دانشکده مهندسی کامپیوتر

دکتر محمدرضا محمدی

بهار ۱۴۰۰

تمرین سری هشتم

یادگیری عمیق

مجتبی نافذ 96431335

مهلت تحویل : ۱ خرداد ۱۴۰۰ ساعت ۲۳:۵۹:۵۹

---

۱. دلیل وقوع Vanishing و Exploding گرادیان را توضیح داده و راه‌حل های موجود برای جلوگیری از هرکدام را شرح دهید.

پدیده ی vanishing یا همان محو شدگی گرادیان زمانی رخ می دهد که تعداد لایه ها زیاد میشود و مشکل اصلی در نوع Optimizer ماست، که با گرادیان کار میکند و با افزایش لایه ها باید تعداد زیادی گرادیان روی lost بگیریم تا تاثیر خطای خروجی روی پارامتر های اولیه را بسنجیم که ضرب این گرادیان ها در نهایت به صفر میل می کند و پارامتر های ابتدایی شبکه train نمی شوند. و در نتیجه خوب همگرا نمی شود.

چند راه حل این مشکل:

۱- استفاده از تابع فعالسازی Relu یا مانند آن: به دلیل خطی بودن آن، در backpropagation آن گرادیان برگشت به عقب خوبی خواهد داشت.(در مقابل tanh, sigmoid گرادیان را محو می کنند.)

۲- استفاده از مقدار دهی مناسب وزن هاست بگونه ای که پتانسیل رخداد محو شدگی گرادیان کمینه شود. به این منظور امروزه از الگوریتم های مقداردهی اولیه ای نظیر Xavier و یا MSRA (که به He initialization یا Kaiming initialization هم معروف است) برای این کار استفاده میشود.

۳- استفاده از batch normalization.

۴- استفاده از optimizer های بهتر که با گرادیان کار نکنند

۵- استفاده از شبکه هایی با معماری بهتر (معماری LSTM ، معماری Resnet و ..)

۶- کاهش نرخ یادگیری

پدیده ی Exploding gradient یا همان انفجار گرادیان زمانی رخ میدهد که تعداد لایه ها زیاد می شود و در backpropagation ما زنجیره ای از ضرب گرادیان در وزن را برای آپدیت پارامتر های ابتدایی شبکه نیاز داریم. که اگر وزن های ما مقادیری بزرگی داشته باشند این زنجیره یک عدد بزرگ می شود و پارامتر های شبکه را به بینهایت سوق میدهد. و شبکه خوب train نمی شود و در نتیجه خوب همگرا نمی شود.

چند راه حل این مشکل:

۱- Clip کنیم. یعنی یک بیشینه برای آن تعیین کنیم. و مقادیر بزرگ را به آن clip می کنیم. به طور نمونه بیشینه را ۱۰ قرار میدهیم زمانی که یک پارامتر در حال آپدیت با مقدار ۲۴۰ بود این عدد را به ۱۰ clip کرده و با مقدار ۱۰ آپدیت می کنیم.

۲- تمام راه های بیان شده برای vanishing برای exploding هم کارایی دارد.

۲. برای پیاده‌سازی هر یک از مسائل زیر، یک معماری از میان انواع معماری‌های RNNای که آموخته‌اید را پیشنهاد داده و دلیل انتخاب خود را نیز توضیح دهید (منظور از میان معماری‌های one to many، many to one، many to many و ترکیب آنهاست).

الف) ترجمه متن زبان فارسی به زبان انگلیسی

ب) دسته‌بندی نظرات کاربران راجع به یک محصول (به دو دسته خوب و بد)

ج) مکالمه با چت‌بات

د) بدست آوردن موضوع یک متن (از میان چند دسته مشخص)

ه) سراییدن یک شعر با استفاده از یک بیت (لینک)

و) تهیه یک سخنرانی با استفاده از یک کلمه

الف) ترجمه ی متن زبان فارسی به زبان انگلیسی:

ترجمه ی بین دو زبان،ما چندین کلمه از زبان ۱ را به چندین کلمه از زبان ۲ تبدیل می کنیم که تعداد کلمه های زبان ۲ احتمال زیاد متفاوت از زبان ۱ است. پس ما مدل many to many غیر متعادل نیاز داریم.

برای ترجمه بین دو زبان بهترین مدل، مدل sequence to sequence است که در واقع ابتدا با یک Many to One مفهوم زبان یک را یاد میگیرد. سپس با یک One to many مفهوم را در زبان دوم بیان می کند.

ب) دسته بندی نظرات کاربران راجع به یک محصول (به دو دسته خوب و بد)

یک Many to One برای مسئله کافی است. چون تعدادی کلمه در ورودی می بیند و خروجی کافیست یک احتمال برای خوب بودن محصول بدهد و بد بودن مکمل آن است و به راحتی به دست می آید.

ج) مکالمه با چت بات

جون طول ورودی و خروجی متفاوت است

مدل sequence to sequence بهتر است. که در واقع ابتدا با یک Many to One مفهوم جمله ما را یاد میگیرد. سپس با یک One to many پاسخ مفهوم را در جمله خودش بیان می کند.

د) به دست آوردن موضوع متن از بین چند دسته بندی

طول ورودی دنباله از کلمات است و خروجی در واقع چند دسته بندی خاص هست که از قبل مشخص است

مدل Many to One به خوبی آن را مدل می کند.

ه) سراییدن یک شعر با استفاده از یک بیت (لینک)

خوب قرار است یک بیت با تعدادی کلمات مشخص ببینیم و یک شعر که ممکن است چندین بیت داشته باشد را بسراییم. البته بستگی به پیاده سازی دارد. به نظر من مدل sequence to sequence بهتر است. که در واقع ابتدا با یک Many to One مفهوم بیت ما را یاد میگیرد. سپس با یک One to many شعر را در خروجی خودش بیان می کند. (مدل های پیچیده تری هم می شود ارائه داد)

و) تهیه ی یک سخنرانی با استفاده از یک کلمه

با استفاده از One to Many می توان این کار را به خوبی انجام داد.

۳. همانطور که می‌دانیم از RNNها برای پیش‌بینی سری زمانی استفاده می‌شود. یک نمونه از سری‌های زمانی، قیمت سهام در فواصل زمانی معین (ساعتی، روزانه و غیره) است. در این تمرین می‌خواهیم از LSTMها برای مدلسازی سری زمانی بازار سهام استفاده کنیم. طبق توضیحات داده شده در فایل نوت‌بوک پیوست شده، یک شبکه LSTM که توانایی پیش‌بینی قیمت سهام را داشته باشد پیاده‌سازی کنید.

اگربه جای پیش‌بینی قیمت سهام، به دنبال پیش‌بینی جهت تغییر قیمت (صعودی و نزولی بودن) سهام باشیم، به نظر شما نتایج آن به نسبت حالت قبل چقدر مورد اطمینان است؟ پاسخ خود را توضیح دهید.

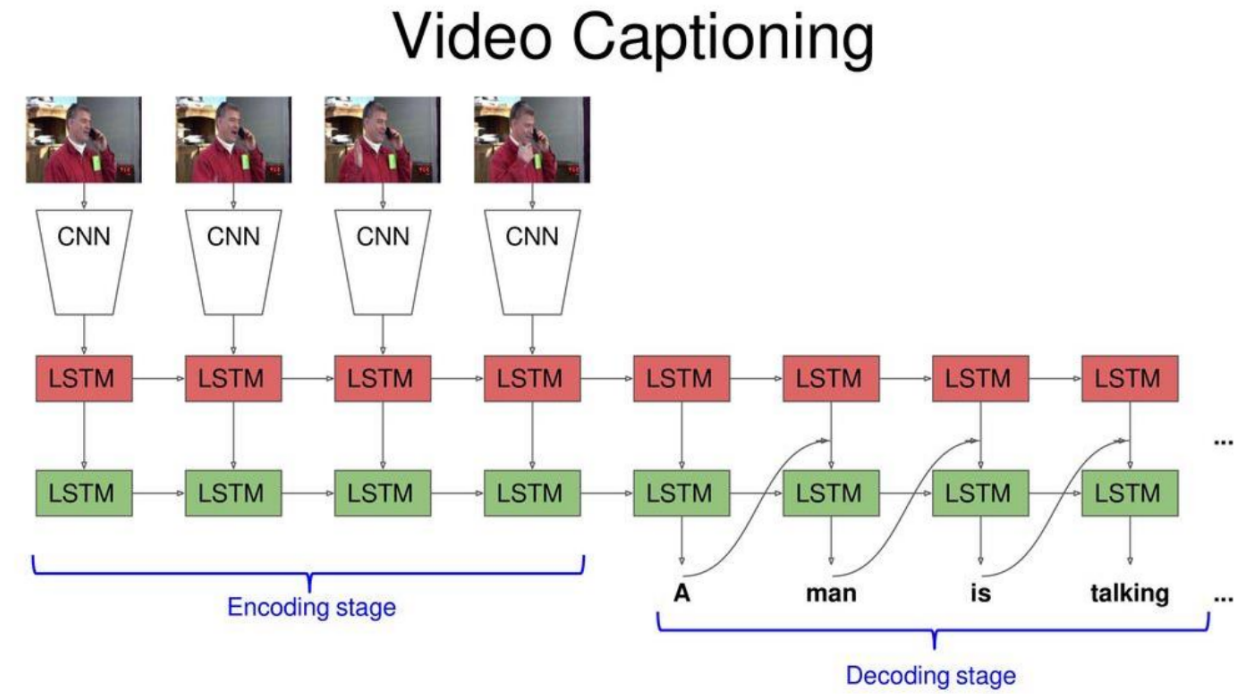
کد ضمیمه شده است.

به نظر من اگر دنبال جهت تغییر باشیم واقعا اصلا قابل اطمینان نیست چون جنس مسئله کاملا در جهت تغییر تصادفی است.

با اینکه ما با گذشته بد حدس میزنیم ولی احتمالا اشتباه خواهیم نمود.

۴. کد مربوط به شبکه نشان داده شده در شکل زیر را بنویسید (نیازی به آموزش مدل نیست، بلکه کافی است مدل را تعریف کنید و summary آن را چاپ کنید). برای بخش CNN می‌توانید از هر شبکه‌ای استفاده کنید و هایپرپارامترهای LSTM را نیز به صورت دلخواه انتخاب کنید. برای اعمال یک (یا چند) لایه غیربازگشتی بر روی یک دنباله می‌توانید از لایه TimeDistributed استفاده کنید.

کد ضمیمه شده.



Venugopalan et al., "Sequence to Sequence - Video to Text", ICCV 2015