COMP6248 Lab 7 Exercise – Transforming Sequences

Wei Chien Teoh (Eugene)
wct1c16@soton.ac.uk

29 April 2021

Introduction

The results are seeded using pytorch_lightning.seed_everything(0) to provide reproducible results.

1 Sequence-to-Sequence Modelling

1.1 Complete and train a sequence-to-sequence model

Listing 1: Encoder class forward method.

```
def forward(self, src):
    # TODO
    embedded = self.embedding(src)
    output, (hidden, cell) =
        self.rnn(embedded) # initial
        (hidden, cell) defaults to zero
    return hidden, cell
```

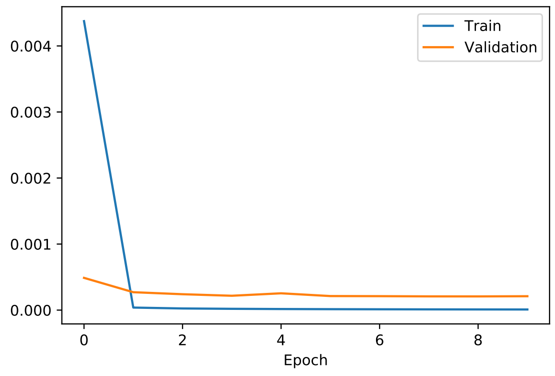

Figure 1: Loss curves for seq2seq model. Runtime: 9m 23s

1.2 Now use it!

Listing 2: Decoded output.

```
answer the following
why is the order of the output reversed
what is the point of teacher forcing
```

1.2.1 Why is the order of the output reversed?

In the seq2seq paper, it is discussed that reversing the source sentences results in better performance and memory utilization. This is possibly due to a "stronger communication link", where the first few words of the source sentences are closer to the first few words of the target sentences. The effects of reversing the output order achieves the same intuitive reasoning to reversing the source order. Listing 3 shows a simple example for the intuition of this phenomenon.[1]

Listing 3: Effects of reversed output order.

```
# original output order
A B C -> alpha beta gamma

# reversed output order
# stronger link between (C, gamma), (B, beta)
A B C -> gamma beta alpha
```

[1] *Nlp - Why Do We Reverse Input When Feeding in Seq2seq Model in Tensorflow(Tf.Reverse(Inputs,[-1]))*. Stack Overflow. URL: https://stackoverflow.com/questions/51003992/why-do-we-reverse-input-when-feeding-in-seq2seq-model-in-tensorflow-tf-reverse (visited on 04/28/2021).

1.2.2 What is the point of teacher forcing?

Teacher forcing[2] is a training strategy for RNN that uses ground truth as an input for the next time step. In sequence modelling, traditional MLL training that uses predicted output as input to the next sequence might provide large loss, which might result in slow convergence or stuck in bad local minima. Teacher forcing allows faster training. However, an extensive amount of teacher forcing might cause small prediction compound in the conditioning context.[3] Hence, in this model, a ratio (defaulting to 50%) is introduced to control the percentage of which teacher forcing occurs.

1.3 Sequence Lengths

The original decode function is modified to work with three set of code in one chunk, shown in Listing 4. Listing 5 shows the output of the function. The output shows missing letters in the start and end of the sentence. This suggests that the model failed to capture and predict long-term structure.

Although LSTMs are developed to better capture long-term structure, it still fails to perfectly achieve it.

The maximum sequence length in the training data is only 6. The short training sequence length is another issue of why the model does not have good performance on larger chunks.

Listing 4: Modified decode function with longer chunks.

```
def decode_long(code):
    out = ''

    # 3 spaces per element
    parts = code.split('␣')
    parts = [parts[i] + '␣' + parts[i+1] + '␣
        ' + parts[i+2] for i in
        range(len(parts)-3)]

    for chunk in parts:
        num = ds.encode_morse('^␣' + chunk +
            '␣$').unsqueeze(1)
        pred = model(num.cuda(), maxlen=2)
        pred = pred[1:].view(-1,
            pred_dim).argmax(-1)
        out +=
            ds.decode_alpha(pred.cpu())[::-1]
    return out
```

Listing 5: Output of decode_long.

```
code:
    .- -. ... .-- . .-. / - .... . / ..-.
        --- .-.. .-.. --- .-- .. -. --.
decode:
    answer the following
decode_long:
    swer the followin
```

[2] Ronald J. Williams and David Zipser. "A Learning Algorithm for Continually Running Fully Recurrent Neural Networks". In: *Neural Computation* 1.2 (June 1989), pp. 270–280. ISSN: 0899-7667. DOI: 10.1162/neco.1989.1.2.270.

[3] Alex Lamb et al. *Professor Forcing: A New Algorithm for Training Recurrent Networks*. Oct. 27, 2016. arXiv: 1610.09038 [cs, stat]. URL: http://arxiv.org/abs/1610.09038 (visited on 04/28/2021).