

Sequence to Sequence modelling

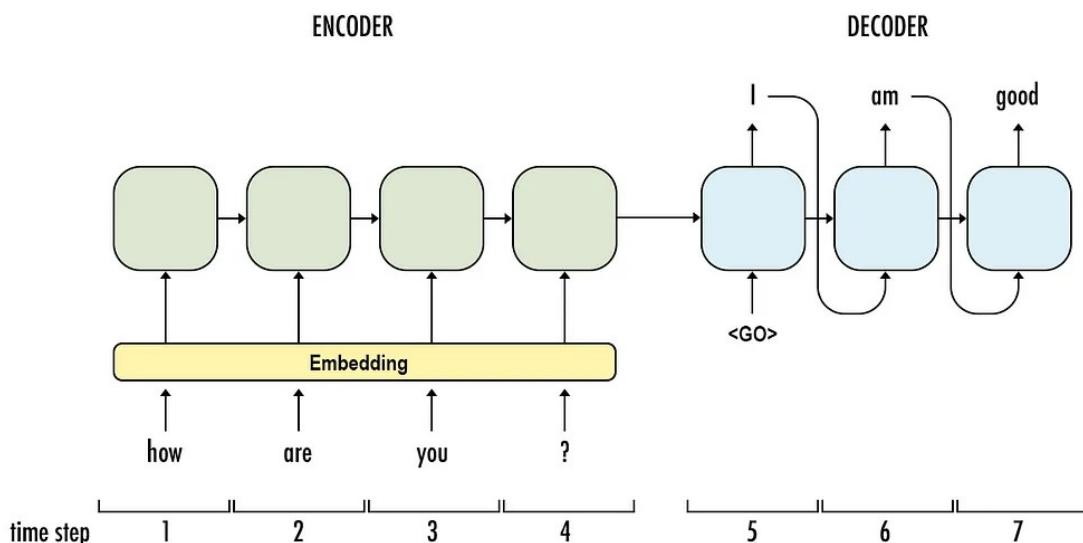
With Teacher Forcing and Attention Mechanism

This notebook explores the implementation of a Sequence-to-Sequence (seq2seq) model with attention and teacher forcing for the task of text summarization.

- **Text Summarization:** The process of condensing a longer piece of text (e.g., an article, document) into a shorter version while preserving the most important information.
- **Seq2seq Models:** A class of neural networks designed to handle sequence-to-sequence tasks, such as machine translation, text summarization, and question answering. They consist of an encoder that processes the input sequence and a decoder that generates the output sequence.
- **Attention Mechanism:** A key component in modern seq2seq models that allows the decoder to focus on different parts of the input sequence when generating each output token. This improves the model's ability to capture long-range dependencies and produce more accurate translations or summaries. (My notebook : <https://www.kaggle.com/code/divyanshvishwkarma/seq2seq-with-attention-mechanism>)
- **Teacher Forcing:** A training technique where the ground truth output tokens are fed as input to the decoder during training. This helps stabilize training and improve the quality of the generated output, especially in the early stages of training. (My notebook : <https://www.kaggle.com/code/divyanshvishwkarma/teacher-forcing-in-seq2seq-tensorflow-and-keras>)

Seq2Seq models

Seq2Seq models are a type of neural network architecture designed to handle tasks involving sequential data, such as machine translation and text summarization. They consist of two main components: an encoder, which processes the input sequence and creates a context vector, and a decoder, which generates the output sequence based on the context vector. Seq2Seq models have revolutionized many NLP tasks by effectively transforming one sequence into another.



About the data

Link : <https://www.kaggle.com/datasets/gowrishankarp/newspaper-text-summarization-cnn-dailymail>

The CNN / DailyMail Dataset is an English-language dataset containing just over 300k unique news articles as written by journalists at CNN and the Daily Mail. The current version supports both extractive and abstractive summarization, though the original version was created for machine reading and comprehension and abstractive question answering.

In [6]: `train.head(10)`

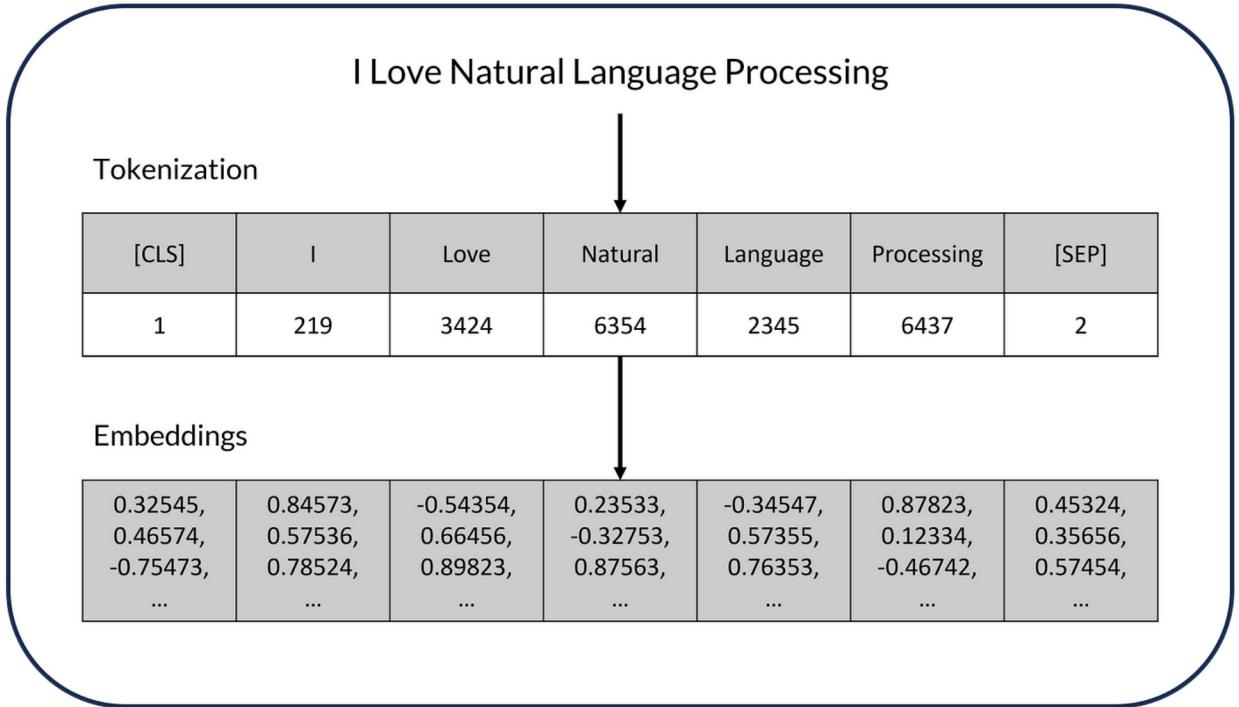
Out[6]:

	article	highlights
0	By . Associated Press . PUBLISHED: . 14:11 EST...	Bishop John Folda, of North Dakota, is taking ...
1	Kabul, Afghanistan (CNN) -- China's top securi...	China's top security official visited Afghani...
2	(CNN) -- Virgin, a leading branded venture cap...	The Virgin Group was founded by Richard Branso...
3	By . Chris Pleasance . Police are hunting for ...	Two men filmed taking iPad from canoe rental o...
4	Baghdad (CNN) -- Radical Iraqi cleric Muqtada ...	Muqtada al-Sadr has been in Iran since 2007 .\...
5	PUBLISHED: . 07:04 EST, 9 January 2014 . . U...	Zhu Sanni, 23, had been left alone at home for...
6	Kabul, Afghanistan (CNN) -- Thousands of bottl...	Official: Bottles are almost exclusively from ...
7	(CNN) -- Tour de France race director Christia...	The 2013 Tour de France will start from the Fr...
8	(CNN) -- Hundreds filed by a casket on Sunday ...	Wes Leonard collapsed after scoring a winning ...
9	Earlier this season I picked Thierry Henry as ...	Sportsmail columnist Martin Keown was honoured...

Preprocessing data

The initial step involves converting the input and target text into sequences of tokens, which can be individual words or sub-word units. This is typically achieved through tokenization techniques. To ensure uniform input shapes for the model, the sequences are then padded with

special tokens (e.g., <PAD>) to achieve equal lengths. Finally, to provide clear boundaries for the model, special start (<START>) and end (<END>) tokens are added to the beginning and end of the target sequences, respectively. This preprocessed data is then ready to be fed into the seq2seq model for training and inference.



```
In [9]: X, y = np.array(train.iloc[:, 0:1]), np.array(train.iloc[:,1:2])
```

Adding "start" and "end" token to the label datapoints

```
In [11]: START = '<start>'  
END = '<end>'  
PAD = '<PAD>'
```

```
In [12]: y = [f"{START} {text} {END}" for text in y]
```

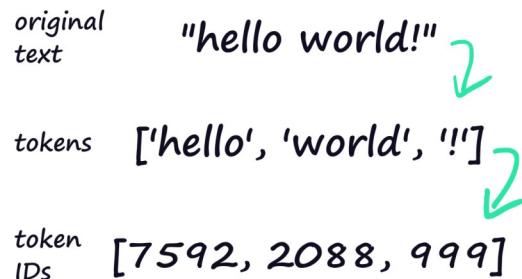
```
In [13]: size = -10  
X_test, y_test = X[size:], y[size:]  
X, y = X[:size], y[:size]
```

Preparing Tokenizer and finding vocabulary size

```
In [14]: e_tk, d_tk = Tokenizer(), Tokenizer()  
e_tk.fit_on_texts(X)  
d_tk.fit_on_texts(y)
```

Converting text to sequences, padding them and finalizing the three series
(enc_inputs, dec_inputs, targets)
analogous to (X, dec_target_input, y)

```
In [17]: enc_inputs = e_tk.texts_to_sequences(X)
targets = d_tk.texts_to_sequences(y)
```



```
In [18]: find_len = lambda x : max([len(seq) for seq in x])+1
input_seq_len, output_seq_len = find_len(enc_inputs), find_len(targets)
input_seq_len, output_seq_len
```

```
Out[18]: (329, 95)
```

```
In [19]: enc_inputs = np.array(pad_sequences(enc_inputs, padding='post', truncating='post', maxlen=output_seq_len))
```

```
In [20]: targets = pad_sequences(targets, padding='post', truncating='post', maxlen = output_seq_len)
```

```
In [21]: dec_inputs = np.array(targets[:, :-1])
targets = np.array(targets[:, 1:])
```

Dimensions of parameter to be passed the model

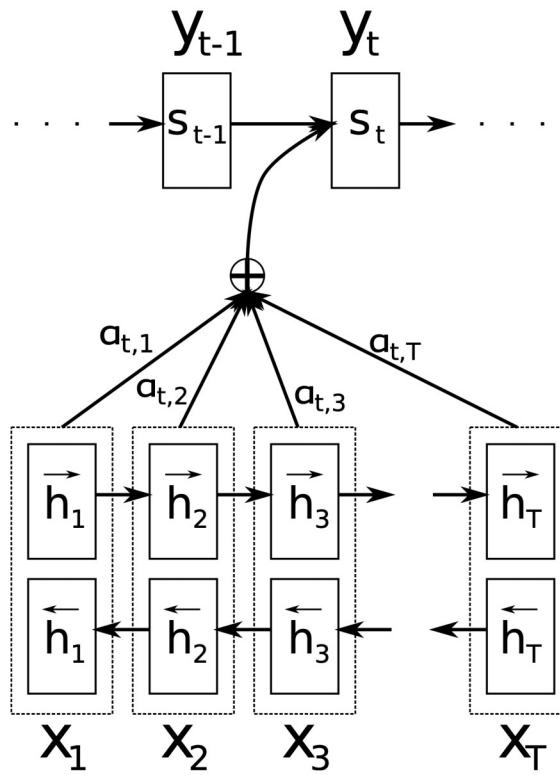
```
In [22]: in_vocab_size, out_vocab_size, input_seq_len, output_seq_len
```

```
Out[22]: (91127, 41940, 329, 95)
```

Attention Mechanism (Bahdanau Attention)

Bahdanau attention, also known as additive attention, is a mechanism designed to improve the performance of sequence-to-sequence models. It works by enabling the model to focus on specific parts of the input sequence when generating each part of the output sequence.

The decoder hidden state $s_{\{t\}}$ (query) at the $t^{\{th\}}$ timestep is passed to all encoder hidden states ($h_{\{1\}}, h_{\{2\}}, \dots, h_{\{T\}}$) to calculate scores. The attention mechanism ensures that the decoder focuses on the most relevant parts of the input (as represented by the keys) when generating the next output token.



```
In [23]: class BahdanauAttention(L.Layer):
    def __init__(self, units):
        super(BahdanauAttention, self).__init__()
        self.W1 = L.Dense(units)
        self.W2 = L.Dense(units)
        self.V = L.Dense(1)

    def call(self, query, values):
        # query - shape == (batch_size, hidden_size) -> decoder hidden state at the current time step
        # values - shape == (batch_size, max_len/timesteps, hidden_size) -> encoder output or previous decoder hidden states
        # here, hidden_size = units, max_len = timesteps
        query = tf.expand_dims(query, axis = 1) # (batch_size, 1, hidden_size)
        score = self.V(tf.nn.tanh(self.W1(query) + self.W2(values))) # (batch_size, max_len, 1)
        attention_weight = tf.nn.softmax(score, axis = 1) # (batch_size, timesteps, 1)
        context = attention_weight*values # (batch_size, timesteps, hidden_size)
        context_vector = tf.reduce_sum(context, axis = 1) # (batch_size, hidden_size)
        return context_vector, attention_weight
```

Model Definition

- Teacher Forcing is implemented in the **train_step** method where we use the actual target sequence as input to the decoder during training.
- The model uses separate Encoder and Decoder classes for better organization.
- A generate method is included for inference, which uses the model's own predictions rather than teacher forcing.
- The architecture uses LSTM cells, but you can easily modify it to use GRU or other RNN cells.

Encoder

```
In [24]: class Encoder(L.Layer):
    def __init__(self, in_vocab, embedding_dim, hidden_units):
        super(Encoder, self).__init__()
        self.embed = L.Embedding(in_vocab, embedding_dim)      # (batch_size, seq_length)
        self.lstm = L.LSTM(hidden_units, return_sequences=True, return_state = True)

    def call(self, inputs):
        # input : (batch_size, seq_length)
        x = self.embed(inputs)                                # (batch_size, seq_length)
        enc_out, hidden_state, cell_state = self.lstm(x)     # O/P (batch_size, seq_length)
        return enc_out, hidden_state, cell_state
```

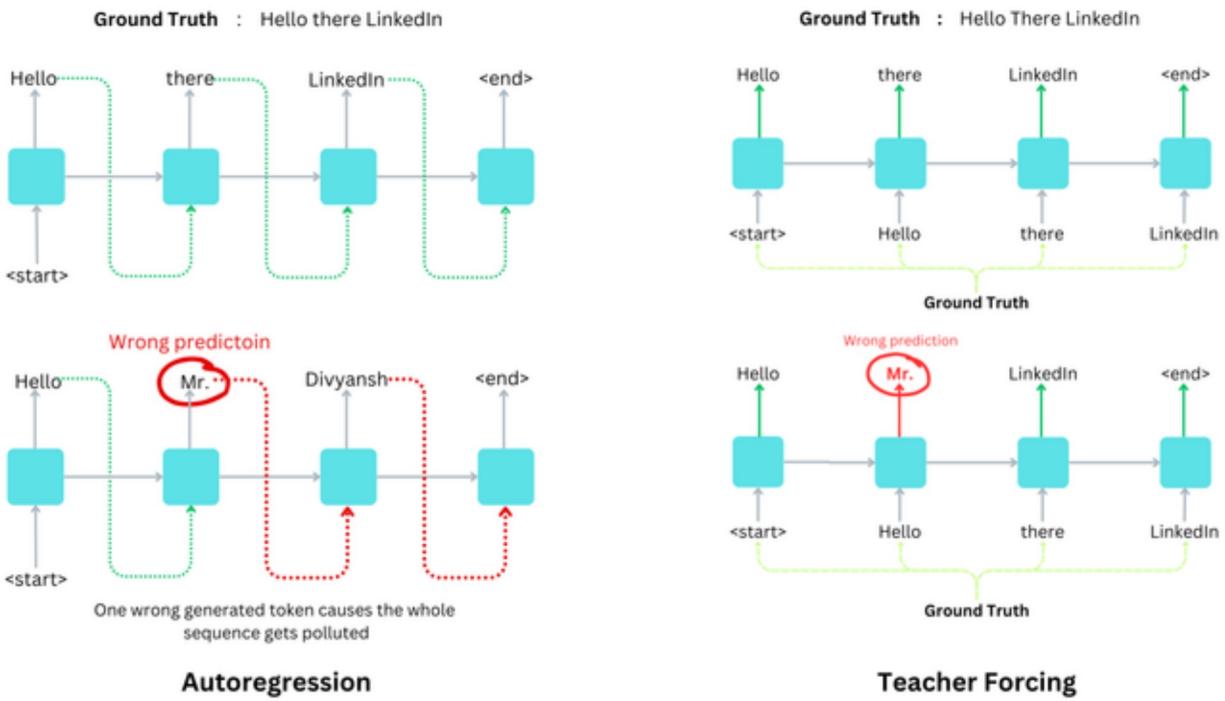
Decoder

```
In [25]: class Decoder(L.Layer):
    def __init__(self, out_vocab, embedding_dim, hidden_units):
        super(Decoder, self).__init__()
        self.embed = L.Embedding(out_vocab, embedding_dim)      # (batch_size, seq_length)
        self.lstm = L.LSTM(hidden_units, return_sequences = True, return_state = True)
        self.dense = L.Dense(out_vocab, activation='softmax')  # (batch_size, seq_length)
        self.attention = BahdanauAttention(64)

    def call(self, inputs, hidden_state, cell_state, enc_output):
        # input : (batch_size, 1)
        x = self.embed(inputs)                                # (batch_size, 1, embedding_dim)
        states = [hidden_state, cell_state]
        context, attention_weights = self.attention(query = hidden_state, values = enc_output)
        dec_out, hidden_state, cell_state = self.lstm(x, initial_state=states)  # O/P
        dec_out = tf.squeeze(dec_out, axis=1)                  # (batch_size, hidden_units)
        # context = tf.expand_dims(context, axis=1)           # (batch_size, 1, embedding_dim)
        inputs = tf.concat([context, dec_out], axis=-1)       # (batch_size, 1, embedding_dim)
        out = self.dense(inputs)                            # (batch_size, 1, out_vocab)
        return out, hidden_state, cell_state
```

Teacher Forcing

Instead of passing the output of the current timestep as the input to the next During training, teacher forcing provides the model with the ground truth (actual) output from the training data instead of feeding the model's own previous output as input. Teacher Forcing makes convergence faster during training especially in the starting epochs.



Note:

This model definition contains extra elements like the serializable decorator and the two config methods, they are optional and can be skipped without any issues. However they are required if you need to save the model in .keras format.

```
In [26]: class Seq2Seq(M.Model):

    def __init__(self, in_vocab, out_vocab, embedding_dim, hidden_units, end_token):
        super(Seq2Seq, self).__init__()

        self.in_vocab = in_vocab
        self.out_vocab = out_vocab
        self.embedding_dim = embedding_dim
        self.hidden_units = hidden_units

        self.encoder = Encoder(in_vocab, embedding_dim, hidden_units)
        self.decoder = Decoder(out_vocab, embedding_dim, hidden_units)
        self.end_token = end_token

    @tf.function
    def train_step(self, inputs):
        (enc_inputs, dec_inputs), targets = inputs           # (batch_size, seq_length)

        with tf.GradientTape() as tape:
            enc_out, hidden_state, cell_state = self.encoder(enc_inputs)          # (batch_size, seq_length, hidden_units)
            seq_len = dec_inputs.shape[1]
            dec_out = tf.TensorArray(tf.float32, seq_len) # (batch_size, seq_len, target_vocab_size)
            mask = tf.TensorArray(tf.bool, size=seq_len)
            for timestep in tf.range(seq_len):
                timestep_input = dec_inputs[:, timestep:timestep+1]      # (batch_size, 1, target_vocab_size)
                timestep_output, hidden_state, cell_state = self.decoder(timestep_input, hidden_state, cell_state)
                dec_out = dec_out.write(timestep, timestep_output)
                is_end = tf.equal(targets[:, timestep], self.end_token) # Creating mask
                mask = mask.write(timestep, tf.logical_not(is_end))
            dec_out = tf.transpose(dec_out.stack(), [1, 0, 2])
            sequence_mask = tf.transpose(mask.stack(), [1, 0])
```

```

        loss = self.compiled_loss(targets, dec_out, sample_weight=tf.cast(sequence_length, tf.float32))
        variables = self.trainable_variables
        gradients = tape.gradient(loss, variables)
        self.optimizer.apply_gradients(zip(gradients, variables))
        self.compiled_metrics.update_state(targets, dec_out) # Update metrics
        return {m.name : m.result() for m in self.metrics}

    @tf.function
    def call(self, inputs, training=False):
        enc_inputs, dec_inputs = inputs
        enc_out, hidden_state, cell_state = self.encoder(enc_inputs) # (batch_size, max_len, embedding_dim)
        seq_len = tf.shape(dec_inputs)[1]
        dec_out = tf.TensorArray(tf.float32, seq_len) # (batch_size, seq_len, target_vocab_size)
        for timestep in tf.range(seq_len):
            timestep_input = dec_inputs[:, timestep:timestep+1] # (batch_size, 1, embedding_dim)
            timestep_output, hidden_state, cell_state = self.decoder(timestep_input, hidden_state, cell_state)
            dec_out = dec_out.write(timestep, timestep_output)
        return tf.transpose(dec_out.stack(), [1, 0, 2])

    def generate(self, enc_inputs, max_len, start, end):
        enc_out, hidden_state, cell_state = self.encoder(enc_inputs)
        dec_in = tf.expand_dims([start], 0) # To get from int -> (1,1) tensor
        result = []
        for _ in range(max_len):
            prediction_logits, hidden_state, cell_state = self.decoder(dec_in, hidden_state, cell_state)
            prediction = tf.argmax(prediction_logits, axis=-1) # return token id
            if prediction == end:
                break
            result.append(prediction.numpy())
            dec_in = tf.expand_dims(prediction, 0)
        return result

```

Model Instance and Training

The model was trained over 40 epochs of training, hyperparameters are 512 embedding dimension and 512 LSTM units in both encoder and decoder.

```
In [27]: model = Seq2Seq(in_vocab=in_vocab_size, out_vocab=out_vocab_size, embedding_dim=1024, encoder_type='lstm', decoder_type='lstm', units=512, latent_dim=512)
```

```
In [28]: model.compile(optimizer='adam', loss='sparse_categorical_crossentropy', metrics=['accuracy'])
```

Model Inference

```
In [36]: summarize(1116)
```

INPUT

(CNN) -- Rescue workers have pulled a body from underneath the rubble of a collapsed apartment building in Cologne, Germany, police spokeswoman Astrid Gelss told CNN. The remains of Cologne's archive building following the collapse on Tuesday afternoon. The apartment building collapsed together with two other buildings on March 3. Initially as many as nine people were reported missing, but after seven of them reported to the authorities that they were safe, search and rescue efforts focused on finding the two remaining missing men. Rescue workers found the body at about 2 a.m. Sunday morning (8 p.m. ET Saturday). It has not been identified. The search for the second missing person is still ongoing, Gelss said. It is still not clear what caused the collapse of the building which contained the city's historical archives, bringing down parts of the two nearby structures. CNN's Per Nyberg contributed to this report.

MODEL OUTPUT

rescue workers pull body from rubble of collapsed building one person still missing after building collapsed in cologne authorities haven't yet determined what brought down the structure which costs about 140 000 people have been killed in india since china watch the cologne more than 100 people have died in september chinese authorities say t hey don't know what sparked the fire zone in september more than 48 million users have been evacuated as heavy than 6 400 people stranded in the east river have been evaucated more than 100 vehicles in the building leg in

GIVEN SUMMARY

Rescue workers pull body from rubble of collapsed building .
One person still missing after building collapsed in Cologne .
Authorities haven't yet determined what brought down the structure .

In [37]: `summarize(100)`

INPUT

Los Angeles (CNN) -- Michael Jackson's brothers are working on a reunion tour to perform their old Jackson 5 songs, Jermaine Jackson said Monday. "Michael would want us to continue on," Jackson said, after he unveiled an exhibit of his brother's wax statues at Madame Tussauds in Hollywood. Saturday marks the second anniversary of the pop icon's sudden death, which the coroner ruled was from an overdose of a surgical anesthetic. "The brothers now are ready just to keep the legacy going and the music," Jackson said. "We're meeting today about putting a tour together just to play the music because it's such wonderful music. We need to continue to play this music because it's healing for us." Jackson did not say which of his brothers were involved in the talks. Recent attempts to reunite the brothers on stage were met with frustration and division, including the effort documented by an A&E Network reality show in 2009. Monday's unveiling was a reunion of sorts for Michael Jackson. It featured wax figures depicting him as a child performing with the Jackson 5, as the "Smooth Criminal" from around 1987 and the "This Is It" Michael Jackson of his last year. Jermaine Jackson said his favorite was the last statue because of how it captured his brother's onstage posture. "Someone was really paying a lot of attention to detail," he said. The Michael Jackson exhibit is on a world tour, spending this summer in Hollywood before heading to Las Vegas in September.

MODEL OUTPUT

jermaine jackson says the brothers are meeting monday to plan a new tour michael would want us to continue on he says jackson was unveiling wax figures of michael jackson on tour at madame tussauds after admitting meeting to meeting that he would need the 59 to go for playing a tour of brothers and a plan to continue in 2005 deal was revealed he will continue a new zealand tour for monday after meeting up in the us he adds he would cooperate fully with the police and he still loved he has been

GIVEN SUMMARY

Jermaine Jackson says the brothers are meeting Monday to plan a new tour .

"Michael would want us to continue on," he says .

Jackson was unveiling wax figures of Michael Jackson on tour at Madame Tussauds .

In [38]: `summarize(210)`

INPUT

Stoke City midfielder Charlie Adam is open to the possibility of a move abroad in the future amid links to Real Sociedad. The 29-year-old has started just five times for Mark Hughes' men in the Premier League this season and a switch to La Liga with Sociedad - who are coached by David Moyes - has been mooted. Adam insists that he is no hurry to leave the Britannia Stadium, but admits the idea of plying his trade outside of England before the end of his career is something that appeals to him. Charlie Adam is open to the possibility of moving abroad to continue his career. The 29-year-old has started just five Premier League games for Stoke this season. 'There's better weather than here,' he told the Stoke Sentinel. 'I would love the experience of playing abroad because it's certainly something different. Hopefully, more British players will go and do it. On rumours of an exit in January, he added: 'It's going to happen when you have good players not playing and the opportunity for clubs to have a look at things [in the transfer window]. 'But there's nothing happening at the moment and I am happy to be here at Stoke.' Adam has been linked with a potential move to David Moyes' Real Sociedad.

MODEL OUTPUT

charlie adam is open to the possibility of moving abroad scottish midfielder has started just five games this season adam has been linked with a move to david moyes' real sociedad press conference will play in scottish champions league final he was also held to break into possibility of scottish football league squad games for possibility of this year is already has charlie possibility that could be open to public run abroad scottish city manager said he is open to remain at the possibility of a leading club to david moyes he will join up

GIVEN SUMMARY

Charlie Adam is open to the possibility of moving abroad.
Scottish midfielder has started just five games this season.
Adam has been linked with a move to David Moyes' Real Sociedad.

In [39]: `summarize(54)`

INPUT

Rio Ferdinand has lifted the lid on Wayne Rooney's temper tantrums, telling how his former Manchester United team-mate used to 'smash up mobile phones in frustration'. Ferdinand and Rooney played alongside each other for 10 years at Old Trafford until the former left the club during the summer. And the QPR defender has shed new light on the current United captain's temper - although he admitted the England striker has now calmed down. Wayne Rooney had a temper in his younger days, reveals former team-mate Rio Ferdinand. The United striker used to 'fly into a rage about the smallest things,' says Ferdinand. Ferdinand acknowledges Rooney has calmed down a bit - although he's still prone to the odd outburst. Possibly Rooney's most famous outburst, against West Ham in 2011 when he swore into a camera. Ferdinand says Rooney would often be arguing on his phone outside the United training ground. In *The Sun*'s serialisation of his autobiography #2sides, Ferdinand said: 'Wayne Rooney always struck me as a very angry young man, always arguing with people outside the training ground, especially on the phone. 'He seemed to fly into a rage about the smallest things and went through mobile phones like they were sweets. He'd smash phones up in frustration, throwing them on the concrete. 'I don't know what happened but he did calm down.' Ferdinand says Rooney (right) struck him as an angry young man, always arguing with people. Ferdinand writes in his autobiography, #2sides, that Rooney went through phones 'like they were sweets' VIDEO Ferdinand book criticises Moyes and Terry.

MODEL OUTPUT

rio ferdinand writes about wayne rooney's temper in his new book ferdinand says rooney struck him as an angry young man he said his former manchester united team mate would often smash phones on the concrete outside their training ground rooney has calmed down now though ferdinand says his suspension is coming to a concrete outside west ham united ferdinand denies his name without assistance the former manchester united striker would have to pay his debut in concrete ferdinand says he is angry that is a concrete to his concrete outside his former team mate

GIVEN SUMMARY

Rio Ferdinand writes about Wayne Rooney's temper in his new book .
Ferdinand says Rooney struck him as an angry young man .
He said his former Manchester United team-mate would often smash phones on the concrete outside their training ground .
Rooney has calmed down now though, Ferdinand says .

In [40]: `summarize(60)`

INPUT

Watford are poised to sign Ecuador right-back Juan Carlos Paredes. The 27-year-old has refused to return to club Sporting Barcelona who plan to fine him. Club president Francisco Hidalgo confirmed: 'The negotiation with the English club is almost closed and we are only missing a few defining details.' Defensive striker: The full-back has refused to go back to his club in order to force through a move to England. World Cup star: Paredes was impressive for Ecuador in Brazil despite their failure to progress. He also insisted that the player has not explained his absence to the club, who have been unable to contact him since the World Cup. Paredes started all three of Ecuador's matches in Brazil and was impressive for his country as they failed to make it past the Group Stages.

MODEL OUTPUT

paredes impressed during world cup for ecuador player hasn't returned to his club who have fined him but sporting barcelona admit they are in negotiations with watford in 2010 world cup winner also offered him for two year as long loan as cameroon boss since may have also been offered in the region but will have further talks with the club have also agreed a further year deal so they have not have offered a deal click here for more transfer news chairman says this season click here for more transfer news chairman our keen

GIVEN SUMMARY

Paredes impressed during World Cup for Ecuador .
Player hasn't returned to his club, who have fined him .
But Sporting Barcelona admit they are in negotiations with Watford .

In [41]: `summarize(513)`

INPUT

Luis Suarez has revealed he couldn't stop crying after Crystal Palace scored three goals in 11 minutes to thwart Liverpool's Premier League title bid. Liverpool head into Sunday's final round of fixtures two points adrift of Manchester City after their 3-3 draw at Selhurst Park on Monday night. Suarez, who was voted Footballer of the Year by the Football Writers' Association last week, was left on the Palace turf in tears after the match, and had to be consoled by captain Steven Gerrard. Blubber: Luis Suarez was left in tears on Monday night after Crystal Palace drew 3-3 with Liverpool on Monday. On the ball: Suarez in training for Liverpool on Saturday ahead of their clash against Newcastle. 'My team-mates were helping me, but I couldn't stop,' revealed Suarez, who has scored 31 goals this season. 'I play every single game from the heart and I was very sad. 'I was sad for my team-mates who have done so much. I was sad that I had missed chances.' Liverpool finished seventh last season, but have launched an inspiring assault on the title, some 24 years after they last won England's top flight. Speaking to the Sunday Mirror, Suarez added: 'It is difficult. At the start of the season, our aim was to finish in the top four and qualify for the Champions League, so it has been good. 'But then we got into a position when we thought we could win the league, but lost against Chelsea and then that happened at Crystal Palace.' Suited and booted: Suarez has scored 31 goals for Liverpool in an award-winning season for the Uruguayan .

MODEL OUTPUT

luis suarez says he was 'sad' he says liverpool could not feature football game between 3 and fa cup winner against liverpool in group f suarez was criticised in a 0 0 draw against liverpool on sunday they would face newcastle united at the anfield hotel in t he summer click here for all the latest liverpool news believes our live updates to fo otball with liverpool side liverpool face west brom at the time on saturday suarez was left out of football squad with ankle injury against liverpool last week click here fo r all the latest

GIVEN SUMMARY

Luis Suarez says he was 'sad' he 'missed chances' in draw at Selhurst Park . Liverpool are two points adrift of Manchester City with one game remaining . The Reds face Newcastle at Anfield with City hosting West Ham on Sunday . Suarez was voted Football of the Year by the Football Writers Association .

In [42]: `summarize(6)`

INPUT

Kabul, Afghanistan (CNN) -- Thousands of bottles of alcohol were destroyed in Kabul this week, in what authorities described as the product of a crackdown on illegal smuggling and sales. The bottles were confiscated over a two-year period in and around the Afghan capital, according to Kabul police and criminal investigations chief Mohammad Zahir. They were taken almost exclusively from "Afghan sources and not foreigners," he said. The illicit items were being stored by Afghan customs officials, who burned the bottles Wednesday after receiving authorization from the city's attorney general's office, he added. Alcohol is largely banned in Afghanistan, and its sales and consumption considered a criminal offense for the country's . Muslims, who constitute roughly 99% of the population. Certain areas that cater to foreigners, however, are permitted to sell it. Zahir said that it was in these areas -- mostly international hotels -- that local sellers had come into possession of the alcohol. CNN's Matiullah Mati contributed to this report .

MODEL OUTPUT

official bottles are almost exclusively from afghan sources and not foreigners alcohol is largely banned in afghanistan certain areas however that cater to foreigners are permitted to sell it claims it is permitted by the cia and that it's some such attacks have been an important part of that both cater for foreigners are permitted to sell de important for both attacks but have been largely banned for the protests including afghan children sources say they are using them about an hour off the phone is now recovering in afghanistan a coalition service official and

GIVEN SUMMARY

Official: Bottles are almost exclusively from "Afghan sources" and not foreigners .
Alcohol is largely banned in Afghanistan .
Certain areas, however, that cater to foreigners are permitted to sell it .

In [43]: `summarize(8)`

INPUT

(CNN) -- Hundreds filed by a casket on Sunday to say goodbye to a small-town hero who died shortly after scoring the winning shot to clinch an undefeated season for his Michigan high school basketball team. Friends, family and even competitors braved a cold Michigan day for a pair of afternoon visitations for 16-year-old Wes Leonard of Fennville. Some piled off of school buses and hugged each other in the crisp winter air. "He was just an all around great guy to play against," Craig Peterson, a player on the Bangor high school basketball team, told CNN affiliate WZZM. "He was always respectful, he didn't run his mouth like most of the people would. He just liked to have fun, and was a great athlete." Leonard's funeral is set for 10:30 a.m. Tuesday at Christ Memorial Church in nearby Holland. Leonard died moments after scoring the winning basket in overtime Thursday, securing a perfect season for the Fennville High School Blackhawks. Leonard scored his team's last four points in the 57-55 victory. An autopsy Friday revealed Leonard died of cardiac arrest due to an enlarged heart, according to a statement from Dr. David A. Start, the Ottawa County chief medical examiner.

MODEL OUTPUT

wes leonard collapsed after scoring a winning basket in overtime his shot secured a perfect season for his high school team an autopsy revealed the teen died of cardiac arrest due to an enlarged heart attack in 2006 the perfect was part of an perfect start to saturday to 2 years into parking lot there an autopsy revealed she was declared dead at the scene cochrane is due to appear in court later this month charges for an undisclosed fee made 10 000 for restitution to attend charges wes wes leonard is saturday for bail

GIVEN SUMMARY

Wes Leonard collapsed after scoring a winning basket in overtime .
His shot secured a perfect season for his high school team .
An autopsy revealed the teen died of cardiac arrest due to an enlarged heart .

Observations

- The outputs are highly aligned to the label data. The results are much better when compared to the model without attention.
 - The model output turns out to be semantically correct most of the time. However it sometimes mixes up or repeats words mid sentence.
 - The model took much longer to train,i.e., 7 minutes as compared to 1.5 minutes for model without attention for 1 epoch of training, largely due to the sequential nature of the architecture.
 - Model complexity is higher, however the results are good enough to justify this added complexity
 - However the enhancement in generated text provided by attention paves way for architectures like transformer to be developed.
 - Accuracy during training (both on train set and validation set), increased largely in a linear fashion.
 - During testing I found that increasing the size of sentences by ~50% increased the training time from 4.5 hours to almost 12.5 hours.
-

References

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. <https://arxiv.org/abs/1409.3215>

Attention Mechanism:

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. 1 <https://arxiv.org/pdf/1409.0473.pdf>

Luong, M.-T., Pham, H., & Manning, C. D. (2015). Effective approaches to attention-based neural machine translation. <https://aclanthology.org/D15-1166/>

- **Original Paper for Teacher Forcing:**

- Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." Advances in neural information processing systems 27 (2014).

- **Review Papers:**

- Wu, Yonghui, et al. "Google's neural machine translation system: Bridging the gap between human and machine translation." arXiv preprint arXiv:1609.08144 (2016).
- Young, Tom, et al. "Recent trends in neural machine translation." arXiv preprint arXiv:1703.01619 (2017).

- **Tutorials and Blog Posts:**

- **Understanding Teacher Forcing in Seq2Seq Models** by Lilian Weng: <https://lilianweng.github.io/>
- **Seq2Seq Tutorial with Neural Networks** by PyTorch: <https://pytorch.org/tutorials/>

- **Research Papers Exploring Alternatives to Teacher Forcing:**

- **Scheduled Sampling for Sequence Prediction with Recurrent Neural Networks** by Samy Bengio et al.: <https://arxiv.org/abs/1506.03099>
- **Improved Training of Sequence to Sequence Models** by Minh-Thang Luong et al.: <https://research.google/pubs/sequence-to-sequence-learning-with-neural-networks/>