به نام خدا

دانشکده مهندسی برق و کامپیوتر

درس یادگیری عمیق

تمرین سری سوم

در این تمرین هدف پیاده سازی تسک question answering می باشد و در آن از شبکه های کانولوشنال و بازگشتی استفاده می شود. در این تمرین از مجموعه داده های DAQUAR (Dataset for question answering) استفاده می‌کنیم. داده ها در لینک زیر قابل مشاهده می‌باشند.

لینک داده های تصاویر:

http://datasets.d2.mpi-inf.mpg.de/mateusz14visual-turing/nyu_depth_images.tar

لینک داده های تست و آموزش و سوالات:

https://www.mpi-inf.mpg.de/departments/computer-vision-and-multimodal-computing/research/vision-and-/language/visual-turing-challenge

دقت کنید که صورت سوالات به صورت زیر هستند

What is the largest brown objects in this image3 ?

برای امبد کردن حتما باید عدد آخر آن را در نظر نگیرید مانند مثال زیر

What is the largest brown objects in this image ?

توضیح مسأله:

ورودی مسأله یک تصویر، و یک سوال مربوط به آن تصویر و۳۲ جواب می باشد و از بین۳۲ کلمه که فقط یکی جواب درست برای سوال مربوطه است جواب را باید انتخاب کنید(دقت کنید که در مجموعه داده ها برای خروجی یک یا ترکیب مشخصی از چند کلمه وجود دارد و ۳۱ کلمه دیگر اصطلاحا دیسترکتور نامیده میشوند و شما باید با استفاده از روش negative sampling انها را انتخاب کنید چند تا از کلمات دیسترکتور(مثلاً ۵) از کلمات موجود در سوال باشند) ابتدا تصویر را با عبور از یک شبکه resnet 18 انکود می کنید و همچنین کلمات سوال را نیز با یک word embedding تبدیل به بردار می کنید (در این مسأله از glove 6B استفاده کنید) سپس یک شبکه بازگشتی (Recurrent)که حالت اولیه (initial state)آن از تصویر ست می‌شود(یعنی با اعمال یک تابع خطی بر خروجیcnn) کلمات را به ترتیب می خواند و خروجی آخر آن مجددا با خروجی تصویر ترکیب (concat)می شود و با استفاده از یک لایه ی شبکه عصبی با تابع غیر خطی Relu به ۳۰۰بعد می رسد (این بعد با بعد بردار مربوط به کلمات برابر است)، سپس با یک تابع خطی دوباره آنها را به یک فضای ۳۰۰ بعدی دیگرنگاشت می کنیم و در نهایت با ضرب داخلی این بردار با تک تک بردارهای سی و دو کلمه ممکن برای جواب (که بردارهای آنها نیز از همان word2vecاولیه آمده) و اعمال تابع softmax و محاسبه cross entropy loss وزن های شبکه (که شامل وزن های word embedding و شبکه encoder عکس نیست) را تغییر می دهیم.در واقع دیکودر به صورت زیر بر روی زیر بر روی ورودی اعمال می‌شوند

Linera(Relu (linear (Concat (Rnn-last-state, cnn-features), to_dim =300)) , to_dim =300)

همچنین پارامترهای انکودر به صورت زیر می باشند

cell type = GRU ,dropout = 0 ,Bidirectional=False ,hidden_size=150 ,Num_layers =1

و پارامترهای شبکه به صورت زیر هستند:

Batch size=32; optimizer= Adam and weight decay= 0.005

جواب سوالات به صورت ترکیب یا حالتی ازقسمت های زیر است:

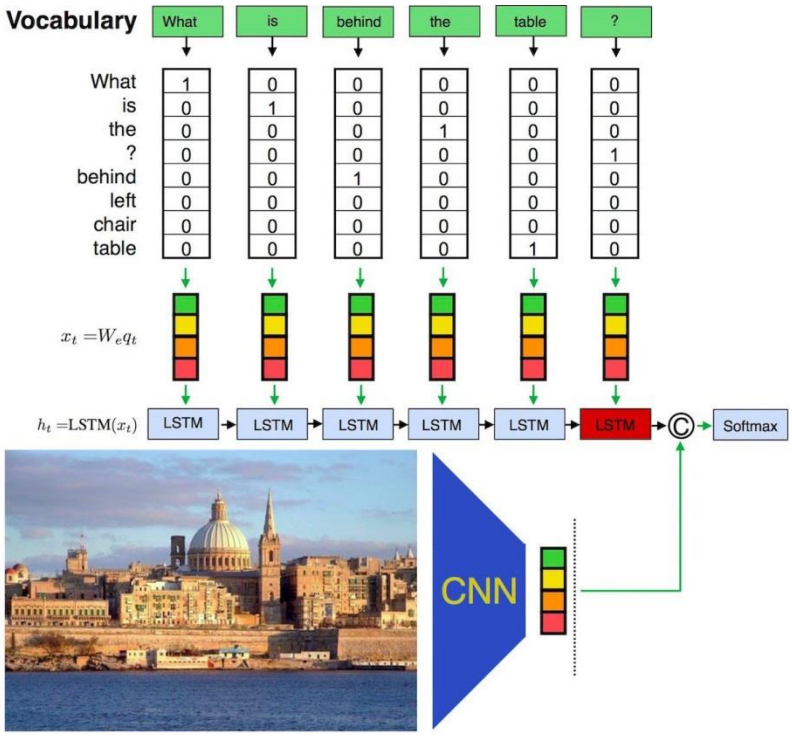
اگر جواب به صورت A_Bباشد میانگین امتیاز جواب Aو جواب Bرا درنظر بگیرید

اگر جواب به صورت A,B,... باشد یکی از جواب ها مثلا Bرا به همراه ۳۱کلمه دیگر در نظر میگیریم و دو جواب صحیح با هم را به عنوان ورودی در نظر نگیرید.

از روی مجموعه داده های اموزش یک پنجم را به عنوان داده های validationانتخاب می کنیم و بقیه داده ها را به عنوان داده های آموزش در نظر می گیریم.

برای validation metric کافی است loss را برای validation set بدست آورید وآن را رسم کنید.

نکته: دقت کنید برای کلماتی که در خارج از مجموعه کلمات هستند باید به صورت رندوم از یک توزیع خاص آنها را انتخاب کنید وتوجه داشته باشید که این بردار باید فقط یک بار انتخاب شود و دفعات بعد برای آن کلمه از همان بردار استفاده شود. تصویر شکل ۱ یک شمای کلی از مسآله را نشان میدهد.



شکل ۱

بخش اول

I. مدل گفته شده در قسمت قبل را پیاده سازی کنید

II. در مدل ارایه شده برای شبکه از مدل Resnet34 استفاده کنید و نتایج را با قسمت اول مقایسه کنید.

III. در مدل ارایه شده برای شبکه از مدل densnet121 استفاده کنید و نتایج را با قسمت اول مقایسه کنید.

IV. در مدل ارایه شده از glove42b استفاده کنید و نتایج را با قسمت اول مقایسه کنید.

V. در مدل ارایه شده برای انکودر سلول ها را به LSTM تغییر دهید و نتایج را با قسمت اول کنید.

VI. در مدل ارایه شده برای انکودر پارامتر hidden size را به 100تغییر دهید و نتایج را با قسمت اول مقایسه کنید.

VII. در مدل ارایه شده برای انکودر پارامتر num_layer را برابر 2قرار دهید و نتایج را با قسمت اول مقایسه کنید.

VIII. در مدل ارایه شده برای انکودر از ساختار دوطرفه(bidirectional) استفاده کنید و تفاوت آن با قسمت اول را توضیح دهید و نتایج را با یکدیگر مقایسه کنید.

IX. در مدل ارایه شده برای انکودر ورودی را به صورت ترکیب (concat) خروجی word2vec و خروجی خطی حاصل از Cnn و همچنین حالت اولیه را به صورت بردار zero در نظر بگیرید و نتایج را باقسمت اول مقایسه کنید.

Input =linear (Concat(Word2vec, Cnn fearture) , to_dim:300)
Initial state = zero vector

لینک های مفید:

https://medium.com/@martinpella/how-to-use-pre-trained-word-embeddings-in-pytorch-71ca59249f76

http://nlp.stanford.edu/data/glove.6B.zip

http://nlp.stanford.edu/data/glove.42B.300d.zip

نکات:

- توجه کنید که نیمی از نمره تمرین مربوط به گزارش می‌باشد. لازم به ذکر است که رعایت اصول نگارشی حائز اهمیت می‌باشد.
- در صورتی که امکان اجرای کد بر روی سیستم خود را ندارید میتوانید از google colab استفاده نمایید.
- در صورتی وجود هرگونه اشکال در اجرای کد نمره صفر برای این تمرین لحاظ خواهد شد. در صورتی که از jupyter notebook استفاده می‌کنید دقت کنید که کد شما به صورت cell به cell اجرا شود.
- گزارش تمرین را حتما به صورت PDF ارسال نمایید.
- کدهای تمرین را به همراه گزارش تمرین در سایت درس آپلود نمایید.
- نحوه نامگذاری تمرین براساس studentnumber_homeworknumber.pdf باشد .
- زبان پیاده‌سازی این تمرین python می‌باشد و از کتابخانه pytorch استفاده کنید.
- هر گونه پرسش پیرامون تمرین را با ایمیلهای sepehr.sameni@gmail.com، raminnakhli@gmail.com و esmaeilfarahng@gmail.com مکاتبه فرمایید.