

گزارش تمرین چهارم درس شبکه‌های عصبی مصنوعی

استاد: دکتر خردپیشه

سید محمد امین حسینی – ۹۹۴۲۲۰۵۷

شرح مسئله:

پیاده‌سازی یک مدل با هدف پیش‌بینی واقعی یا جعلی بودن توییت‌هایی با محتوای وقایع طبیعی.

دو مدل در این تمرین پیاده‌سازی شده:

Transformers و BERT -۲

LSTM -۱

دیتاست و پیش‌پردازش:

مجموعه‌ای از حدوداً ۷۶۰۰ توییت به زبان انگلیسی که در آن در مورد یک واقعه‌ی طبیعی نوشته شده است.

	id	keyword	location	text	target
31	48	ablaze	Birmingham	@bbcmtd Wholesale Markets ablaze http://t.co/l...	1
32	49	ablaze	Est. September 2012 - Bristol	We always try to bring the heavy. #metal #RT h...	0
33	50	ablaze	AFRICA	#AFRICANBAZE: Breaking news: Nigeria flag set a...	1
34	52	ablaze	Philadelphia, PA	Crying out for more! Set me ablaze	0
35	53	ablaze	London, UK	On plus side LOOK AT THE SKY LAST NIGHT IT WAS...	0

در مرحله پاک‌سازی و پیش‌پردازش، مجموعه‌ای از توابع بر روی این دیتاست اعمال شده که به شرح زیر است:

- کوچک‌سازی همه حروف در متن
- برگرداندن مخفف‌ها به حالت معمولی و جدا از هم
- پاک کردن URLها و تگ‌های HTML
- پاک کردن کاراکترهای که در در مجموعه ASCII وجود ندارند (مانند Û، Ò)
- پاک کردن کاراکترهای خاص مثل نمادها، ایموجی‌ها و دیگر کاراکترهای گرافیکی
- پاک کردن علائم نگارشی
- جایگزین کردن اصطلاحات، مخفف‌ها و غلط‌های تایپی و با حالت درست و یا رایج
- پاک کردن کلمات پرتکرار بدون اطلاعات مفید (stop words)
- POS Tagging
- Snowball Stemmer
- Lemmatization

حالت نهایی دیتاست:

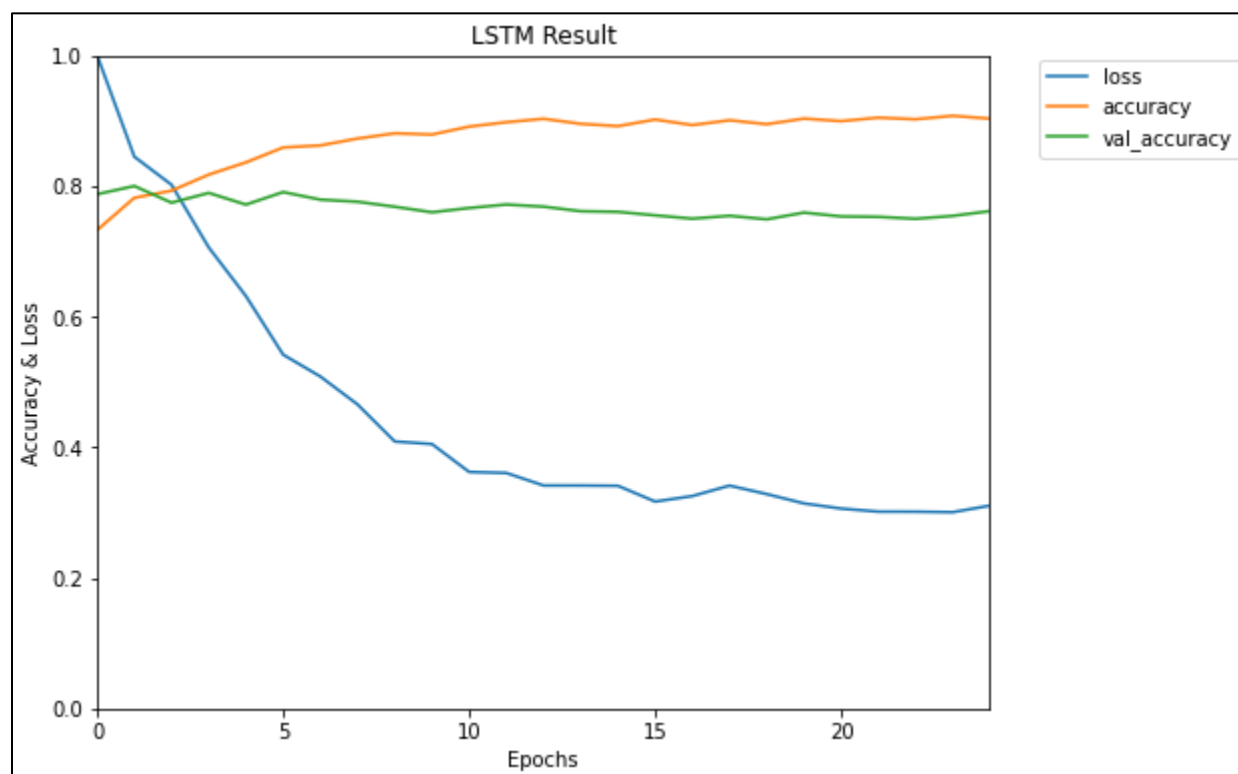
target		lemmatize_text
0	1	deed reason earthquake may allah forgive u
1	1	forest fire near la ronge sask canada
2	1	resident ask shelter place notified officer ev...
3	1	13000 people receive wildfire evacuation order...
4	1	get send photo ruby alaska smoke wildfire pour...

:LSTM

RNNها از حافظه کوتاه مدت رنج می‌برند. اگر یک داده‌ی دنباله‌ای بیش از اندازه طولانی باشد، این شبکه‌ها برای انتقال اطلاعات از مراحل قبلی به مراحل بعدی دچار مشکل خواهند شد. LSTM به عنوان راه حلی برای حافظه کوتاه مدت ایجاد شده است. این نوع شبکه سازوکارهای درونی به نام دروازه دارد که می‌توانند جریان اطلاعات را تنظیم کنند. این دروازه‌ها اهمیت نگهداری یا حذف داده‌ها را آموزش می‌بینند. با این کار، اطلاعات مفید را از زنجیره طولانی دنباله‌ها برای پیش بینی نهایی منتقل کند.

شبکه استفاده شده در این تمرین شامل ۲ لایه Bidirectional LSTM با ۲۵۶ واحد پنهان و ۲ لایه Dropout و دولایه Fully Connected است. برای آماده‌سازی توکن‌ها از وکتورهای ۲۰۰ بعدی از پیش آموزش دیده‌ی GloVe استفاده شده است. این شبکه با نرخ یادگیری ۰/۰۰۱ در ۲۵ epoch آموزش دیده که نتایج آن در ادامه آمده است:

Train Loss: 0.182		Train Acc: 90.28%		
Val. Acc: 76.84%				
Train Loss: 0.182		Train Acc: 89.51%	Train Loss: 0.533	Train Acc: 73.25%
Val. Acc: 76.13%			Val. Acc: 78.74%	
Train Loss: 0.182		Train Acc: 89.16%	Train Loss: 0.450	Train Acc: 78.18%
Val. Acc: 76.04%			Val. Acc: 79.99%	
Train Loss: 0.169		Train Acc: 90.17%	Train Loss: 0.427	Train Acc: 79.28%
Val. Acc: 75.50%			Val. Acc: 77.44%	
Train Loss: 0.173		Train Acc: 89.32%	Train Loss: 0.376	Train Acc: 81.74%
Val. Acc: 75.00%			Val. Acc: 78.94%	
Train Loss: 0.182		Train Acc: 90.06%	Train Loss: 0.337	Train Acc: 83.61%
Val. Acc: 75.44%			Val. Acc: 77.14%	
Train Loss: 0.175		Train Acc: 89.46%	Train Loss: 0.289	Train Acc: 85.89%
Val. Acc: 74.90%			Val. Acc: 79.07%	
Train Loss: 0.167		Train Acc: 90.32%	Train Loss: 0.271	Train Acc: 86.21%
Val. Acc: 75.92%			Val. Acc: 77.91%	
Train Loss: 0.163		Train Acc: 89.93%	Train Loss: 0.248	Train Acc: 87.26%
Val. Acc: 75.34%			Val. Acc: 77.59%	
Train Loss: 0.161		Train Acc: 90.46%	Train Loss: 0.218	Train Acc: 88.06%
Val. Acc: 75.27%			Val. Acc: 76.82%	
Train Loss: 0.161		Train Acc: 90.21%	Train Loss: 0.216	Train Acc: 87.87%
Val. Acc: 74.98%			Val. Acc: 75.97%	
Train Loss: 0.160		Train Acc: 90.74%	Train Loss: 0.193	Train Acc: 89.09%
Val. Acc: 75.42%			Val. Acc: 76.62%	
Train Loss: 0.166		Train Acc: 90.31%	Train Loss: 0.192	Train Acc: 89.76%
Val. Acc: 76.15%			Val. Acc: 77.16%	



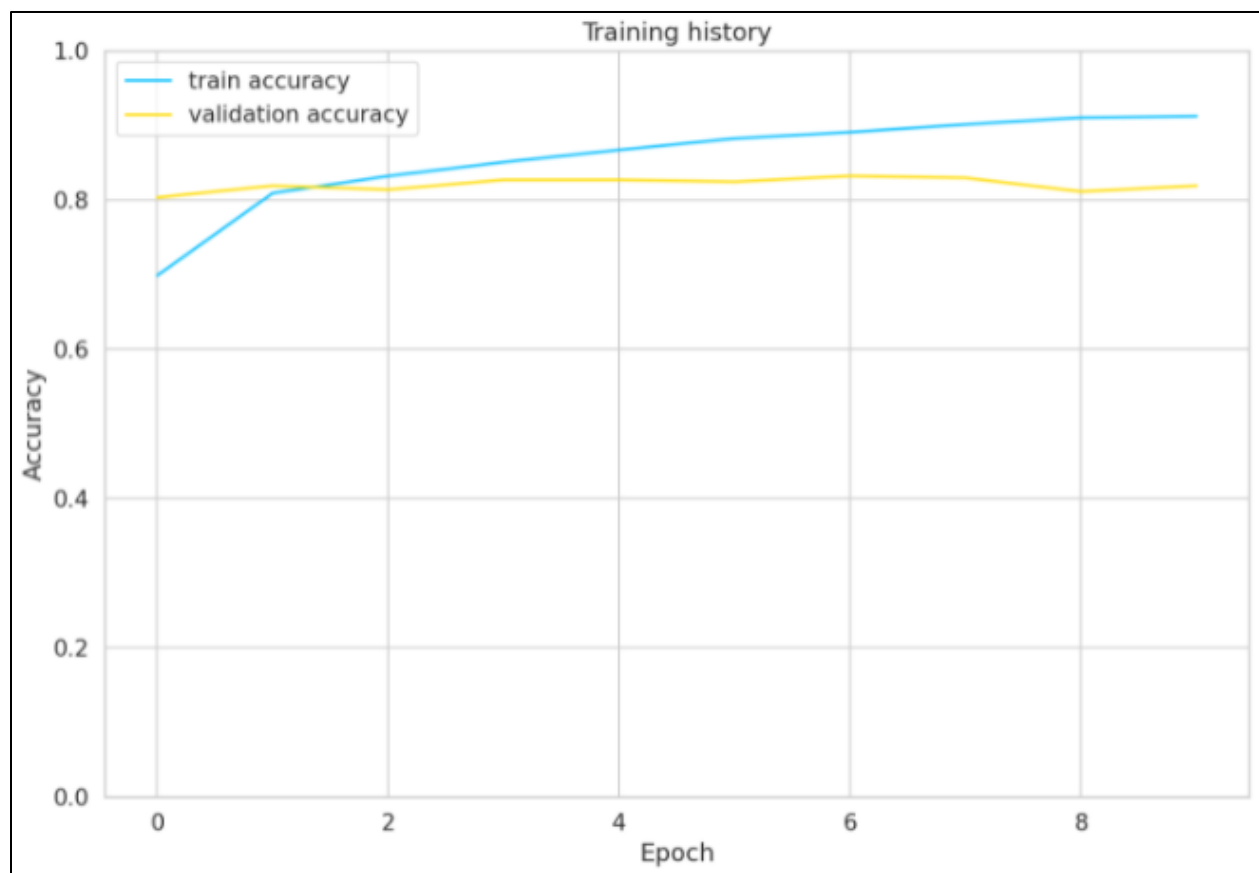
BERT و Transformers:

BERT (Bidirectional Encoder Representations from Transformers) [مقاله‌ای است که اخیراً توسط محققان Google AI Language منتشر شده](#) که بهترین نتایج را در طیف گسترده‌ای از مسائل NLP، از جمله پاسخ‌گویی به سوال (SQuAD v1.1)، استنتاج زبان طبیعی (MNLI) و سایر موارد ارائه داده است.

نوآوری اصلی BERT استفاده از آموزش دو طرفه ترانسفورمر، یک مدل توجه (Attention) جدید و محبوب، برای مدل سازی زبان است. این برخلاف تلاش‌های قبلی است که به دنباله‌ای از متن فقط از چپ به راست یا آموزش ترکیبی چپ به راست و راست به چپ نگاه می‌کرد. نتایج این مقاله نشان می‌دهد که یک مدل زبانی که به طور دو طرفه آموزش دیده است، می‌تواند نسبت به مدل‌های تک جهت زبان، مفهوم عمیق‌تری از زمینه و جریان زبان ارائه دهد. در این مقاله، محققان جزئیات تکنیکی جدید به نام Masked LM (MLM) را ارائه می‌دهند که امکان آموزش دو طرفه را در مدل‌هایی که قبلاً غیرممکن بود، فراهم می‌کند.

در این تمرین برای پیش‌بینی واقعی یا جعلی بودن توییت‌ها، این شبکه با نرخ یادگیری 2×10^{-5} در ۱۰ epoch آموزش دیده است که نتایج آن به صورت زیر است:

Epoch 6/10	Epoch 1/10
Train loss 0.345 accuracy 0.882	Train loss 0.593 accuracy 0.698
Val loss 0.476 accuracy 0.824	Val loss 0.494 accuracy 0.803
Epoch 7/10	Epoch 2/10
Train loss 0.326 accuracy 0.890	Train loss 0.478 accuracy 0.808
Val loss 0.469 accuracy 0.832	Val loss 0.471 accuracy 0.818
Epoch 8/10	Epoch 3/10
Train loss 0.301 accuracy 0.901	Train loss 0.439 accuracy 0.831
Val loss 0.478 accuracy 0.829	Val loss 0.460 accuracy 0.813
Epoch 9/10	Epoch 4/10
Train loss 0.284 accuracy 0.910	Train loss 0.399 accuracy 0.850
Val loss 0.516 accuracy 0.811	Val loss 0.444 accuracy 0.826
Epoch 10/10	Epoch 5/10
Train loss 0.276 accuracy 0.912	Train loss 0.376 accuracy 0.866
Val loss 0.560 accuracy 0.8188	Val loss 0.443 accuracy 0.826



	precision	recall	f1-score	support
0	0.81	0.87	0.84	209
1	0.82	0.75	0.78	172
accuracy			0.81	381
macro avg	0.81	0.81	0.81	381
weighted avg	0.81	0.81	0.81	381

