



تمرین سری پنجم
درس یادگیری عمیق

نام مدرس: دکتر محمدی
دستیاران آموزشی مرتبط: حاجی زاده-محمودی
مهلت تحویل: ۱۴۰۲/۰۲/۲۳

۱- از مدل های seq-to-seq برای حل مسائل مشترک بینایی ماشین و پردازش زبان استفاده می شود، یکی از این مسائل Video Captioning می باشد که از نوع many-to-many این مدل ها استفاده می شود، در مقاله زیر برای حل این مساله از مدل های RNN در کنار مدل های CNN به صورت many-to-many استفاده شده است. نحوه حل مساله در این مقاله را توضیح دهید. آیا می توان ایده ی مطرح شده در این مقاله را برای مساله Image Captioning استفاده کرد؟ توضیح دهید. (حاجی زاده)

[CAM-RNN: Co-Attention Model Based RNN for Video Captioning](#)

۲- به سوالات زیر پاسخ دهید.

الف) اگر در GRU فقط گیت update داشته باشیم و گیت reset را حذف کنیم، چه اتفاقی می افتد و خروجی چه تغییری می کند؟

ب) فرض کنید در GRU می خواهیم خروجی هر مرحله فقط به یکی از ورودی های مرحله قبل بستگی داشته باشد. به عنوان مثال برای تولید h_t فقط از x_t استفاده کرده باشیم به طوری که $t_prime < t$ است. در این حالت مقادیر مناسب برای گیت های reset و update چه باید باشد؟ (محمودی)

۳- با توجه به مدل های زیر به سوالات پاسخ دهید. (حاجی زاده)

```
model = keras.models.Sequential()
model.add(keras.layers.Embedding(input_dim=1000, output_dim=128, input_length=200))
model.add(keras.layers.SimpleRNN(units=128, return_sequences=False))
model.add(keras.layers.Dense(units=20, activation="softmax"))
```

معماری مدل ۱

```
model2 = keras.models.Sequential()
model2.add(keras.layers.Input(shape=(None, 128)))
model2.add(keras.layers.SimpleRNN(units=128, return_sequences=False))
model2.add(keras.layers.Dense(units=20, activation="softmax"))
```

معماری مدل ۲



تمرین سری پنجم
درس یادگیری عمیق

نام مدرس: دکتر محمدی
دستیاران آموزشی مرتبط: حاجی زاده-محمودی
مهلت تحویل: ۱۴۰۲/۰۲/۲۳

الف) در مدل دوم اگر تابع فعالیت softmax به تابع sigmoid تغییر یابد تعداد عملیاتی (ضرب، جمع، تفریق و جمع) که کاهش می یابد را محاسبه کنید.

ب) اگر یک corpus در اختیار داشته باشیم و یک لایه Dense به عنوان لایه ی بعد از Input در مدل دوم اضافه کنیم ظرفیت مدل جدید (از نظر تعمیم پذیری و توانایی یادگیری) را با مدل ۱ مقایسه کنید.

۴- فایل نوت بوک پیوست شده برای حل مساله Image Captioning می باشد، در این فایل به سوالات زیر پاسخ دهید. (حاجی زاده- محمدی)

الف) بخش مربوط به DataLoader را تکمیل کنید . در این بخش با داشتن مسیر مربوط به تصاویر و فایل انوتیشن باید به موارد زیر دست بیابید.

- واژه نامه ^۱
- اندازه واژه نامه
- تمییز کردن و توکنایز کردن caption تصاویر

خروجی این بخش یک تصویر و caption متناظر خواهد بود. توابع پایه این کلاس تعریف شده اند این توابع را تکمیل کنید در صورت نیاز توابع دیگری نیز اضافه کنید.

ب) مدل را آموزش دهید و معیار های BLEU-1 و BLEU-2 را گزارش کنید. (برای [آشنایی](#) و [ماژول](#) آماده این معیار می توانید از لینک های قرار داده شده استفاده کنید)

ج) در بخش Model برای بخش RNN از یک لایه ی امبدینگ استفاده شده است وزن های آن لایه را با وزن های پیش آموخته [Golve](#) تنظیم کنید و دوباره آموزش مدل را انجام دهید. معیار BLEU-1 و BLEU-2 را برای این مرحله نیز گزارش کنید.



تمرین سری پنجم
درس یادگیری عمیق

نام مدرس: دکتر محمدی
دستیاران آموزشی مرتبط: حاجی زاده-محمودی
مهلت تحویل: ۲۳/۰۲/۱۴۰۲

(د) مرحله (ب) و (ج) را با یکدیگر مقایسه و تحلیل کنید. (ذکر حداقل ۲ مورد نیاز هست)

(ه) در مدل برای بخش استخراج ویژگی از تصاویر از یک مدل پیش آموخته resnet استفاده شده است، لایه های این مدل فریز شده اند (گرادیانی به لایه ها بازگردانده نمی شود و وزن های این مدل بروزرسانی نمی گردد)، در این بخش برخی از لایه های این مدل را از فریز در بیاورید و مدل را تنظیم مجدد^۲ نمایید و برای بخش امبدینگ هم از وزن های Glove استفاده کنید. معیار BLEU-1, BLEU-2 را در این بخش با بخش ج مقایسه کنید و تحلیل کنید. (امتیازی)

(و) در مدل های فوق (بخش های (ب)، (ج) و (ه)) Gradient Vanishing اتفاق افتاده است یا خیر؟ در صورت وجود یا عدم وجود علت را توضیح دهید.

(ز) فرض کنید ورودی مدل بازگشتی در سطح کاراکتر باشد، مزایا و معایب آن را بیان کنید.

لطفا سند قوانین انجام و تحویل تمرین های درس را مطالعه و موارد خواسته شده را رعایت فرمایید

موفق و سلامت باشید