Capstone_Project

February 14, 2021

1 Capstone Project

1.1 Neural translation model

1.1.1 Instructions

In this notebook, you will create a neural network that translates from English to German. You will use concepts from throughout this course, including building more flexible model architectures, freezing layers, data processing pipeline and sequence modelling.

This project is peer-assessed. Within this notebook you will find instructions in each section for how to complete the project. Pay close attention to the instructions as the peer review will be carried out according to a grading rubric that checks key parts of the project instructions. Feel free to add extra cells into the notebook as required.

1.1.2 How to submit

When you have completed the Capstone project notebook, you will submit a pdf of the notebook for peer review. First ensure that the notebook has been fully executed from beginning to end, and all of the cell outputs are visible. This is important, as the grading rubric depends on the reviewer being able to view the outputs of your notebook. Save the notebook as a pdf (you could download the notebook with File -> Download .ipynb, open the notebook locally, and then File -> Download as -> PDF via LaTeX), and then submit this pdf for review.

1.1.3 Let's get started!

We'll start by running some imports, and loading the dataset. For this project you are free to make further imports throughout the notebook as you wish.

[ ]:
```
import tensorflow as tf
import tensorflow_hub as hub
from tensorflow.keras.layers import Layer, Concatenate, Input, Masking, LSTM,␣
↪Embedding, Dense
from tensorflow.keras.models import Model
import unicodedata
import re
from IPython.display import Image
import numpy as np
import json
from sklearn.model_selection import train_test_split
```

```python
import matplotlib.pyplot as plt
import time
import datetime
import sys
```

For the capstone project, you will use a language dataset from http://www.manythings.org/anki/ to build a neural translation model. This dataset consists of over 200,000 pairs of sentences in English and German. In order to make the training quicker, we will restrict to our dataset to 20,000 pairs. Feel free to change this if you wish - the size of the dataset used is not part of the grading rubric.

Your goal is to develop a neural translation model from English to German, making use of a pre-trained English word embedding module.

Import the data The dataset is available for download as a zip file at the following link:

https://drive.google.com/open?id=1KczOciG7sYY7SB9UlBeRP1T9659b121Q

You should store the unzipped folder in Drive for use in this Colab notebook.

[ ]:
```python
# Run this cell to connect to your Drive folder

from google.colab import drive
drive.mount('/content/gdrive')
```

```
Drive already mounted at /content/gdrive; to attempt to forcibly remount, call
drive.mount("/content/gdrive", force_remount=True).
```

[ ]:
```python
# Run this cell to load the dataset

NUM_EXAMPLES = 125000
data_examples = []
path = 'gdrive/MyDrive/Colab Notebooks/data/deu2.txt'
with open(path, 'r', encoding='utf8') as f:
    for line in f.readlines():
        if len(data_examples) < NUM_EXAMPLES:
            data_examples.append(line)
        else:
            break
```

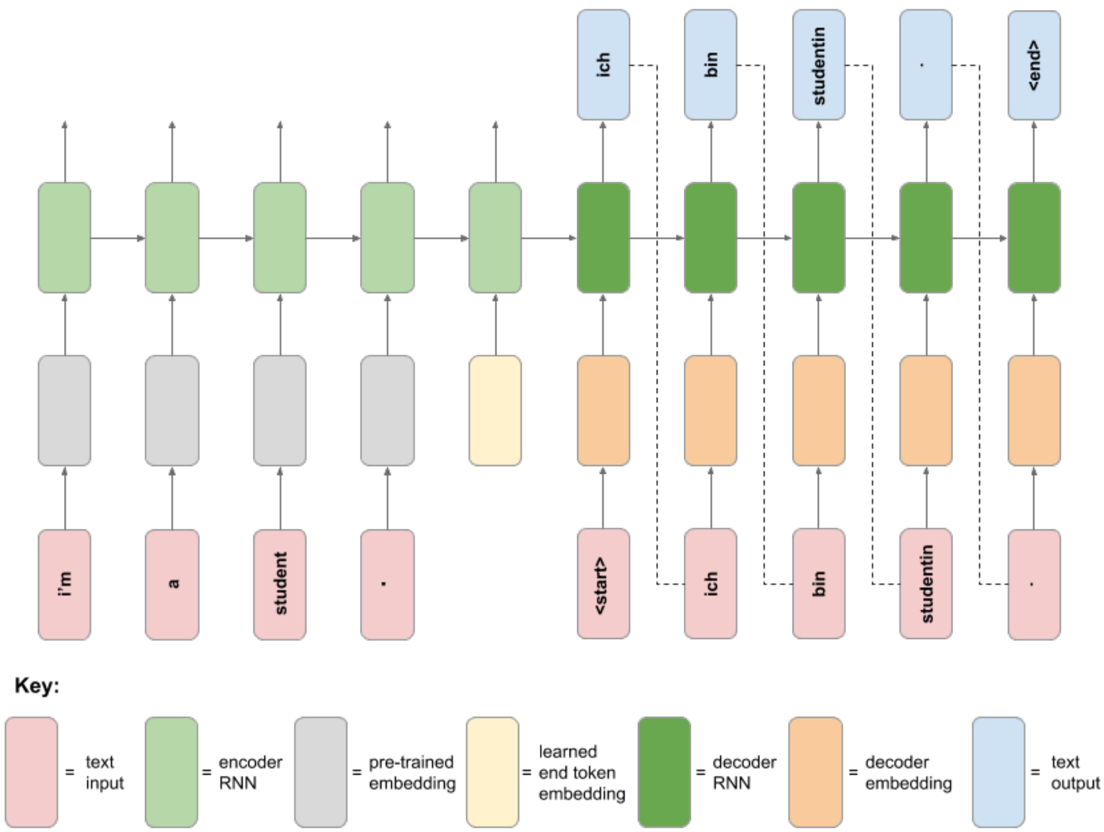
[ ]:
```python
data_examples[:3]
```

[ ]:
```
['Hi.\tHallo!\tCC-BY 2.0 (France) Attribution: tatoeba.org #538123 (CM) &
#380701 (cburgmer)\n',
 'Hi.\tGrüß Gott!\tCC-BY 2.0 (France) Attribution: tatoeba.org #538123 (CM) &
#659813 (Esperantostern)\n',
 'Run!\tLauf!\tCC-BY 2.0 (France) Attribution: tatoeba.org #906328 (papabear) &
#941078 (Fingerhut)\n']
```

```python
eng_data = []
deu_data = []

for line in data_examples:
  sentence = line.split('\t')
  eng_data.append(sentence[0])
  deu_data.append(sentence[1])
```

The custom translation model The following is a schematic of the custom translation model architecture you will develop in this project.

```python
# Run this cell to download and view a schematic diagram for the neural translation model

!wget -q -O neural_translation_model.png --no-check-certificate "https://docs.google.com/uc?export=download&id=1XsS1VlXoaEo-RbYNilJ9jcscNZvsSPmd"
Image("neural_translation_model.png")
```



The custom model consists of an encoder RNN and a decoder RNN. The encoder takes words of an English sentence as input, and uses a pre-trained word embedding to embed the words into a 128-dimensional space. To indicate the end of the input sentence, a special end token (in the same 128-dimensional space) is passed in as an input. This token is a TensorFlow Variable that is learned in the training phase (unlike the pre-trained word embedding, which is frozen).

The decoder RNN takes the internal state of the encoder network as its initial state. A start token is passed in as the first input, which is embedded using a learned German word embedding. The decoder RNN then makes a prediction for the next German word, which during inference is then passed in as the following input, and this process is repeated until the special <end> token is emitted from the decoder.

1.2 1. Text preprocessing

- Create separate lists of English and German sentences, and preprocess them using the preprocess_sentence function provided for you above.
- Add a special "<start>" and "<end>" token to the beginning and end of every German sentence.
- Use the Tokenizer class from the tf.keras.preprocessing.text module to tokenize the German sentences, ensuring that no character filters are applied. *Hint: use the Tokenizer's ``filter'' keyword argument.*
- Print out at least 5 randomly chosen examples of (preprocessed) English and German sentence pairs. For the German sentence, print out the text (with start and end tokens) as well as the tokenized sequence.
- Pad the end of the tokenized German sequences with zeros, and batch the complete set of sequences into one numpy array.

[ ]:
```
eng_data = list(map(preprocess_sentence, eng_data))
deu_data = list(map(preprocess_sentence, deu_data))
```

[ ]:
```
for i,line in enumerate(deu_data):
  string = '<start> ' + line + ' <end>'
  deu_data[i] = string

# deu_data = list(deu_data)
deu_data[:3]
```

[ ]:
```
['<start> hallo ! <end>',
 '<start> gruess gott ! <end>',
 '<start> lauf ! <end>']
```

[ ]:
```
deu_data_tk = []

for sentence in deu_data:
  deu_data_tk.append(sentence.split())

print(deu_data_tk[:3])
```

```
[['<start>', 'hallo', '!', '<end>'], ['<start>', 'gruess', 'gott', '!',
'<end>'], ['<start>', 'lauf', '!', '<end>']]
```

```
[ ]: tokenizer = tf.keras.preprocessing.text.Tokenizer(filters=None)

     tokenizer.fit_on_texts(deu_data_tk)
```

```
[ ]: deu_data_tk = tokenizer.texts_to_sequences(deu_data_tk)
```

```
[ ]: random_sample = np.random.choice(len(deu_data), 5)

     for n,i in enumerate(random_sample):
       print('Sample #', n+1)
       print('English sentence:', eng_data[i])
       print('German sentence:', deu_data[i])
       print('German sentence (tokenized):', deu_data_tk[i],'\n\n')
```

```
Sample # 1
English sentence: i'll show you that i am right .
German sentence: <start> ich werde euch beweisen , dass ich recht habe . <end>
German sentence (tokenized): [1, 4, 65, 72, 1702, 8, 28, 4, 195, 21, 3, 2]


Sample # 2
English sentence: tom has been pretending .
German sentence: <start> tom hat nur so getan , als ob . <end>
German sentence (tokenized): [1, 5, 16, 100, 54, 153, 8, 78, 344, 3, 2]


Sample # 3
English sentence: how fast were you driving ?
German sentence: <start> wie schnell bist du gefahren ? <end>
German sentence (tokenized): [1, 29, 213, 56, 12, 841, 6, 2]


Sample # 4
English sentence: she turned off the lights .
German sentence: <start> sie loeschte die lichter . <end>
German sentence (tokenized): [1, 10, 5677, 18, 3389, 3, 2]


Sample # 5
English sentence: you should have worked harder .
German sentence: <start> du haettest dir bei der arbeit mehr muehe geben sollen
! <end>
German sentence (tokenized): [1, 12, 501, 46, 123, 25, 138, 85, 677, 223, 236,
17, 2]
```

```
[ ]: deu_data_tk = tf.keras.preprocessing.sequence.pad_sequences(deu_data_tk,␣
     ↪padding='post')
```

1.3 2. Prepare the data

Load the embedding layer As part of the dataset preproceessing for this project, you will use a pre-trained English word embedding module from TensorFlow Hub. The URL for the module is https://tfhub.dev/google/tf2-preview/nnlm-en-dim128-with-normalization/1.

This embedding takes a batch of text tokens in a 1-D tensor of strings as input. It then embeds the separate tokens into a 128-dimensional space.

The code to load and test the embedding layer is provided for you below.

NB: this model can also be used as a sentence embedding module. The module will process each token by removing punctuation and splitting on spaces. It then averages the word embeddings over a sentence to give a single embedding vector. However, we will use it only as a word embedding module, and will pass each word in the input sentence as a separate token.

```
[ ]: # Load embedding module from Tensorflow Hub

     embedding_layer = hub.KerasLayer("https://tfhub.dev/google/tf2-preview/
     ↪nnlm-en-dim128/1",
                                      output_shape=[128], input_shape=[], dtype=tf.
     ↪string)
```

```
[ ]: # Test the layer

     embedding_layer(tf.constant(["these", "aren't", "the", "droids", "you're",␣
     ↪"looking", "for"])).shape
```

```
[ ]: TensorShape([7, 128])
```

You should now prepare the training and validation Datasets.

- Create a random training and validation set split of the data, reserving e.g. 20% of the data for validation (NB: each English dataset example is a single sentence string, and each German dataset example is a sequence of padded integer tokens).
- Load the training and validation sets into a tf.data.Dataset object, passing in a tuple of English and German data for both training and validation sets.
- Create a function to map over the datasets that splits each English sentence at spaces. Apply this function to both Dataset objects using the map method. *Hint: look at the tf.strings.split function.*
- Create a function to map over the datasets that embeds each sequence of English words using the loaded embedding layer/model. Apply this function to both Dataset objects using the map method.

- Create a function to filter out dataset examples where the English sentence is greater than or equal to than 13 (embedded) tokens in length. Apply this function to both Dataset objects using the filter method.
- Create a function to map over the datasets that pads each English sequence of embeddings with some distinct padding value before the sequence, so that each sequence is length 13. Apply this function to both Dataset objects using the map method. *Hint: look at the tf.pad function. You can extract a Tensor shape using tf.shape; you might also find the tf.math.maximum function useful.*
- Batch both training and validation Datasets with a batch size of 16.
- Print the element_spec property for the training and validation Datasets.
- Using the Dataset .take(1) method, print the shape of the English data example from the training Dataset.
- Using the Dataset .take(1) method, print the German data example Tensor from the validation Dataset.

[ ]:
```
X_train, X_val, y_train, y_val = train_test_split(eng_data, deu_data_tk,␣
↪test_size=0.2)
```

[ ]:
```
train_dataset = tf.data.Dataset.from_tensor_slices((X_train, y_train))
val_dataset = tf.data.Dataset.from_tensor_slices((X_val, y_val))
```

[ ]:
```
def map_string_split(eng, deu):
  eng = tf.strings.split(eng, sep=' ')
  return eng, deu
```

[ ]:
```
train_dataset = train_dataset.map(map_string_split)
val_dataset = val_dataset.map(map_string_split)
```

[ ]:
```
def map_embeddings(eng, deu):
  eng = embedding_layer(eng)
  return eng, deu
```

[ ]:
```
train_dataset = train_dataset.map(map_embeddings)
val_dataset = val_dataset.map(map_embeddings)
```

[ ]:
```
def filter_fn(eng,deu):
    return (tf.shape(eng)[0]) < 13
```

[ ]:
```
train_dataset = train_dataset.filter(filter_fn)
val_dataset = val_dataset.filter(filter_fn)
```

[ ]:
```
def map_to_pad(eng, deu):
  size = tf.shape(eng)[0]
  lacking = 13 - size
  padded = tf.pad(eng,
                  paddings=[[lacking,0],[0,0]],
                  mode='CONSTANT', constant_values=0.0)
```

```
    return padded, deu
```

[ ]:
```
train_dataset = train_dataset.map(map_to_pad)
val_dataset = val_dataset.map(map_to_pad)
```

[ ]:
```
train_dataset = train_dataset.batch(16, drop_remainder=True)
val_dataset = val_dataset.batch(16, drop_remainder=True)
```

[ ]:
```
print(train_dataset.element_spec)
print(val_dataset.element_spec)
```

```
(TensorSpec(shape=(16, None, 128), dtype=tf.float32, name=None),
TensorSpec(shape=(16, 23), dtype=tf.int32, name=None))
(TensorSpec(shape=(16, None, 128), dtype=tf.float32, name=None),
TensorSpec(shape=(16, 23), dtype=tf.int32, name=None))
```

[ ]:
```
print(next(iter(train_dataset.take(1)))[0].shape) # First eng training example␣
 ↪shape
```

```
(16, 13, 128)
```

[ ]:
```
print(next(iter(val_dataset.take(1)))[1].shape) # First deu validating example␣
 ↪shape
```

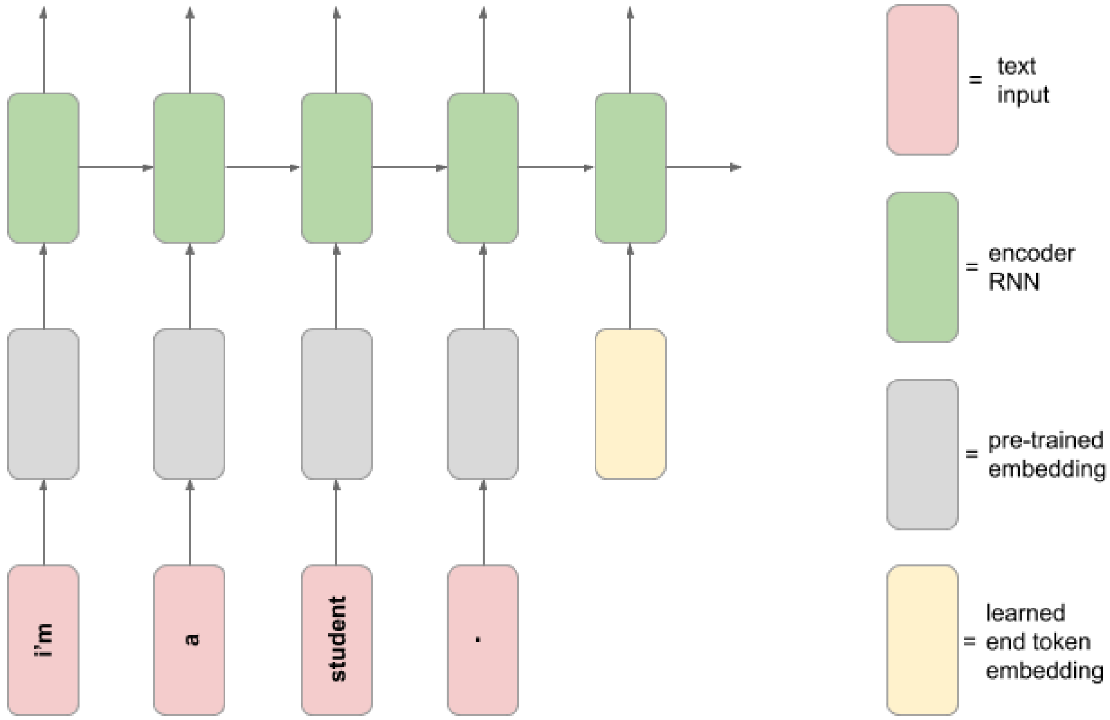
```
(16, 23)
```

1.4 3. Create the custom layer

You will now create a custom layer to add the learned end token embedding to the encoder model:

[ ]:
```
# Run this cell to download and view a schematic diagram for the encoder model

!wget -q -O neural_translation_model.png --no-check-certificate "https://docs.
 ↪google.com/uc?export=download&id=1JrtNOzUJDaOWrK4C-xv-4wUuZaI12sQI"
Image("neural_translation_model.png")
```

[ ]:



You should now build the custom layer. * Using layer subclassing, create a custom layer that takes a batch of English data examples from one of the Datasets, and adds a learned embedded `end' token to the end of each sequence. * This layer should create a TensorFlow Variable (that will be learned during training) that is 128-dimensional (the size of the embedding space). *Hint: you may find it helpful in the call method to use the tf.tile function to replicate the end token embedding across every element in the batch.* * Using the Dataset .take(1) method, extract a batch of English data examples from the training Dataset and print the shape. Test the custom layer by calling the layer on the English data batch Tensor and print the resulting Tensor shape (the layer should increase the sequence length by one).

[ ]:
```
class AddEndToken(Layer):

  def __init__(self):
    super(AddEndToken, self).__init__()
    self.emb = tf.Variable(initial_value=tf.random.uniform(shape=(128,)),
                           trainable=True)

  def call(self, inputs):
    x = tf.tile(tf.reshape(self.emb, (1, 1, -1)), [tf.shape(inputs)[0],1, 1])
    # This creates a tensor of the repeated learned embedding of size (batch␣
↪size, 1)
    # Now we have to concatenate this tensor at the right of our batch
```

```
    return Concatenate(axis = 1)([inputs, x])

end_tk_layer = AddEndToken()
```

[ ]:
```
eng_batch, _ = next(iter(train_dataset.take(1)))
print(eng_batch.shape) # Shape of the first batch of eng training examples
```

```
(16, 13, 128)
```

[ ]:
```
modified_batch = end_tk_layer(eng_batch)
```

[ ]:
```
print(modified_batch.shape)
```

```
(16, 14, 128)
```

1.5 4. Build the encoder network

The encoder network follows the schematic diagram above. You should now build the RNN encoder model. * Using the functional API, build the encoder network according to the following spec: * The model will take a batch of sequences of embedded English words as input, as given by the Dataset objects. * The next layer in the encoder will be the custom layer you created previously, to add a learned end token embedding to the end of the English sequence. * This is followed by a Masking layer, with the mask_value set to the distinct padding value you used when you padded the English sequences with the Dataset preprocessing above. * The final layer is an LSTM layer with 512 units, which also returns the hidden and cell states. * The encoder is a multi-output model. There should be two output Tensors of this model: the hidden state and cell states of the LSTM layer. The output of the LSTM layer is unused. * Using the Dataset .take(1) method, extract a batch of English data examples from the training Dataset and test the encoder model by calling it on the English data Tensor, and print the shape of the resulting Tensor outputs. * Print the model summary for the encoder network.

[ ]:
```
def encoder():
  inputs = Input(batch_shape=(None, 13, 128))
  x = end_tk_layer(inputs)
  x = Masking(mask_value=0.0)(x)
  x, hidden_state, cell_state = LSTM(512, return_state=True)(x)
  model = Model(inputs = inputs, outputs = [hidden_state, cell_state])
  return model
```

[ ]:
```
EncoderModule = encoder()
```

[ ]:
```
EncoderModule.summary()
```

```
Model: "model"

_________________________________________________________________
Layer (type)                 Output Shape              Param #
```

```
=================================================================
input_1 (InputLayer)         [(None, 13, 128)]         0
_________________________________________________________________
add_end_token (AddEndToken)  (None, 14, 128)           128
_________________________________________________________________
masking (Masking)            (None, 14, 128)           0
_________________________________________________________________
lstm (LSTM)                  [(None, 512), (None, 512) 1312768
=================================================================
Total params: 1,312,896
Trainable params: 1,312,896
Non-trainable params: 0
_________________________________________________________________
```

[ ]:
```
eng_batch, _ = next(iter(train_dataset.take(1)))
```

[ ]:
```
print('Input shape: ', eng_batch.shape)
```

```
Input shape:  (16, 13, 128)
```

[ ]:
```
eng_batch, _ = next(iter(train_dataset.take(1)))
output = EncoderModule(eng_batch)
```

[ ]:
```
for i,tensor in enumerate(output):
  print('Output tensor #', str(i+1), '-->', tf.shape(tensor))
```

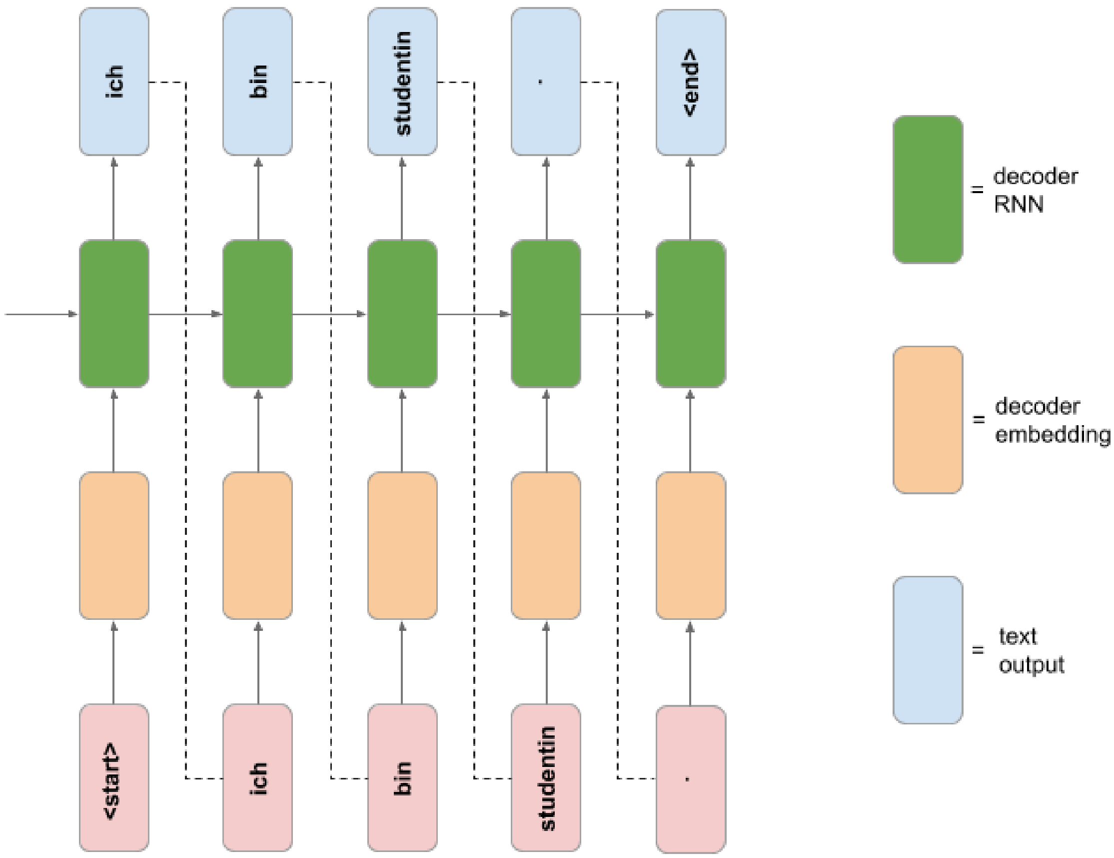
```
Output tensor # 1 --> tf.Tensor([ 16 512], shape=(2,), dtype=int32)
Output tensor # 2 --> tf.Tensor([ 16 512], shape=(2,), dtype=int32)
```

1.6 5. Build the decoder network

The decoder network follows the schematic diagram below.

[ ]:
```
# Run this cell to download and view a schematic diagram for the decoder model

!wget -q -O neural_translation_model.png --no-check-certificate "https://docs.
↪google.com/uc?export=download&id=1DTeaXD8tA8RjkpVrB2mr9csSBOY4LQiW"
Image("neural_translation_model.png")
```

[ ]:



You should now build the RNN decoder model. * Using Model subclassing, build the decoder network according to the following spec: * The initializer should create the following layers: * An Embedding layer with vocabulary size set to the number of unique German tokens, embedding dimension 128, and set to mask zero values in the input. * An LSTM layer with 512 units, that returns its hidden and cell states, and also returns sequences. * A Dense layer with number of units equal to the number of unique German tokens, and no activation function. * The call method should include the usual inputs argument, as well as the additional keyword arguments hidden_state and cell_state. The default value for these keyword arguments should be None. * The call method should pass the inputs through the Embedding layer, and then through the LSTM layer. If the hidden_state and cell_state arguments are provided, these should be used for the initial state of the LSTM layer. *Hint: use the* initial_state *keyword argument when calling the LSTM layer on its input.* * The call method should pass the LSTM output sequence through the Dense layer, and return the resulting Tensor, along with the hidden and cell states of the LSTM layer. * Using the Dataset .take(1) method, extract a batch of English and German data examples from the training Dataset. Test the decoder model by first calling the encoder model on the English data Tensor to get the hidden and cell states, and then call the decoder model on the German data Tensor and hidden and cell states, and print the shape of the resulting decoder Tensor outputs. * Print the model summary for the decoder network.

[ ]:
```python
unique_words = len(tokenizer.word_index) + 1
```

[ ]:
```python
class RNNDecoder(Model):

  def __init__(self, **kwargs):
    super(RNNDecoder, self).__init__(**kwargs)
    self.embedding_layer = Embedding(input_dim=unique_words, output_dim=128,␣
↪mask_zero=True)
    self.lstm_layer = LSTM(512, return_state=True, return_sequences=True)
    self.dense_layer = Dense(units=unique_words)

  def call(self, inputs, hidden_state=None, cell_states=None):
    x = self.embedding_layer(inputs)
    if (hidden_state is not None) and (cell_states is not None):
      x, h_state, c_state = self.lstm_layer(x, initial_state=[hidden_state,␣
↪cell_states])
    else:
      x, h_state, c_state = self.lstm_layer(x)
    return self.dense_layer(x), h_state, c_state

DecoderModule = RNNDecoder()
```

[ ]:
```python
eng_batch, deu_batch = next(iter(train_dataset.take(1)))
hidden_state, cell_states = EncoderModule(eng_batch)
output_tensors = DecoderModule(deu_batch, hidden_state, cell_states)

for i,tensor in enumerate(output_tensors):
  print('Output tensor #', str(i+1), '-->', tf.shape(tensor))
```

```
Output tensor # 1 --> tf.Tensor([   16    23 20242], shape=(3,), dtype=int32)
Output tensor # 2 --> tf.Tensor([ 16 512], shape=(2,), dtype=int32)
Output tensor # 3 --> tf.Tensor([ 16 512], shape=(2,), dtype=int32)
```

[ ]:
```python
DecoderModule.summary()
```

```
Model: "rnn_decoder"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding (Embedding)        multiple                  2590976
_________________________________________________________________
lstm_1 (LSTM)                multiple                  1312768
_________________________________________________________________
dense (Dense)                multiple                  10384146
=================================================================
Total params: 14,287,890
Trainable params: 14,287,890
Non-trainable params: 0
```

1.7 6. Make a custom training loop

You should now write a custom training loop to train your custom neural translation model. * Define a function that takes a Tensor batch of German data (as extracted from the training Dataset), and returns a tuple containing German inputs and outputs for the decoder model (refer to schematic diagram above). * Define a function that computes the forward and backward pass for your translation model. This function should take an English input, German input and German output as arguments, and should do the following: * Pass the English input into the encoder, to get the hidden and cell states of the encoder LSTM. * These hidden and cell states are then passed into the decoder, along with the German inputs, which returns a sequence of outputs (the hidden and cell state outputs of the decoder LSTM are unused in this function). * The loss should then be computed between the decoder outputs and the German output function argument. * The function returns the loss and gradients with respect to the encoder and decoder's trainable variables. * Decorate the function with @tf.function * Define and run a custom training loop for a number of epochs (for you to choose) that does the following: * Iterates through the training dataset, and creates decoder inputs and outputs from the German sequences. * Updates the parameters of the translation model using the gradients of the function above and an optimizer object. * Every epoch, compute the validation loss on a number of batches from the validation and save the epoch training and validation losses. * Plot the learning curves for loss vs epoch for both training and validation sets.

Hint: This model is computationally demanding to train. The quality of the model or length of training is not a factor in the grading rubric. However, to obtain a better model we recommend using the GPU accelerator hardware on Colab.

[ ]:
```
def preprocess_deu(deu_batch):
  start_token = tokenizer.word_index.get('<start>')
  end_token = tokenizer.word_index.get('<end>')

  deu_input = tf.where(deu_batch == end_token, x=0, y=deu_batch)[:, :-1]
  deu_output = tf.where(deu_batch == start_token, x=0, y=deu_batch)[:, 1:]

  return (deu_input, deu_output)
```

[ ]:
```
opt = tf.keras.optimizers.Adam()
loss_fn = tf.keras.losses.SparseCategoricalCrossentropy(from_logits=True)
all_trainable_variables = EncoderModule.trainable_variables+DecoderModule.
 ↪trainable_variables
```

[ ]:
```
# @tf.function
def compute_grads(eng_input, deu_input, deu_output):

  with tf.GradientTape() as tape:
```

```
    hidden_s, cell_s = EncoderModule(eng_input)
    seq_output, _, _ = DecoderModule(deu_input, hidden_s, cell_s)
    batch_loss = loss_fn(deu_output, seq_output)
    grads = tape.gradient(batch_loss, all_trainable_variables)

  return batch_loss, grads
```

[ ]:
```
def custom_checkpoint():
  EncoderModule.save_weights('gdrive/MyDrive/Colab Notebooks/data/encoder',␣
↪overwrite=True, save_format='h5')
  DecoderModule.save_weights('gdrive/MyDrive/Colab Notebooks/data/decoder',␣
↪overwrite=True, save_format='h5')
```

[ ]:
```
def training(train_dataset, val_dataset, loss_fn, compute_grads, opt,␣
↪num_epochs=5, verbose=True):

  train_loss = []
  val_loss = []
  training_start_time = time.time()

  print("Starting training...\n")


  for epoch in range(num_epochs):

    if verbose:
      print('--------- Epoch #{:02d} ---------'.format(epoch+1))

    epoch_start_time = time.time()
    epoch_loss_mean = tf.keras.metrics.Mean()
    val_epoch_loss_mean = tf.keras.metrics.Mean()

    for eng, deu in train_dataset:

      if verbose:
        duration = int(time.time() - epoch_start_time)
        conversion = datetime.timedelta(seconds=duration)
        sys.stdout.write("\r" + 'Duration: ' + str(conversion))
        sys.stdout.flush()

      deu_input, deu_output = preprocess_deu(deu)
      batch_loss, grads = compute_grads(eng, deu_input, deu_output)
      opt.apply_gradients(zip(grads, all_trainable_variables))
      epoch_loss_mean(batch_loss)


    for eng, deu in val_dataset:
```

```
    if verbose:
      duration = int(time.time() - epoch_start_time)
      conversion = datetime.timedelta(seconds=duration)
      sys.stdout.write("\r" + 'Duration: ' + str(conversion))
      sys.stdout.flush()

    deu_input, deu_output = preprocess_deu(deu)
    # We won't call the compute_grads functions to obtein the val loss
    # because this would be computationally expensive and we just need
    # to do the forward prop to get the loss

    #forward prop:
    hidden_s, cell_s = EncoderModule(eng)
    seq_output, _, _ = DecoderModule(deu_input, hidden_s, cell_s)
    batch_loss = loss_fn(deu_output, seq_output)
    val_epoch_loss_mean(batch_loss)

  train_loss.append(epoch_loss_mean.result().numpy())
  val_loss.append(val_epoch_loss_mean.result().numpy())
  sys.stdout.flush()


  if verbose:
    print("\nTraining loss: {:.3f}".format(epoch_loss_mean.result()))
    print("Validation loss: {:.3f}\n\n".format(val_epoch_loss_mean.result()))

  if epoch > 0:
    val_change_ratio = val_loss[epoch]/val_loss[epoch-1]
    val_train_ratio = val_epoch_loss_mean.result()/epoch_loss_mean.result()

    if val_change_ratio > 0.95 or val_train_ratio > 1.6:
      print('Early Stopping Callback activated!')
      if verbose:
        duration = int(time.time() - training_start_time)
        conversion = datetime.timedelta(seconds=duration)
        print('\n\nTraining completed!')
        print('\nTotal duration: ' + str(conversion))

      history = (train_loss, val_loss)

      return history

    else:
      custom_checkpoint()
      # If the val_loss/train_loss is < 1.6 it will save the weights
```

```
    if verbose:
      duration = int(time.time() - training_start_time)
      conversion = datetime.timedelta(seconds=duration)
      print('\n\nTraining completed!')
      print('\nTotal duration: ' + str(conversion))

    history = (train_loss, val_loss)

    return history
```

[ ]:
```
history = training(train_dataset, val_dataset, loss_fn, compute_grads, opt,␣
 ↪num_epochs=5, verbose=True)
```

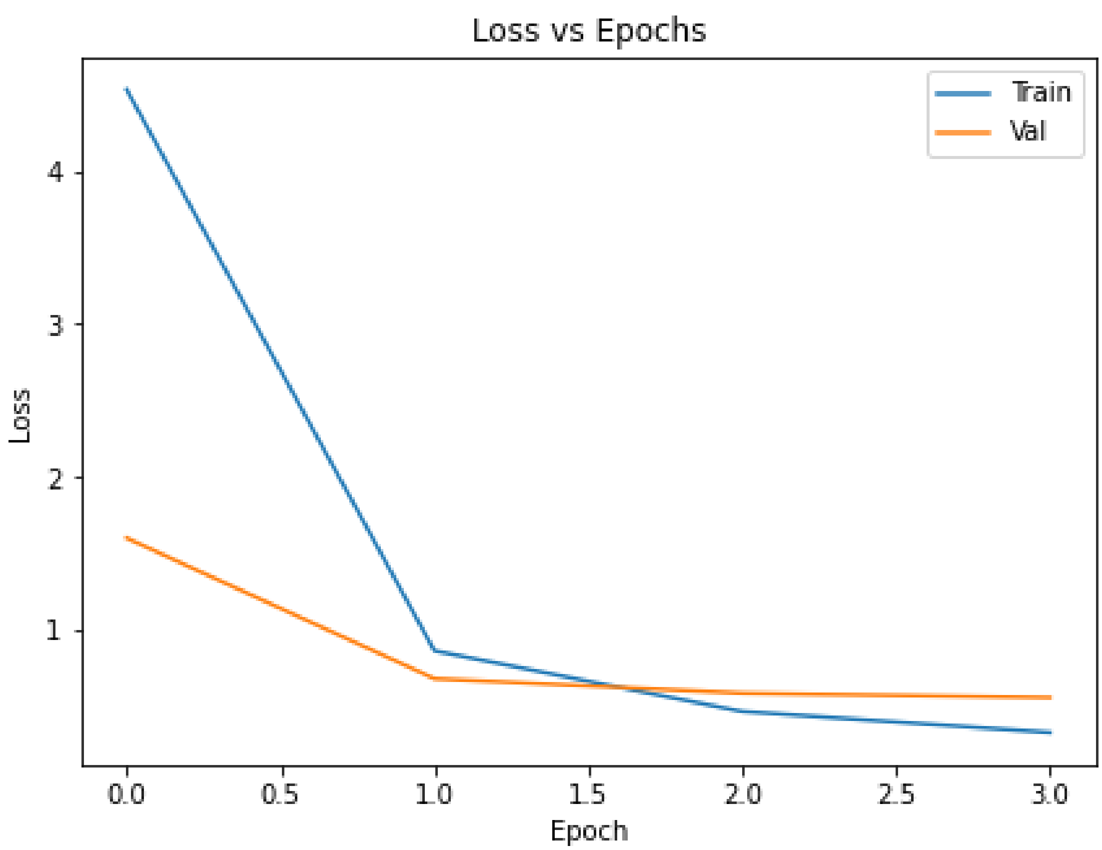
```
Starting training…

--------- Epoch #01 ---------
Duration: 0:17:32
Training loss: 4.532
Validation loss: 1.603


--------- Epoch #02 ---------
Duration: 0:17:30
Training loss: 0.866
Validation loss: 0.684


--------- Epoch #03 ---------
Duration: 0:17:23
Training loss: 0.470
Validation loss: 0.590


--------- Epoch #04 ---------
Duration: 0:17:09
Training loss: 0.333
Validation loss: 0.562


Early Stopping Callback activated!


Training completed!

Total duration: 1:09:37
```

[ ]:

```python
# In case of Early Stopping, we load the weights saved before the overfitting:

EncoderModule.load_weights('gdrive/MyDrive/Colab Notebooks/data/encoder')
DecoderModule.load_weights('gdrive/MyDrive/Colab Notebooks/data/decoder')
```

[ ]:

```python
train_loss, val_loss = history

plt.figure(figsize=(7, 5))

plt.plot(train_loss)
plt.plot(val_loss)

plt.title('Loss vs Epochs')

plt.legend(['Train', 'Val'])
plt.ylabel('Loss')
plt.xlabel('Epoch')

plt.show()
```



1.8 7. Use the model to translate

Now it's time to put your model into practice! You should run your translation for five randomly sampled English sentences from the dataset. For each sentence, the process is as follows: * Preprocess and embed the English sentence according to the model requirements. * Pass the embedded sentence through the encoder to get the encoder hidden and cell states. * Starting with the special "<start>" token, use this token and the final encoder hidden and cell states to get the one-step prediction from the decoder, as well as the decoder's updated hidden and cell states. * Create a loop to get the next step prediction and updated hidden and cell states from the decoder, using the most recent hidden and cell states. Terminate the loop when the "<end>" token is emitted, or when the sentence has reached a maximum length. * Decode the output token sequence into German text and print the English text and the model's German translation.

[ ]:
```
random_sample = np.random.choice(len(deu_data), 5)

sample_eng = []
sample_deu = []
sentence_maxlen = train_dataset.element_spec[1].shape[1]

for i in random_sample:
  sample_eng.append(eng_data[i])
  sample_deu.append(' '.join(deu_data[i].split(' ')[1:-1]))

# English samples preprocessing
sample_eng_tensor = list(map(lambda x: tf.strings.split(x), sample_eng))
sample_eng_tensor = list(map(lambda x: embedding_layer(x), sample_eng_tensor))
sample_eng_tensor = list(map(lambda x: tf.pad(x, paddings=[[13-x.shape[0],␣
↪0],[0, 0]]),sample_eng_tensor))
sample_eng_tensor = list(map(lambda x: tf.expand_dims(x, axis=0),␣
↪sample_eng_tensor))

print('TRANSLATIONS:\n')

for n, eng_sentence in enumerate(sample_eng_tensor):

  translation = []
  hidden_state, cell_state = EncoderModule(eng_sentence)
  input_token = tf.Variable([[tokenizer.word_index['<start>']]])
  for m in range(sentence_maxlen):
    seq_output, hidden_state, cell_state = DecoderModule(input_token,␣
↪hidden_state, cell_state)
    prediction_index = int(np.argmax(seq_output.numpy(), axis=2))
    word = tokenizer.index_word[prediction_index]

    if word == '<end>':
      break
```

```
    else:
      input_token = tf.Variable([[prediction_index]])
      translation.append(word)

  translation = ' '.join(translation)
  print('\n\n~~~~~~~~~~~~~~~~~~~~~~ Sample #', n+1, '~~~~~~~~~~~~~~~~~~~~~~')
  print('English sentence:', sample_eng[n])
  print('German sentence (actual):', sample_deu[n])
  print('Translation (predicted):',translation)
```

```
TRANSLATIONS:



~~~~~~~~~~~~~~~~~~~~~~ Sample # 1 ~~~~~~~~~~~~~~~~~~~~~~
English sentence: the food is delicious .
German sentence (actual): das essen ist koestlich .
Translation (predicted): das essen ist koestlich .


~~~~~~~~~~~~~~~~~~~~~~ Sample # 2 ~~~~~~~~~~~~~~~~~~~~~~
English sentence: only fifty people came .
German sentence (actual): nur fuenfzig leute waren gekommen .
Translation (predicted): nur der beiden waren er gekommen .


~~~~~~~~~~~~~~~~~~~~~~ Sample # 3 ~~~~~~~~~~~~~~~~~~~~~~
English sentence: take the road on the left .
German sentence (actual): nehmen sie die strasse links .
Translation (predicted): nehmen sie die strasse links !


~~~~~~~~~~~~~~~~~~~~~~ Sample # 4 ~~~~~~~~~~~~~~~~~~~~~~
English sentence: i don't care what happens .
German sentence (actual): es ist mir gleich , was passiert .
Translation (predicted): es ist mir egal , was es geht .


~~~~~~~~~~~~~~~~~~~~~~ Sample # 5 ~~~~~~~~~~~~~~~~~~~~~~
English sentence: tom is good at this .
German sentence (actual): tom kann das gut .
Translation (predicted): tom ist gut darin .
```