



به نام خدا
درس یادگیری عمیق
تمرین سری سوم
استاد درس : دکتر محمدرضا محمدی
دستیاران : محمد مصطفی رستم خانی ، علیرضا حقانی
دانشگاه علم و صنعت ایران، دانشکده مهندسی کامپیوتر
نیمسال دوم تحصیلی ۱۴۰۳ - ۱۴۰۴

مهلت تحویل : ۱۴۰۴/۰۲/۰۵

لطفا به نکات موجود در سند قوانین انجام و تحویل تمرین ها دقت فرمایید.

سوالات تئوری



۱. از مدل های Sequence-to-Sequence برای حل مسائل مشترک بینایی ماشین و پردازش زبان مانند مسئله Video Captioning و از نوع Many-to-Many آنها استفاده می شود. در مقاله CAM-RNN زیر برای حل این مسئله از معماری های CNN و RNN استفاده شده است. نحوه حل مسئله را در این مقاله توضیح دهید. آیا ایده ی این مقاله را برای مسئله Image Captioning می توان استفاده کرد؟ توضیح دهید (۱۰ نمره).

RNN یک متد که برای تسک video captioning است. در آن از co-attention به همراه RNN استفاده شده است تا یک caption توصیفی و دقیق تولید کنند. رویکرد های معمولی بر اساس RNN هستند که کلمه به کلمه برای ویدیو، کپشن تولید میکنند و کلمه فعلی، بر اساس محتوای تصویر و کلمه قبلی تولید شده، تولید میشود. در CAM-RNN از تصویر و متن، یکسری feature تولید میشود و RNN به عنوان decoder عمل کرده و برای ویدیو کپشن تولید میکند. CAM-RNN از سه ماژول تشکیل شده است: visual attention module و text attention module و balancing gate. در حین تولید کپشن، visual attention به صورت تطبیقی معلوم میکند که کدام قسمت از کدام فریم برجسته ترین است و باید به آن توجه کرد و به کدام فریم مرتبط ترین به کپشن است و باید به آن توجه کرد. Text attention module به صورت اتوماتیک به مرتبط ترین کلمه یا phrase تولید شده قبلی توجه میکند. علاوه بر آن بین این دو attention module از یک balancing gate استفاده شده است تا بتوان تاثیر هر کدام از text feature و visual feature را در هنگام تولید کپشن، کنترل

و تنظیم کرد. به صورت دقیق تر این روش شامل بخش های زیر است:

- یک visual attention module که شامل دو لایه است و برای استخراج ویژگی های قوی تر و بهتر استفاده شده است و مرتبط ترین فریم و برجسته ترین ناحیه در آن فریم را مشخص میکند.
- یک phrase-level text attention module که در هنگام تولید کپشن، به مرتبط ترین قسمت از phrase های تولید شده قبلی توجه میکند و میتواند text feature های دقیق تری به دست آورد.
- یک balancing gate که برای تنظیم تاثیر visual features در هنگام فرآیند تولید کپشن است و برای تولید کلمات غیر بصری مناسب است.

معماری این سیستم در زیر آورده شده است:

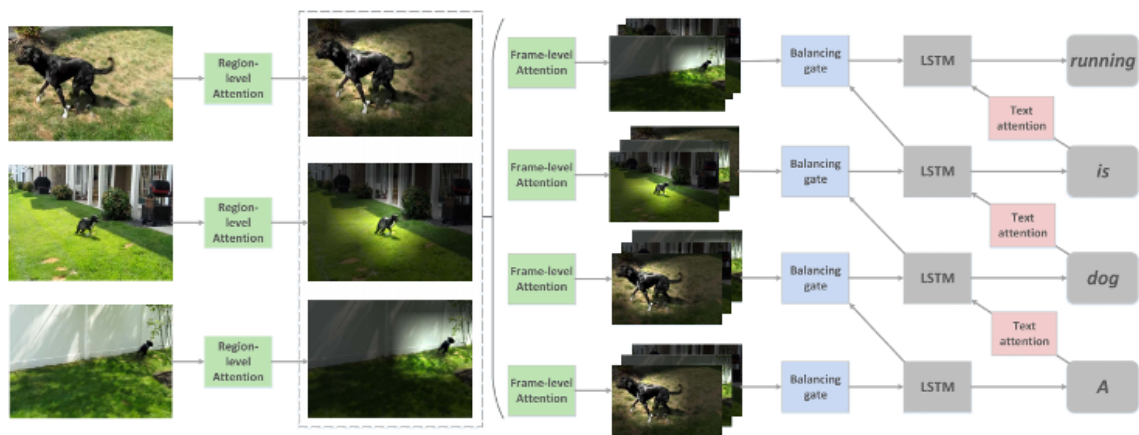


Fig. 2. The architecture of the CAM-RNN (co-attention model based RNN). Specifically, the CAM is composed of three parts as follows. 1) The visual attention module contains the region-level attention layer and frame-level attention layer, which can adaptively focus on the most correlated visual features. 2) The text attention layer operates on previously generated words, which can automatically attend to the most correlated text features. 3) The balancing gate is designed to adaptively regulate the influence of the visual features. Finally, with the encoded visual feature and text feature, LSTM is employed as the

با توجه به اینکه در image captioning میتوان تصویر را به صورت فیلمی که تنها یک فریم دارد، تصور کرد میتوان به این صورت برای آن کپشن تولید کرد ولی دیگر frame-level attention کاربردی ندارد زیرا تنها یک فریم داریم که میتوانیم به آن توجه کنیم.

۲. با فرض اینکه ابعاد داده ورودی ۱۲۸ و ابعاد بردار نهان ۶۴ باشد، تعداد پارامترهای یک لایه Simple RNN و GRU و LSTM و BiLSTM را با ذکر کامل جزئیات محاسبات به دست آورده و با هم مقایسه کنید (۱۲ نمره).

$$\text{Simple RNN} = 128 * 64 + 64 * 64 + 64 = 8192 + 4096 + 64 = 12352$$

$$\text{GRU} = 12352 * 3 = 37056$$

$$\text{LSTM} = 12352 * 4 = 49408$$

$$\text{BiLSTM} = 49408 * 2 = 98816$$



۳. برای حل هر یک از مسائل زیر، یک معماری مناسب از بین انواع معماری‌های RNN که تاکنون آموخته‌اید، انتخاب کنید. معماری‌های مورد نظر شامل One-to-Many، Many-to-One و Many-to-Many هستند. پس از انتخاب، دلیل خود را برای استفاده از آن معماری به‌طور واضح توضیح دهید (۱۸ نمره).

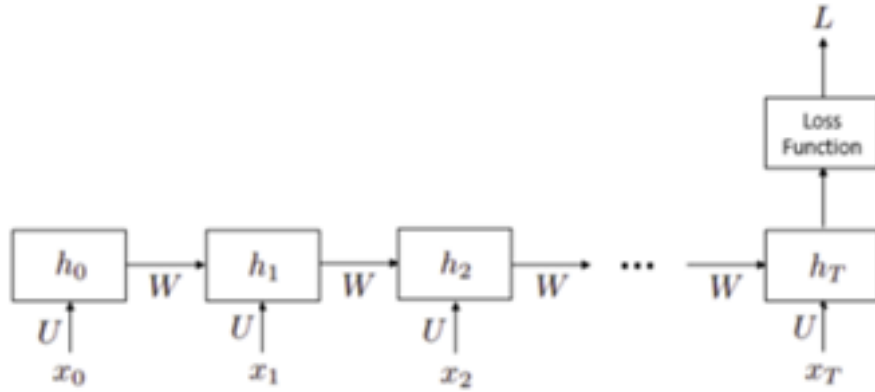
- many-to-many : Text Summarization
- many-to-many : Machine Translation
- many-to-many : Video Captioning
- many-to-one : Sentiment Analysis
- many-to-many : Automatic Speech Recognition
- many-to-many, many-to-one : Question Answering
- many-to-many : Text-to-Speech
- many-to-many : Paraphrase Generation
- many-to-many : Code Translation

سوال امتیازی



۴. با توجه به شبکه نشان داده‌شده در شکل زیر، به سوالات مطرح‌شده پاسخ دهید. برای سادگی، فرض کنید تمام مقادیر شامل ورودی‌ها، وزن‌ها و خروجی‌ها اسکالر هستند. همچنین تمام توابع فعال‌ساز را از نوع سیگموئید σ در نظر بگیرید (۲۰ نمره).

- ابتدا گرادینان h_t را بر حسب گرادینان $h_{(t+1)}$ بنویسید.



$$h_{t+1} = \sigma(Ux_{t+1} + Wh_t)$$

$$\frac{\partial L}{\partial h_t} = \frac{\partial L}{\partial h_{t+1}} \frac{\partial h_{t+1}}{\partial h_t} = \frac{\partial L}{\partial h_{t+1}} \sigma'(Ux_{t+1} + Wh_t) \cdot W$$

- حال با استفاده از رابطه قسمت قبل، و با قاعده مشتق زنجیری، گرادیان h_0 را بر حسب گرادیان h_T بنویسید.

$$\begin{aligned} \frac{\partial L}{\partial h_0} &= \frac{\partial L}{\partial h_T} \frac{\partial h_T}{\partial h_{T-1}} \cdots \frac{\partial h_2}{\partial h_1} \frac{\partial h_1}{\partial h_0} \\ &= \frac{\partial L}{\partial h_T} (\sigma'(Ux_T + Wh_{T-1}) \cdot W) \cdots (\sigma'(Ux_2 + Wh_1) \cdot W) (\sigma'(Ux_1 + Wh_0) \cdot W) \end{aligned}$$

- برش گرادیان توسط مقدار و توسط اندازه را توضیح دهید. برتری برش توسط اندازه را به برش توسط مقدار توضیح دهید.

در برش توسط مقدار ما یک آستانه مانند T در نظر گرفته و هر عنصر در بردار گرادیان که بزرگتر از T باشد را به T کاهش می‌دهیم. عناصر دیگر بدون تغییر باقی می‌مانند. در برش توسط اندازه ما نیز با یک آستانه مانند T این بار بر روی اندازه بردار گرادیان قرار داده، و اگر اندازه بردار گرادیان بیشتر از T باشد، بردار گرادیان را طوری نرمال کرده که اندازه آن برابر T شود. در برش توسط مقدار، چگونگی جهت بردار گرادیان در برخی جهات تقطیع کرده و در برخی دیگر از جهات نگه می‌داریم، این کار باعث می‌شود بردار گرادیان نهایی در جهت بردار گرادیان ابتدایی نباشد در حالی که در برش توسط اندازه چون تمام عناصر گرادیان به یک اندازه $scale$ می‌شوند، این اتفاق رخ نخواهد داد.

سوالات عملی



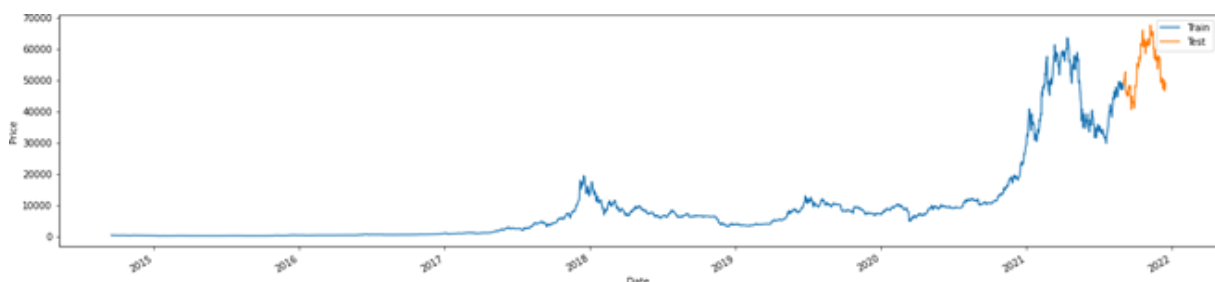
۵. به نوت‌بوک `divan_hafez.ipynb` مراجعه کرده و قسمت های مشخص شده را کامل کنید. در این نوت‌بوک قرار است با استفاده از یک RNN متن های مشابه با غزل های دیوان حافظ تولید کنید (۳۰ نمره).

در نظر داشته باشید تنها اجازه ویرایش قسمت های مشخص شده را دارید.
به نوت‌بوک `divan_hafez.ipynb` مراجعه کنید.



۶. در سالهای اخیر توجه به رمزارزها بسیار گسترش یافته است. یکی از معروفترین رمزارزهای موجود در بازار Bitcoin است. در این سوال قصد داریم قیمت Bitcoin را در آینده پیش‌بینی کنیم. برای این کار مراحل زیر را دنبال نمایید (۳۰ نمره).

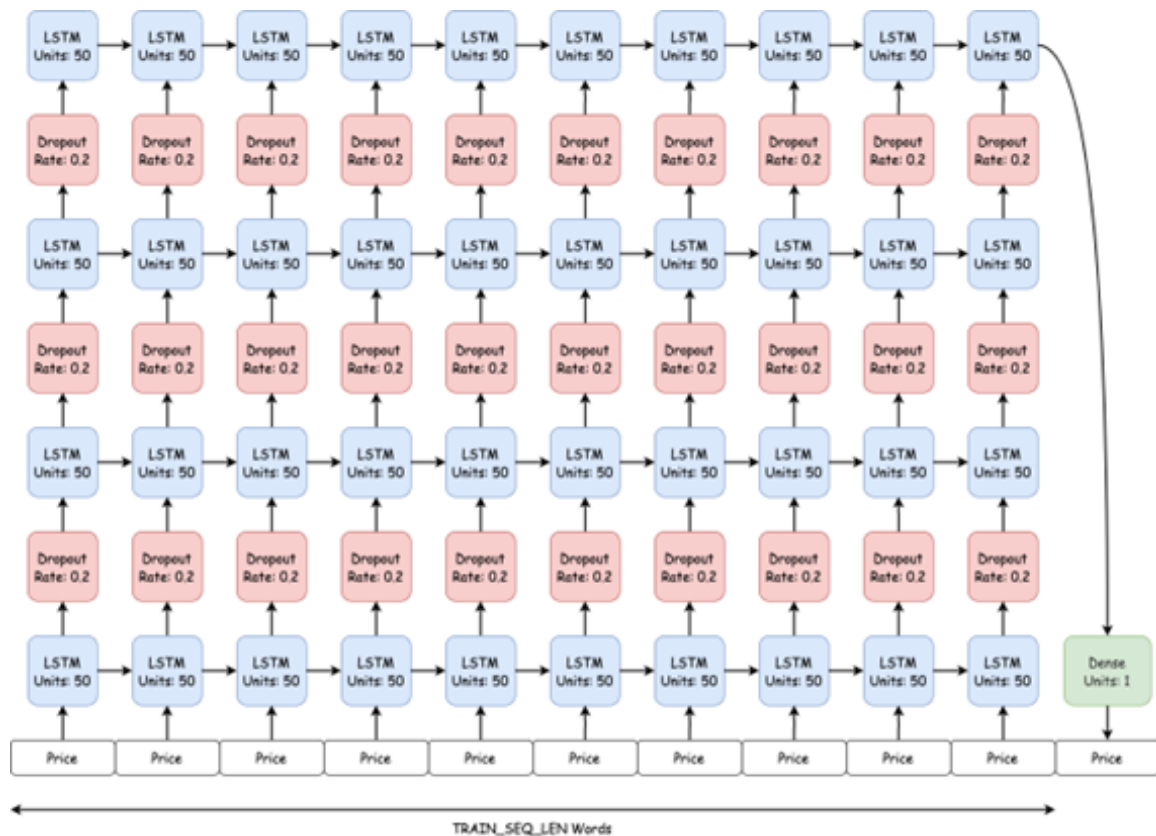
- ابتدا لازم است کتابخانه `yfinance` را نصب نمایید. این کتابخانه را میتوانید به کمک این [لینک](#) نصب نمایید.
- حال میتوانید قیمت Bitcoin را دانلود نمایید. برای این کار از تابع `download` موجود در این [لینک](#) استفاده کنید. لازم به ذکر است نماد شاخص مورد نظر برابر با `BTC-USD` است و تاریخ ذخیره‌سازی برای داده‌های آموزشی را برابر با 0.8 ابتدایی داده ها از لحاظ زمانی قرار دهید.
- برای آزمایش درستی مراحل فوق، نمودار این شاخص را بر حسب زمان رسم نمایید و به هر یک از داده‌های آموزشی و آزمایشی رنگ متفاوتی اختصاص دهید. نمودار حاصل مشابه با نمودار زیر خواهد بود.



- مقادیر محور `y` نمودار فوق را با استفاده از تابع `MinMaxScaler` کتابخانه `scikit-learn` نرمالیزه کنید. توجه داشته باشید تنها از داده‌های آموزشی برای تنظیم مقیاس استفاده کنید و داده‌های آزمون را براساس معیار داده‌های آموزش مقیاس‌شان تنظیم می‌شود.

- در مرحله بعد داده‌های مورد نیاز برای آموزش و آزمایش مدل را تهیه می‌نماییم. برای این کار متغیری تعریف کنید که نشان دهنده تعداد داده‌های گذشته برای پیش‌بینی داده مشخصی باشد. به عنوان مثال اگر این متغیر را برابر با ۶۰ قرار دهید، یعنی از ۶۰ داده گذشته در پیش‌بینی آن داده استفاده شده‌است.

- مدل را مشابه با معماری شکل زیر بسازید.



- حال این مدل را با بهینه‌ساز AdamW و تابع ضرر MSE و BatchSize=32 به تعداد ۱۰۰ اپیاک آموزش دهید.
- پس از آموزش مدل، پیش‌بینی را بر روی داده‌های آزمون انجام دهید و نمودار را رسم نمایید. در این نمودار که بر حسب زمان رسم می‌شود، هر دو مقدار واقعی و پیش‌بینی را با رنگ‌های متفاوت رسم کنید.
- در نهایت، به صورت متوالی آینده را از زمانی که داده‌ها به پایان می‌رسند پیش‌بینی کنید.
- افزایش یا کاهش متغیر تعریف شده در مرحله تهیه داده مورد نیاز برای آموزش مدل یعنی تعداد داده‌های گذشته برای پیش‌بینی داده‌های مشخص چه مزایا یا معایبی دارد؟ شرح دهید.

به نوتبوک bitco.in/bitcoin_price_prediction_lstm_pytorch.ipynb مراجعه کنید.