



تمرین سری چهارم درس شبکه های عصبی

استاد: سعید رضا خرد پیشه

ارائه دهنده : دانا نمیرانیان

۳۱ خرداد ۱۴۰۲

## ۱ تمرین :

What are the main difficulties when training RNNs? How can you handle them?

۱- Vanishing and exploding gradients: که در حالت vanishing مقادیر گرادیان ها در هنگام backpropagation بسیار کوچک میشوند و در حالت exploding مقادیر گرادیان ها بزرگ میشوند و به همین دلیل یادگیری را برای شبکه ی عصبی بسیار مشکل میکند در حالت اول که گرادیان ها بسیار کوچک میشوند و باعث میشود که در شبکه های RNN اطلاعات لایه های اولی به لایه های بعدی خوب انتقال پیدا نکند و در حالت دوم که گرادیان ها مقادیر بزرگی دارند ممکن است همگرایی به جواب نهایی را در مرحله یاد گیری دچار مشکل کند .

یکی از راه حل هاب اولیه ان استفاده از نوع دیگری از شبکه های RNN به اسم LSTM یا long short term memory و gated recurrent units GRU میباشد که هر دوی انها دارای قابلیت به خاطر سپاری بیشتری در وابستگی های بین داده ها در داده های صفی دارد . راه حل بعدی استفاده از روش gradient clipping میباشد که برای گرادیان ها یک استانه و حد بالا مشخص میکنیم که مقدار گرادیان ها نباید از مقدار استانه تجاوز کند .

۲- پیچیدگی محاسبات : شبکه های عصبی RNN به دلیل این که صفی از داده ها را پردازش میکنند ممکن است در هنگام پردازش صف های طولانی مقدار محاسبات افزایش پیدا کند به دلیل این که در این نوع شبکه های عصبی هر داده ورودی به صورت صف پردازش میشود و به صورت موازی پردازش نمیشود . برای حل این مشکل میتوان داده ها را به صورت دسته ای وارد کرد و بعد از ان که همه ی داده های داخل یک دسته وارد شدند مقدار گرادیان ها بروز رسانی شوند. یک روش دیگر هم این است که معماری شبکه را تا حد ممکن ساده کنیم و از تعداد لایه های کمتری استفاده کنیم و روش اخر هم میتوان از stacked recurrent networks استفاده کرد که میتواند موثر تر از روش های معمولی باشد.

۳- مشکل بعدی که ممکن است به وجود آید به خاطر سپاری وابستگی بین داده ها long term dependencies در صف های طولانی میباشد که در این قسمت هم یکی از روش ها استفاده از شبکه های RNN به اسم LSTM یا long short term memory و gated recurrent units GRU میباشد که وابستگی های بیشتری را به خاطر میسپارند.

## ۲ تمرین :

How can 1D convolutional layers be beneficial when used in conjunction with RNNs?

۱- لایه های convolutional میتوانند در استخراج ویژگی ها از (feature) داده های صفی کمک کننده باشند به اصن صورت که فیلتر را در هر قسمت از صف برده و ویژگی های ان قسمت از صف را استخراج میکنیم در حالتی که در شبکه RNN معمولی نمیتوان از هر قسمت صف را که خواستیم انتخاب کنیم و ویژگی های مهم انرا استخراج کنیم.

۲- کاهش بعد مسئله : به دلیل این که با استفاده از استخراج ویژگی ها از صف ها تعداد ویژگی های مهم تر را استخراج کرده و بعد و پیچیدگی مسئله را کاهش دهند.

۳- به دلیل این که لایه های convolutional میتوانند از تکنیک های مختلف regularization مثل dropout بهره بگیرند میتوانند به جلوگیری از overfit کمک کنند

۴- افزایش سرعت با استفاده از موازی سازی : در شبکه های عصبی به دلیل این که داده ها به صورت صف و طولی پردازش میشوند روش زمانبر است و اما لایه های convolutional میتوانند به صورت موازی پردازش شوند و سرعت یادگیری را بالا ببرند.  
RNN

## ۳ تمرین :

What are the pros and cons of using a stateful RNN versus a stateless RNN?

statefull به شبکه های عصبی ای گفته میشود که اطلاعات به دست آمده از صف داده قبلی را نگه داری میکنند از مزایای ان میتوان به ۱- نگهداری بیشتر اطلاعات داده های پیشین و به خاطر سپاری بهتر وابستگی های میان داده ها long dependencies در صف داده ها میباشد ۲- همچنین ساینز شبکه و پیچیدگی ان کمتر میباشد و در نتیجه زمان آموزش ان نیز کمتر

میشود. از معایب آن میتوان به ۱- پیاده سازی این شبکه ها سخت میباشد مخصوصا آماده کردن داده ها برای آموزش آنها. ۲- داده ها را نمیتوان shuffle کرد به دلیل این که ترتیب داده ها در آن اهمیت دارد.

stateless هم به شبکه عصبی گفته میشود که در آن internal state شبکه بعد از هر batch ریست میشود به این معنی که شبکه دیگر اطلاعات به دست آمده از داده های پیشین را در خود ذخیره نمیکند. مزایای آن شامل ۱- پیاده سازی آن ها راحت تر میباشد و برای صف داده هایی به کار میرود که صف قبلی از داده ها به صف بعدی وابسته نباشد. معایب آن شامل: ۱- نمیتواند اطلاعات از صف داده های پیشین را به خاطر پسپارد. اگر صف داده ای که به شبکه داده میشود به صف داده ی قبل از آن وابسته باشد این یک عیب به حساب میاید. ۲- مشکل vanishing gradient ممکن است به وجود آید که در آن مقدار گرادیان ها بسیار کوچک میشود و عمل یادگیری و اپدیت وزن های شبکه را با مشکل مواجه میکند

## ۴ تمرین :

Why do people use encoder-decoder RNNs rather than plain sequence-to-sequence RNNs for automatic translation?

۱- یکی از دلایل اصلی آن این میباشد که در هنگام ترجمه یک متن توسط sequence to sequence کل متن در یک لحظه بررسی نمیشود و تنها کلمه به کلمه بررسی میشود که این یک نقص است به دلیل این که در هنگام ترجمه باید به کل متن توجه کرد و معنی کلمه را نسبت به کل متن ترجمه کرد. اما در encoder-decoder RNN چون ویژگی های اصلی متن و وابستگی های بین کلمات آنها توسط encoder در یک بردار ذخیره میشود که این بردار شامل اطلاعات تمامی قسمت های جمله است و سپس توسط decoder به یک جمله ی دیگر ترجمه میشود که decoder آنرا به صورت کلی تر میبیند و ترجمه بهتری میدهد.