

Como enseñar a las máquinas a leer y comprender

Sanchez Ramos, Lesly Dashiel
Universidad Nacional
de Ingeniería
Lima, Perú
lesly.sanchez.r@uni.pe

Yagi Vasquéz, Gladys Alesandra
Universidad Nacional
de Ingeniería
Lima, Perú
gladys.yagi.v@uni.pe

Resumen—Los modelos RNN utilizados en máquinas con el propósito de enseñarles a leer y comprender han ido en aumento en los últimos años. Sin embargo, sigue siendo un reto implementar un modelo completamente fiable para esta clase de propósito. Además, esta clase de modelo trabajan con un gran conjunto de datos.

En el presente trabajo se implementó un modelo basado en RNN para enseñarle a una máquina a leer y comprender un libro de niños.

Palabras Claves: NLP, RNN, Seq2Seq, CBT.

I. INTRODUCCIÓN:

Los avances en los algoritmos de Deep Learning, con el objetivo de darle a una máquina la capacidad de leer y comprender el lenguaje en documentos, han ido en aumento [1]. Esto se debe a que en los últimos años los conjuntos de datos para entrenamiento y prueba a gran escala para evaluar este tipo de capacidad han aumentado, así como la creación de frameworks, tales como ParlAI, que nos ofrecen una amplia gama de herramientas para entrenar modelos de lectura y comprensión así como modelos de diálogo y respuesta.

La capacidad de las máquinas para leer y comprender un determinado documento puede ser testada mediante su habilidad para poder responder preguntas y respuestas por medio de contenido de documentos, predecir palabras en historias incompletas, resumir reseñas de productos o predecir si una premisa y su hipótesis se contradicen o son neutrales entre sí.

Teaching Machines to Read and Comprehend [1] es uno de los papers más famosos y mencionados que implementa modelos basados en redes neuronales para probar la capacidad de un sistema para responder preguntas planteadas sobre el contenido de un documento que han visto. Uno de esos modelos es el Deep LSTM Reader, donde por medio un encoder y decoder Deep LSTM, se procesan las consultas y luego el documento, dando como resultado un modelo que

procesa cada par de consultas del documento como única secuencia larga.

Un conjunto de datos famoso por ser utilizado para la comprensión lectora de las máquinas es el CNN/Daily Mail.

Original Version	Anonymised Version
Context The BBC producer allegedly struck by Jeremy Clarkson will not press charges against the "Top Gear" host, his lawyer said Friday. Clarkson, who hosted one of the most-watched television shows in the world, was dropped by the BBC Wednesday after an internal investigation by the British broadcaster found he had subjected producer Oisin Tymon "to an unprovoked physical and verbal attack." ...	the <i>ent381</i> producer allegedly struck by <i>ent212</i> will not press charges against the " <i>ent153</i> " host, his lawyer said Friday. <i>ent212</i> , who hosted one of the most-watched television shows in the world, was dropped by the <i>ent381</i> Wednesday after an internal investigation by the <i>ent180</i> broadcaster found he had subjected producer <i>ent193</i> "to an unprovoked physical and verbal attack." ...
Query Producer X will not press charges against Jeremy Clarkson, his lawyer says.	producer X will not press charges against <i>ent212</i> , his lawyer says.
Answer Oisin Tymon	<i>ent193</i>

Figura 1. Versión original y anonimizada de un artículo extraído del conjunto de datos de validación del Daily Mail. Las entidades anonimizadas están constantemente permutándose durante la prueba y entrenamiento.

De la Figura 1 podemos observar como se extraen las entidades de oraciones clave de los artículos de las noticias para ser reemplazados con espacios en blanco. Para así poder entrenar a la máquina en responder preguntas dependiendo del texto restante.

En este trabajo, se buscará analizar un modelo de red neuronal recurrente que sea capaz de completar oraciones a partir de fragmentos de un libro para niño, se pretende que esto sea posible mediante el uso del framework ParlAI, la cual también es una plataforma que nos ofrece tutoriales de como añadir nuestros propios datasets así o usar los que nos proporcionan, crear nuestros propios modelos, entrenarlos y validarlos, así como implementar agentes que se encargen de diversas tareas.

Demostraremos facilidad que tiene esta plataforma así como la eficacia de nuestro modelo para esta clase de trabajos.

II. OBJETIVO DEL ESTUDIO

Comprender y analizar las redes neuronales recurrentes y como estas permiten que una máquina pueda leer y comprender un cuento para niños.

II-A. Objetivos Específicos

1. Implementar el modelo Seq2Seq en el framework ParlAI para enseñarle a una máquina a leer y comprender.
2. Evaluar la efectividad del modelo Seq2Seq para completar oraciones de un libro para niños.

III. MARCO TEÓRICO

III-A. Natural Language Processing

El procesamiento del lenguaje natural (NLP) es un campo de la ciencia de la computación que se ocupa del uso de las computadoras para la extracción, interpretación, análisis y manipulación del lenguaje humano. Algo relacionado al NLP es el Question answering (QA), el cual es una tarea del NLP encargada de responder a preguntas en el lenguaje natural, aprovechando los métodos de recuperación de información y representación del conocimiento [2]. Siendo de esta manera la forma en la que la máquina aprenderá a leer y comprender .

III-B. Recurrent Neural Network

Las redes neuronales recurrentes (RNN) tienen la capacidad de modelar la información contextual en secuencias largas. Para el procesamiento del lenguaje natural se desea capturar las dependencias a largo plazo; además de mantener el orden contextual entre las palabras para resolver el significado global de un texto. Otra ventaja de las RNN es su capacidad para aprender representaciones de secuencias de longitud variable, como frases, documentos y muestras de habla [2].

III-C. Long Short-Term Memory

La memoria a largo y corto plazo (LSTM) utiliza puertas de entrada, salida y olvido para controlar la propagación del gradiente en la memoria de la red neuronal recurrente (RNN). Esto permite proteger las celdas de memoria que lleven algún estado oculto al siguiente paso. Los estados de activación permiten que la red aprenda las condiciones para cuando olvidar o mantener la información en la celda de memoria [2].

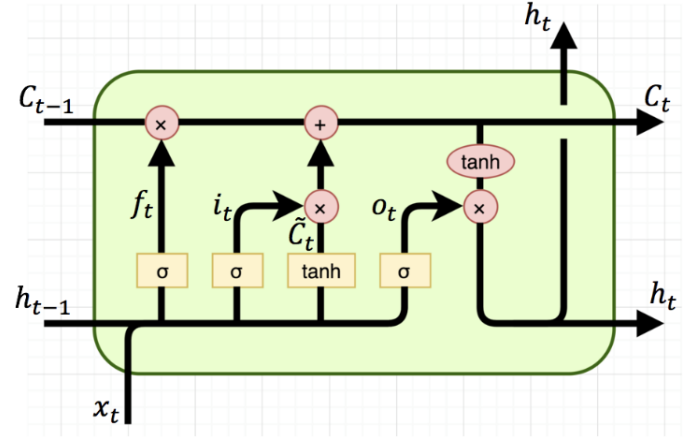


Figura 2. Diagrama de una celda LSTM

III-D. Seq2Seq

El modelo Sequence to Sequence se utiliza cuando tenemos una entrada de longitud variable que no es necesariamente una entidad asociada a otra. Un modelo Seq2Seq se compone más comunmente de dos RNN que trabajan de manera cooperativa: encoder y decoder. Ambas siendo modelos LSTM. Este modelo es usado comúnmente para máquinas de traducción, resumen de texto, modelos de conversación, predicción de palabras, y más.

III-D1. Encoder: Se encargará de codificar el input palabra por palabra en un vector de estado o de contexto (en el caso de LSTM estas son llamadas estado oculto o vectores de estado de las celdas).

III-D2. Decoder: Este es el estado oculto final por parte del codificador. Se encarga de encapsular todos los elementos de entrada para poder realizar una predicción.

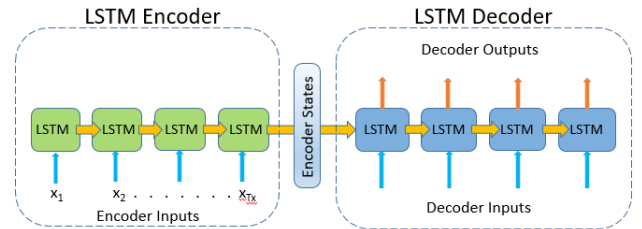


Figura 3. Modelo encoder-decoder con LSTM unidireccionales

En la Figura 2, el encoder codifica una secuencia de entrada pasándolas por las celdas LSTM, este encoder consta de los estados ocultos; luego, se pasa a la decodificación como estados iniciales junto con los

inputs del decodificador para poder hacer predicciones.

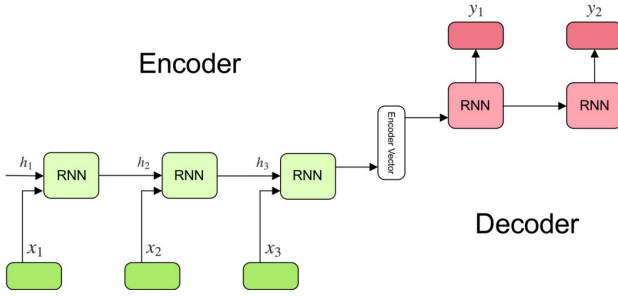


Figura 4. Encoder-decoder

Como podemos observar en la Figura 4, cada rectángulo representa una capa oculta en un paso de tiempo t .

La fórmula para calcular los elementos de salida de la capa oculta en cada paso de tiempo t es [4]:

$$h_t = f(W^{(hh)}h_{t-1} + W^{(hx)}x_t) \quad (1)$$

A continuación, se detalla los parámetros asociados a esta fórmula:

- $x_1, \dots, x_{t-1}, x_t, x_{t+1}, \dots, x_T$: Los vectores de palabras correspondientes a un documento con T palabras.
- W^{hx} : Matriz de pesos utilizada para condicionar el vector de palabras de entrada x_t .
- W^{hh} : Matriz de pesos utilizada para condicionar el vector de palabras de salida h_{t-1} .
- h_{t-1} : Salida de la función no lineal en el paso de tiempo anterior $t - 1$.
- σ : Función de no lineal (sigmoide).

Además, tenemos que la función para calcular la distribución de probabilidad de salida sobre el vocabulario en cada paso del tiempo t es:

$$y_t = \text{softmax}(W^s h_t) \quad (2)$$

Donde W_s es la matriz de pesos utilizada de acuerdo al vocabulario y a la palabra de salida h_{t-1} . e y_t es la siguiente palabra predicha dado el contexto del documento hasta el momento (es decir h_{t-1}) y el último vector palabra observado x_t .

IV. ESTADO DEL ARTE:

En la actualidad existen novedosas arquitecturas de redes neuronales para tareas de comprensión lectora de la máquina. En esta sección describiremos algunas de estas.

IV-A. Deep LSTM Reader

Deep LSTM Reader propuesto en el artículo [Teaching Machines to Read and Comprehend](#) [1] es un modelo que pone a prueba la capacidad de los Deep LSTM

encoder para manejar secuencias largas de inputs. Este modelo se alimenta palabra por palabra de artículos del CNN al Deep LSTM encoder y luego con un delimitador también se introducen consultas al codificador. Como resultado, se obtiene un modelo que procesa cada par de consultas del artículo como una única secuencia larga. Dado el artículo y la consulta, el modelo predice el token del artículo que responde a la consulta.

IV-B. ReasoNet

ReasoNet propuesto en el artículo [ReasoNet: Learning to Stop Reading in Machine Comprehension](#) [3] es una arquitectura de red neuronal llamada red de razonamiento la cual tiene el fin de realizar las tareas de comprensión de la máquina. Las ReasoNets hacen uso de múltiples giros para explotar eficientemente y razonar sobre la relación entre las consultas, los documentos y las respuestas. Con el uso de aprendizaje por refuerzo, las ReasoNets pueden determinar dinámicamente continuar con el proceso de comprensión tras digerir los resultados intermedios, o terminar la lectura cuando concluye que la información existente es adecuada para producir una respuesta.

IV-C. R-NET

ReasoNet propuesto en el artículo [R-Net: Machine Reading Comprehension with self-matching networks](#) [5] es un modelo de redes neuronales de extremo a extremo para responder a preguntas del estilo de la comprensión lectora, cuyo objetivo es responder a preguntas de un pasaje dado. En este artículo primero se empareja la pregunta y el pasaje con redes recurrentes basadas en la atención para obtener la representación del pasaje consciente de la pregunta. Luego propone un mecanismo de atención auto-ajustada para refinar la representación comparando el pasaje consigo mismo, lo que codifica eficazmente la información de todo el pasaje. Por último, emplea redes de punteros para localizar las posiciones de las respuestas de los pasajes.

V. METODOLOGÍA:

En esta sección se describirá los pasos del desarrollo del proyecto y las herramientas necesarias.

V-A. ParlAI

ParlAI es un framework de software de código abierto para la investigación de diálogos implementada en Python, disponible en <http://parl.ai>. Su objetivo es proporcionar un marco unificado para compartir, capacitar y probar modelos de diálogo; integración de Amazon Mechanical Turk para la recopilación de datos, la evaluación humana y el aprendizaje por refuerzo; además cuenta con un [repositorio](#) de modelos de aprendizaje automático para comparar con otros modelos y mejorar las arquitecturas existentes. También ofrece más de 20 tareas, incluye conjuntos de datos

populares como SQuAD, bAbI tasks, MCTest, WikiQA, QACNN, QADailyMail, CBT, bAbI Dialog, Ubuntu, OpenSubtitles y VQA.

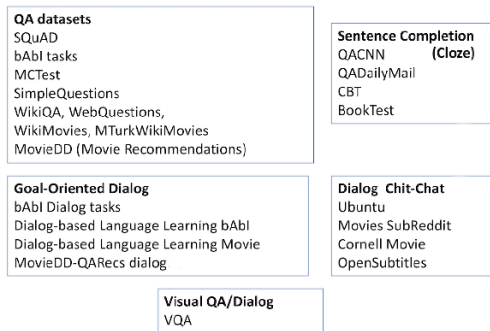


Figura 5. Tareas en la primera versión de ParlAI. Fuente: [ParlAI: A Dialog Research Software Platform](#)

Además de la amplia gama de datasets disponibles, ofrece una amplia gama de ayudantes para crear nuestros propios agentes.

ParlAI integra varios modelos, incluidos modelos neuronales como redes de memoria, Seq2seq y attentive LSTMs, los cuales implementaremos en este proyecto.

VI. EXPERIMENTACIÓN Y RESULTADOS

VI-A. Conjunto de Datos

Para la evaluación del modelo se tomó el dataset [Children'S Book Test \(CBT\)](#), ya que este era el de menor tamaño y esto haría que el modelo cargue más rapido el dataset, en comparación a otros datasets conocidos como CNN y Daily Mail.

CBT es un dataset que está diseñado para medir directamente qué tan bien los modelos lingüísticos pueden explotar un contexto lingüístico más amplio, por esta razón la usaremos para que la máquina aprender a leer y comprender. Cabe mencionar que la CBT se construye a partir de libros que están disponibles gratuitamente, gracias al Proyecto Gutenberg.

VI-B. Experimentación

Entrenamos el modelo con un batch y tres batches, obtuvimos los siguientes resultados:

1. Para un batch

Evaluación para los datos de validación

Evaluación para los datos de test

2. Para tres batches

Evaluación para los datos de validación Evaluación para los datos de test

```

03:50:52 | running eval: valid
04:11:08 | eval completed in 1216.06s
04:11:08 | valid:

```

	accuracy	bleu-4	clen	ctpb	ctps	ctrunc	ctrunclen	exps	exs	f1	gpu_mem	llen	loss	lr	ltpb	ltps
all	.000625	6.25e-13	512.3	512.3	3370	0	0	6.579	8000	.000625	.2580	2.013	9.236	1	2.013	13.24
cbt:CN	.0010	1e-12	522.1			0	0	2000	.0010			2.004	11.07			
cbt:NE	0	0	489			0	0	2000	0			2.047	10.06			
cbt:P	.0015	1.5e-12	528.9			0	0	2000	.0015			2	5.704			
cbt:V	0	0	509.2			0	0	2000	0			2	10.04			
	litrunc	litrunclen	ppl	token_acc	token_em		total_train_updates	tpb	tps							
all	0	0	27608	.3525	.000625		50 514.3	3384								
cbt:CN	0	0	63955	.3403	.0010											
cbt:NE	0	0	23331	.2694	0											
cbt:P	0	0	325	.4740	.0015											
cbt:V	0	0	22820	.3262	0											

Figura 6. Tabla 1

```

04:56:25 | test:

```

	accuracy	bleu-4	clen	ctpb	ctps	ctrunc	ctrunclen	exps	exs	f1	gpu_mem	llen	loss	lr	ltpb	ltps
all	.0003	3.368e-13	524.4	524.4	3456	0	0	6.591	10000	.0003667	.1937	2.029	9.459	1	2.029	13.37
cbt:CN	.0006	8e-13	539.3			0	0	2500	.0006			2.004	11.01			
cbt:NE	0	1.472e-13	503.7			0	0	2500	.0002667			2.13	10.97			
cbt:P	.0004	4e-13	537.1			0	0	2500	.0004			2	5.86			
cbt:V	0	0	517.3			0	0	2500	0			2.001	9.969			
	litrunc	litrunclen	ppl	token_acc	token_em		total_train_updates	tpb	tps							
all	0	0	35458	.3458	.0003		50 526.4	3469								
cbt:CN	0	0	61960	.3154	.0006											
cbt:NE	0	0	58171	.2527	0											
cbt:P	0	0	359.8	.4698	.0004											
cbt:V	0	0	21349	.3455	0											

Figura 7. Tabla 2

VII. ANÁLISIS

Observamos las tablas 1 y 3 para el análisis de la métrica accuracy que nos da la precisión del modelo. Analizando las tablas obtenidas al entrenar el modelo con los datos de evaluación de uno y tres lotes, se nota un incremento del accuracy, de 0.000625 % a 0.007875 % .

Ahora observamos las tablas 2 y 4 para el análisis de la métrica accuracy, analizando las tablas obtenidas con los datos de test de uno y tres lotes, se nota un incremento del accuracy, de 0.0003 % a 0.0088 % . Ambos incrementos del accuracy ocurrieron por la modificación de la cantidad de lotes, ya que esto hizo que el modelo se beneficie.

VIII. CONCLUSIONES

Se logró comprender como trabajan las RNN en diferentes modelos y como estas permiten que las máquinas sean capaces de aprender a leer y comprender.

- Se logró implementar el modelo Seq2Seq en el ParlAI para poder leer y comprender un libro para niños.
- Se concluyó que el modelo Seq2Seq no era el mejor para esta clase de tareas dado que su accuracy fue mucho menor en comparación a otros modelos.

REFERENCIAS

- [1] Teaching machines to read and comprehend. *Avances en los sistemas de procesamiento de información neuronal*.
- [2] Uday Kamath, John Liu, and James Whitaker. *Deep learning for NLP and speech recognition*, volume 84. Springer, 2019.
- [3] Yelong Shen, Po-Sen Huang, Jianfeng Gao, and Weizhu Chen. Reasonet: Learning to stop reading in machine comprehension. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1047–1055, 2017.
- [4] Richard Socher and Richard Socher Mundra. Cs 224d: Deep learning for nlp4. 2015.
- [5] Wenhui Wang, Nan Yang, Furu Wei, Baobao Chang, and Ming Zhou. R-net: Machine reading comprehension with self-matching networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 189–198, 2017.

```

05:24:48 | valid:
accuracy bleu-4 clen ctpb ctps ctrunc ctruncflen exps exs f1 gpu_mem llen loss lr ltpb ltps \
all .007875 7.892e-12 512.3 1536 8201 0 0 16.01 8000 .007938 .1936 2.013 9.903 1 6.035 32.22 \
cvt:CN .0010 1e-12 522.1 0 0 2000 .0010 2.004 11.36
cvt:NE 0 6.767e-14 489 0 0 2000 .00025 2.047 10.13
cvt:P .0305 3.05e-11 528.9 0 0 2000 .0305 2 7.11
cvt:V 0 0 500.2 0 0 2000 0 2 11.02
litrunc ltruncflen ppl token_acc token_em total_train_updates tpb tps
all 0 0 43190 .3975 .00775 51 1542 8233
cvt:CN 0 0 85669 .3787 .0010
cvt:NE 0 0 25088 .3464 0
cvt:P 0 0 1224 .4975 .0300
cvt:V 0 0 60780 .3675 0

```

Figura 8. Tabla 3

```

05:35:27 | test:
accuracy bleu-4 clen ctpb ctps ctrunc ctruncflen exps exs f1 gpu_mem llen loss lr ltpb ltps \
all .0088 8.8e-12 524.4 1572 8216 0 0 15.67 10000 .0088 .1938 2.029 10 1 6.083 31.79
cvt:CN 0 0 539.4 0 0 2500 0 2.084 11.25
cvt:NE 0 0 503.7 0 0 2500 0 2.11 10.7
cvt:P .0352 3.52e-11 537.1 0 0 2500 .0352 2 7.159
cvt:V 0 0 517.3 0 0 2500 0 2.001 10.9
litrunc ltruncflen ppl token_acc token_em total_train_updates tpb tps
all 0 0 44372 .3888 .0087 51 1578 8248
cvt:CN 0 0 76860 .3615 0
cvt:NE 0 0 44272 .3062 0
cvt:P 0 0 1286 .4996 .0348
cvt:V 0 0 54270 .3880 0

```

Figura 9. Tabla 4