

Práctica Texto – Creación de un modelo de Traducción ES_EN

*Esta práctica está subida al repositorio de [GitHub](#). Todas las explicaciones de cómo funcionan las distintas partes del código se han ido explicando en el Notebook utilizado, mientras que en este informe solamente están las estructuras y resultados básicos obtenidos.

El objetivo de esta práctica es la elaboración de un modelo de traducción de Castellano a Inglés utilizando como fuente de datos pares de frases traducidas inglés-castellano.

Para esta práctica hemos utilizado 3 modelos distintos

- RNN
- GRU
- LSTM

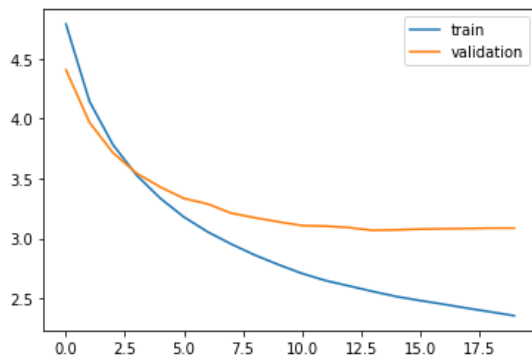
De los cuales hemos probado dos arquitecturas distintas, una con una longitud de las *features* de 8 elementos, y otra de 16, siendo las *units* de los modelos siempre de 128 elementos (por limitaciones de computación).

Para entrenar los modelos hemos utilizado el set de datos al completo, de más de 130k pares de frases, utilizando un 2% de los elementos para Test (también por limitaciones de computación).

RNN

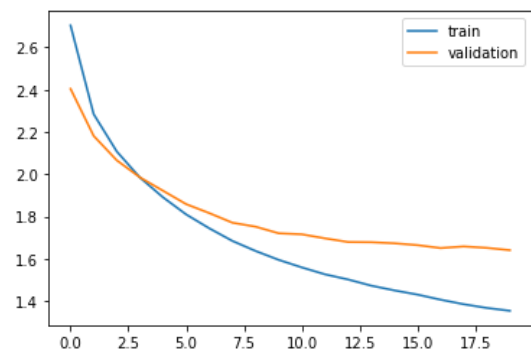
Model #1

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 8, 128)	3511168
simple_rnn (SimpleRNN)	(None, 128)	32896
repeat_vector (RepeatVector)	(None, 8, 128)	0
simple_rnn_1 (SimpleRNN)	(None, 8, 128)	32896
dense (Dense)	(None, 8, 13731)	1771299
Total params: 5,348,259		
Trainable params: 5,348,259		
Non-trainable params: 0		



Model #2

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 16, 128)	3511168
simple_rnn (SimpleRNN)	(None, 128)	32896
repeat_vector (RepeatVector)	(None, 16, 128)	0
simple_rnn_1 (SimpleRNN)	(None, 16, 128)	32896
dense (Dense)	(None, 16, 13731)	1771299
Total params: 5,348,259		
Trainable params: 5,348,259		
Non-trainable params: 0		



Predicciones

Model #1

	spanish	english	predicted
0	le ha perdonado todo	she has forgiven him for everything	she has everything for
1	me gustaría tener una respuesta	id like to have an answer	i like to a
2	es seguro que él vendrá	he is sure to come	that is he come
3	descubrimos ruinas de una civilización antigua	we discovered relics of an ancient civilization	i have an the
4	soy soltero	i am single	im a
5	estaba muy cansado de manera que me fui temprano a la cama	i was very tired so i went to bed early	to early i go to bed
6	él es capaz de hablar diez lenguas	he is able to speak ten languages	he often swim to languages
7	debiste haber venido más pronto	you should have come earlier	you should wait come here
8	me acuerdo diciéndotelo así	i remember telling you so	i remember it
9	¿necesitas el libro	do you need the book	do you need the book
10	¿compraste un lavarropas	did you buy a washing machine	did you buy a washing machine
11	tom es un manitas ¿a que sí	tom is very crafty isnt he	tom is a than to
12	¿a qué distancia está de nueva york	how far is it to new york	what car is this new this new
13	¿puedo ir al cine esta noche	may i go to the movies tonight	can i go the tonight
14	nos conocemos el uno al otro desde que éramos niños	we have known each other since we were children	we the we were a

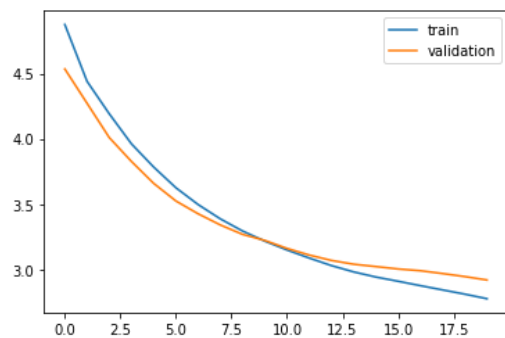
Model #2

	spanish	english	predicted
0	le ha perdonado todo	she has forgiven him for everything	he has everything
1	me gustaría tener una respuesta	id like to have an answer	i like to a
2	es seguro que él vendrá	he is sure to come	its is sure he come
3	descubrimos ruinas de una civilización antigua	we discovered relics of an ancient civilization	i have a of
4	soy soltero	i am single	im am
5	estaba muy cansado de manera que me fui temprano a la cama	i was very tired so i went to bed early	it was tired bed i bed
6	él es capaz de hablar diez lenguas	he is able to speak ten languages	he is speak to the languages
7	debiste haber venido más pronto	you should have come earlier	you should have to more
8	me acuerdo diciéndotelo así	i remember telling you so	i remember that it
9	¿necesitas el libro	do you need the book	do you need the book
10	¿compraste un lavarropas	did you buy a washing machine	did you buy a car
11	tom es un manitas ¿a que sí	tom is very crafty isnt he	tom is a to tom mary
12	¿a qué distancia está de nueva york	how far is it to new york	is the new is
13	¿puedo ir al cine esta noche	may i go to the movies tonight	can i go the tonight
14	nos conocemos el uno al otro desde que éramos niños	we have known each other since we were children	we were the we

GRU

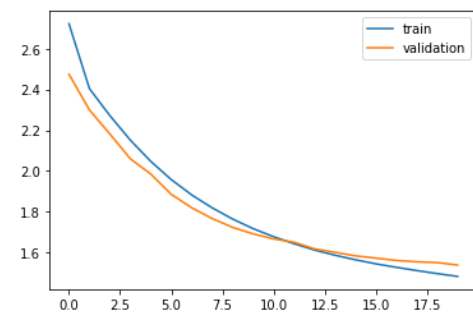
Model #1

Layer (type)	Output Shape	Param #
embedding_2 (Embedding)	(None, 8, 128)	3511168
gru_2 (GRU)	(None, 128)	99072
dropout_2 (Dropout)	(None, 128)	0
repeat_vector_2 (RepeatVecto	(None, 8, 128)	0
gru_3 (GRU)	(None, 8, 128)	99072
dropout_3 (Dropout)	(None, 8, 128)	0
dense_2 (Dense)	(None, 8, 13731)	1771299
Total params: 5,480,611		
Trainable params: 5,480,611		
Non-trainable params: 0		



Model #2

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 16, 128)	3511168
gru (GRU)	(None, 128)	99072
dropout (Dropout)	(None, 128)	0
repeat_vector (RepeatVector)	(None, 16, 128)	0
gru_1 (GRU)	(None, 16, 128)	99072
dropout_1 (Dropout)	(None, 16, 128)	0
dense (Dense)	(None, 16, 13731)	1771299
Total params: 5,480,611		
Trainable params: 5,480,611		
Non-trainable params: 0		



Predicciones

Model #1

	spanish	english	predicted
0	le ha perdonado todo	she has forgiven him for everything	she has everything
1	me gustaría tener una respuesta	id like to have an answer	id like to have a answer
2	es seguro que él vendrá	he is sure to come	he is sure he will come
3	descubrimos ruinas de una civilización antigua	we discovered relics of an ancient civilization	we a of a
4	soy soltero	i am single	im am
5	estaba muy cansado de manera que me fui temprano a la cama	i was very tired so i went to bed early	i was i to bed
6	él es capaz de hablar diez lenguas	he is able to speak ten languages	he is speak to speak ten
7	debiste haber venido más pronto	you should have come earlier	you should have come early
8	me acuerdo diciéndotelo así	i remember telling you so	i hate that it
9	¿necesitas el libro	do you need the book	do you need the book
10	¿compraste un lavarropas	did you buy a washing machine	did you buy a new
11	tom es un manitas ¿a que sí	tom is very crafty isnt he	tom is a than isnt
12	¿a qué distancia está de nueva york	how far is it to new york	what is the to in this new
13	¿puedo ir al cine esta noche	may i go to the movies tonight	i go to tonight
14	nos conocemos el uno al otro desde que éramos niños	we have known each other since we were children	we of the we were

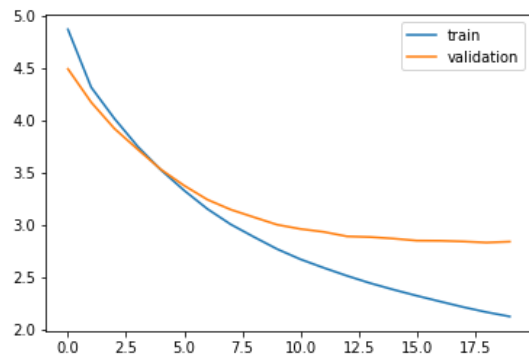
Model #2

	spanish	english	predicted
0	le ha perdonado todo	she has forgiven him for everything	he has everything
1	me gustaría tener una respuesta	id like to have an answer	id like to have a answer
2	es seguro que él vendrá	he is sure to come	he is that he come
3	descubrimos ruinas de una civilización antigua	we discovered relics of an ancient civilization	we were a of the
4	soy soltero	i am single	im am
5	estaba muy cansado de manera que me fui temprano a la cama	i was very tired so i went to bed early	i was very tired i to go bed
6	él es capaz de hablar diez lenguas	he is able to speak ten languages	he is to speak or
7	debiste haber venido más pronto	you should have come earlier	you should be to come
8	me acuerdo diciéndotelo así	i remember telling you so	i remember you that
9	¿necesitas el libro	do you need the book	do you need the book
10	¿compraste un lavarropas	did you buy a washing machine	did you buy a buy
11	tom es un manitas ¿a que sí	tom is very crafty isnt he	tom is a is than he
12	¿a qué distancia está de nueva york	how far is it to new york	how does it new
13	¿puedo ir al cine esta noche	may i go to the movies tonight	can i go to the
14	nos conocemos el uno al otro desde que éramos niños	we have known each other since we were children	we have the we

LSTM

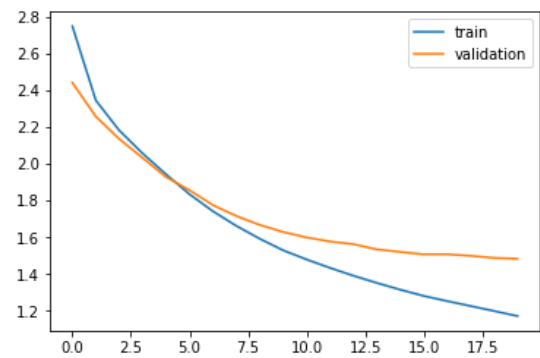
Model #1

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 8, 128)	3511168
lstm (LSTM)	(None, 128)	131584
repeat_vector (RepeatVector)	(None, 8, 128)	0
lstm_1 (LSTM)	(None, 8, 128)	131584
dense (Dense)	(None, 8, 13731)	1771299
Total params: 5,545,635		
Trainable params: 5,545,635		
Non-trainable params: 0		



Model #2

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 16, 128)	3511168
lstm (LSTM)	(None, 128)	131584
repeat_vector (RepeatVector)	(None, 16, 128)	0
lstm_1 (LSTM)	(None, 16, 128)	131584
dense (Dense)	(None, 16, 13731)	1771299
Total params: 5,545,635		
Trainable params: 5,545,635		
Non-trainable params: 0		



Predicciones

Model #1

	spanish	english	predicted
0	le ha perdonado todo	she has forgiven him for everything	he has up about everything
1	me gustaría tener una respuesta	id like to have an answer	id like to a
2	es seguro que él vendrá	he is sure to come	its is that he will come
3	descubrimos ruinas de una civilización antigua	we discovered relics of an ancient civilization	we the old of a old
4	soy soltero	i am single	im am
5	estaba muy cansado de manera que me fui temprano a la cama	i was very tired so i went to bed early	i go bed
6	él es capaz de hablar diez lenguas	he is able to speak ten languages	he is able to speak languages
7	debiste haber venido más pronto	you should have come earlier	you should have come again
8	me acuerdo diciéndotelo así	i remember telling you so	i remember myself that so
9	¿necesitas el libro	do you need the book	do you need the book
10	¿compraste un lavarropas	did you buy a washing machine	did you buy a washing machine
11	tom es un manitas ¿a que sí	tom is very crafty isnt he	tom is an terrible you
12	¿a qué distancia está de nueva york	how far is it to new york	what did you is from new in town
13	¿puedo ir al cine esta noche	may i go to the movies tonight	can i go to the tonight
14	nos conocemos el uno al otro desde que éramos niños	we have known each other since we were children	have the other they were

Model #2

	spanish	english	predicted
0	le ha perdonado todo	she has forgiven him for everything	he has everything
1	me gustaría tener una respuesta	id like to have an answer	id like to answer
2	es seguro que él vendrá	he is sure to come	its sure he come
3	descubrimos ruinas de una civilización antigua	we discovered relics of an ancient civilization	we were a of found a born
4	soy soltero	i am single	im am
5	estaba muy cansado de manera que me fui temprano a la cama	i was very tired so i went to bed early	i was tired that i to bed
6	él es capaz de hablar diez lenguas	he is able to speak ten languages	he is able to languages
7	debiste haber venido más pronto	you should have come earlier	you must have be soon
8	me acuerdo diciéndotelo así	i remember telling you so	i remember that
9	¿necesitas el libro	do you need the book	do you need the book
10	¿compraste un lavarropas	did you buy a washing machine	did you buy a machine
11	tom es un manitas ¿a que sí	tom is very crafty isnt he	tom is a than you
12	¿a qué distancia está de nueva york	how far is it to new york	how of is new
13	¿puedo ir al cine esta noche	may i go to the movies tonight	can i go to the movies
14	nos conocemos el uno al otro desde que éramos niños	we have known each other since we were children	we know the other who

Conclusiones

Primero de todo, los resultados obtenidos con el tipo de modelo #2 en los tres casos ha sido ligeramente mejor a todos los del tipo #1, como se puede observar tanto en las gráficas de Loss como comparando los resultados de predicción (aunque en estos resultados la diferencia apreciada es pequeña).

Por otro lado, tanto los modelos de RNN como de LSTM tienden al sobreaprendizaje, mientras que el de GRU no. Además, viendo las tendencias de las curvas podría haberse entrenado con más epochs el modelo de GRU.

Por último, si tuviéramos que elegir uno de los tres modelos sería el de GRU, dado que sus traducciones son ligeramente mejores al resto, pero aún así este método de usar modelos Seq2Seq lo consideramos (por lo menos con lo que hemos ejecutado) poco válido para conseguir resultados de calidad.