9.2990 - val_accuracy: 0.2221
Epoch 43/50
27/27 [=====] - 4s 131ms/step - loss: 0.0568 - accuracy: 0.9843 - val_loss: 9.3605 - val_accuracy: 0.2212
Epoch 44/50
27/27 [=====] - 3s 128ms/step - loss: 0.0572 - accuracy: 0.9838 - val_loss: 9.4876 - val_accuracy: 0.2142
Epoch 45/50
27/27 [=====] - 3s 128ms/step - loss: 0.0549 - accuracy: 0.9838 - val_loss: 9.3251 - val_accuracy: 0.2206
Epoch 46/50
27/27 [=====] - 3s 129ms/step - loss: 0.0540 - accuracy: 0.9842 - val_loss: 9.4504 - val_accuracy: 0.2166
Epoch 47/50
27/27 [=====] - 3s 128ms/step - loss: 0.0538 - accuracy: 0.9842 - val_loss: 9.5272 - val_accuracy: 0.2181
Epoch 48/50
27/27 [=====] - 3s 129ms/step - loss: 0.0543 - accuracy: 0.9833 - val_loss: 9.4760 - val_accuracy: 0.2227
Epoch 49/50
27/27 [=====] - 3s 127ms/step - loss: 0.0534 - accuracy: 0.9844 - val_loss: 9.4911 - val_accuracy: 0.2227
Epoch 50/50
27/27 [=====] - 3s 127ms/step - loss: 0.0530 - accuracy: 0.9845 - val_loss: 9.4238 - val_accuracy: 0.2212
```





نتیجه دقت برای مجموعه آزمون:

8/8 [=====] - 0s 38ms/step - loss: 9.3051 - accuracy: 0.2305
[9.30506706237793, 0.23046875]

حال به علت نصف و نیمه نبودن کلمات، کلمات به درستی تولید شده‌اند. علاوه بر این همانطور که بیان شد اشعار تولید شده از نظر معنایی نیز بهترند. همانطور که پیش از این بیان کردم و از نتایج پیداست، اوفیت شدن موجب نتایج جذاب مجموعه آزمون شده است و گرنه صحت هم برای اعتبارسنجی و هم برای آزمون تنها حدود ۲۰ درصد است و حتی معیار Bleu نیز برای آزمون در حدود ۲۷ درصد است که تنها کمی بهتر از نتیجه‌ی حالت کاراکتر به کاراکتر شده است.

ایا آلت و لشکر و رای پاک	یکی نامور بود نامش سباک
جهانديده با داد و فرمانروا	که در شهر چهارم بد او پادشا
چو آگه شد از پیش بهمن برفت	مر او را خجسته پسر بود هفت
ایا لشکر و کوس و با دار و گیر	ز چهارم بیامد سوی اردشیر
ز باره	چو چشمش به روی سپید رسید

یر از خون دل و روی چون سندروس	یکی نامه بنوشت دیگر بطوس
کزو دید نیرو و بخت و هنر	تخت آفرین کرد بر دادگر
خداوند دیهیم شاهنشهی	خداوند پیروزی و فرهی
به خشکی چو پیل و نهنگ اندر آب	بی پشه تا یر و جنگ عقاب
دم خویش بی رای او نشمرد	ز پیمان و فرمان او نگذرد

بپیوندم از گفته باستان	کنون پرشگفتی یکی داستان
چه بازی نمود ای پسر گوش دار	نگه کن که مر سام را روزگار
دلش بود جوینده کام را	نبود ایچ فرزند مرسام را
ز گلبرگ رخ داشت و ز مشک موی	نگاری بد اندر شبستان اوی
که خورشید چهر و برومند بود	از آن ماهش امید فرزند بود
	ز سام

سر مهران رایه آغوش دار	کنون شیرین بار بد گوش دار
به پرداخت بی داد و بی کام گاه	چو آگاه شد بار بد زانک شاه
یر از آب مرگان و دل یر ز خون	ز چهارم بیامد سوی طیسفون
شده لعل رخسار او شنبیلید	بیامد بدان خاتمه او را بدید
خروشان	ز ماتی همی بود در پیش شاه

شگفتی بماند اندران نامدار	چو بشنید شد نامه را خواستار
رخ تاجور گشت همچون زریر	چو آن نامه برخواند مرد دبیر
به گفتار مشتاق و نندی مکن	بدو گفت کای مرد چیره سخن
چنان هم نماید همی راه تو	بزرگی نماید همی شاه تو
نباشم ز گوینده همدستان	کسی باز خواهد ز هندوستان
	به لشکر همی گوید این

برو آفرین کرد بهرام و هور	چو بر تخت بنشست بهرام گور
جهادار و بیدار و بیننده را	پرستش گرفت آفریننده را
خداوند افزونی و کمتری	خداوند پیروزی و برتری
کزویست گیتی سراسر به پای	خداوند داد و خداوند رای
ازو یافتم کافریدست بخت	ازان پس چنین گفت کاین تاج و تخت
	بدو هستم امید و هم زو هراس

نتیجه Bleu در سطح کلمه برابر است با:

0.277072192513369

به علت بزرگی ماتریس درهم‌ریختگی قادر به نمایش یا ذخیره‌ی آن نبودم. در اینجا عباراتی که بیشترین دفعات به درستی در جای خود تشخیص داده شده‌اند را می‌آورم:

(نتایج از راست به چپ است)

```
[ 3. 3. 3. 3. 5. 8. 9. 22. 76. 104.]  
['/', '*', 'آمد', 'جو', 'بد', 'از', 'که', 'به', 'ز', 'و']
```

برای راحتی کار در نوشتن کد از / به جای t/ و از * به جای n/ استفاده کرده بودم.

مقدار temperature را برای هر دو روش برابر با ۰/۳ قرار دادم که در مقدار لوجیت تاثیر گذاشته و در کلام ساده، موجب می‌گردد انتخاب کاراکتر/کلمه به راحتی انجام نگیرد.

اکنون آنچه را که در رابطه با اورفیت شدن گفتم در نتایج روش کلمه به کلمه منتهی تنها با ده تکرار نشان می‌دهم.

```
Epoch 1/10  
27/27 [=====] - 12s 152ms/step - loss: 7.2168 - accuracy: 0.0666 - val_loss: 6.4645 - val_accuracy: 0.0879  
Epoch 2/10  
27/27 [=====] - 4s 133ms/step - loss: 6.1438 - accuracy: 0.0906 - val_loss: 6.2723 - val_accuracy: 0.1091  
Epoch 3/10  
27/27 [=====] - 3s 126ms/step - loss: 5.7148 - accuracy: 0.1283 - val_loss: 6.0900 - val_accuracy: 0.1504  
Epoch 4/10  
27/27 [=====] - 3s 123ms/step - loss: 5.2282 - accuracy: 0.2019 - val_loss: 5.9278 - val_accuracy: 0.2056  
Epoch 5/10  
27/27 [=====] - 3s 122ms/step - loss: 4.7498 - accuracy: 0.2380 - val_loss: 5.8819 - val_accuracy: 0.2212  
Epoch 6/10  
27/27 [=====] - 3s 124ms/step - loss: 4.2762 - accuracy: 0.2668 - val_loss: 5.9722 - val_accuracy: 0.2215  
Epoch 7/10  
27/27 [=====] - 3s 125ms/step - loss: 3.8187 - accuracy: 0.3038 - val_loss: 6.1363 - val_accuracy: 0.2218  
Epoch 8/10  
27/27 [=====] - 3s 127ms/step - loss: 3.3350 - accuracy: 0.3563 - val_loss: 6.2924 - val_accuracy: 0.2188  
Epoch 9/10  
27/27 [=====] - 3s 126ms/step - loss: 2.8890 - accuracy: 0.4202 - val_loss: 6.4373 - val_accuracy: 0.2169  
Epoch 10/10  
27/27 [=====] - 3s 125ms/step - loss: 2.5048 - accuracy: 0.4789 - val_loss: 6.6248 - val_accuracy: 0.2169
```

برای آزمون:

```
8/8 [=====] - 0s 28ms/step - loss: 6.5163 - accuracy: 0.2237  
[6.516274452209473, 0.22369025647640228]
```

و حالا اشعار تولید شده:

ایا آلت و لشکر و رای پاک	یکی نامور بود نامش سبک
جهانپیده با داد و فرمانروا	که در شهر چهارم بد او پادشا
چو آگه شد از پیش بهمن برفت	مر او را خجسته پسر بود هفت
ایا لشکر و کوس و با دار و گیر	ز چهارم بیامد سوی اردشیر
ز بازه	چو چشمش به روی سپید رسید

که با احمد سهل بودی به اهله	یکی نامه بنوشت دیگر بطوس
که ای شیردل مهتر نامدار	همه نامداران خویش
کز و وپسه خواهد همی	همه موبدان را به زه کرد و راد
ز گیتی همی جست گردان و تیر	چو از کوه اندر آمد پدید
ز گیتی همی جست هر کس به دست	همی گفت هرکس که دارید خویش

بپیوندم از گفته باستان	کنون پرسگفتی یکی داستان
چه بازی نمود ای پسر گوش دار	نگه کن که مر سام را روزگار
دلش بود جوینده کام را	نبود ایچ فرزند مرسام را
ز گلبرگ رخ داشت و ز مشک موی	نگاری بد اندر سیستان اوی
که خورشید چهر و برومند بود	از آن ماهش امید فرزند بود
	ز سام

کز و روشنایی گرفتست روز	کنون شیرین خردمند خسرو شدست
همی کرد با موبدان	چو آمد ز درگاه مهراب شاه
بزرگان و سالاروش بخردان	به نزدیک آن کرگ باید شدن
که در پادشاهی بجند ز جای	چنان بد که روزی یکی تندباد
که از کشتگان خاک شد شهریار	ز گیتی همی نام او بود ماخ
	چو آمد به نزدیک شاه

شگفتی بماند اندران نامدار	چو بشنید شد نامه را خواستار
رخ تاجور گشت همچون زریر	چو آن نامه برخواند مرد دبیر
به گفتار مستجاب و تندی مکن	بدو گفت کای مرد چیره سخن
چنان هم نماید همی راه نو	بزرگی نماید همی شاه نو
نیاتم ز گوینده همدستان	کسی باز خواهد ز هندوستان
	به لشکر همی گوید این

به کردار آتش ببیمود راه	چو خورشید تابان برآمد ز نخچیرگاه
که بگشای کشور نگاه ایچ بود	به جایی که درویش بد یزدگرد
به دل شاه با موبدان	چو آمد به نزدیک مازندران
بزرگان و توران سپاه	به نزدیک آن کرگ باید شدن
به یک شب به سر بر نهاد	چو آمد به نزدیک مازندران
	ز هر سو نهادند روی

۵- به طور کلی، در صورتی که می‌شد بر مسئله‌ی عدم حفظ وابستگی‌های طولانی مدت فائق آمد، و مسئله‌ی وزن، آهنگ و قافیه‌ی شعر مطرح بود، روش کاراکتر به کاراکتر را انتخاب می‌کردم. روش کلمه به کلمه به این علت که می‌تواند همنشینی کلمات را بیابد، قادر به دریافت وزن نیز هست. علاوه بر این دریافت معنا نیز در سطح کلمه و بالاتر صورت می‌گیرد. احتمال می‌دهم در صورتی که

معماری بهتری بود، نتایج کلمه به کلمه "معنا"ی بهتری پیدا می کردند. اگر منظور سوال از "اشعار مشابه اشعار شاهنامه" از نظر وزن، آهنگ و قافیه است، کاراکتر به کاراکتر را انتخاب می کنم.

البته در اینجا هر دو روش نتیجه تقریباً یکسان داشتند.

در صورت داشتن نوعی معماری به صورت سلسله مراتبی که در لایه های پایین تر ارتباط حروف و در لایه های بالاتر ارتباط کلمات و جملات را می آموخت، نتیجه احتمالاً بهتر می شد.