

UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD
DEL CUSCO
FACULTAD DE INGENIERÍA ELÉCTRICA, ELECTRÓNICA, INFORMÁTICA Y
MECÁNICA
ESCUELA PROFESIONAL DE INGENIERÍA INFORMÁTICA Y DE SISTEMAS



TESIS

CHATBOT GENERATIVO EN EL IDIOMA ESPAÑOL UTILIZANDO
LA ARQUITECTURA DE RED NEURONAL TRANSFORMER

Para optar al título profesional de:

INGENIERO INFORMÁTICO Y DE SISTEMAS

Presentado por:

BR. DEYBIS YOUL PAUCAR PINTO

BR. MARYCEL USCAMAITA QUISPETUPA

Asesor:

DR. RONY VILLAFUERTE SERNA

CUSCO - PERÚ

2019

- Entrenamiento “*OpenSubtitles*”: Utilizaron el dataset “*OpenSubtitles*” con un vocabulario total de 100K y debido a problemas de memoria, limitaron el `batch_size` a 2048. El modelo fue entrenado para un total de 1M steps, lo que llevó aproximadamente 5,5 días en una GPU Nvidia GeForce GTX 1070.
- Su último entrenamiento consistió en combinar el conjunto de datos del entrenamiento “*OpenSubtitles*” con los parámetros del modelo entrenado de “*Cornell Movie-Dialog with speakers*” teniendo un total de vocabulario de 103k y por problemas de memoria, limitaron el `batch_size` a 2048. El modelo recibió entrenamiento adicional para los 675K steps, que tardaron aproximadamente 3.5 días en una GPU Nvidia GeForce GTX 1070.

Comentario: Esta publicación científica solo nos ayudará a comprender en profundidad los modelos recientes relacionados con los chatbots, como son las redes LSTM, redes Encoder-Decoder, la atención y el transformer, ya que el investigador no presenta una conclusión definida.

3. (Vaswani et al., 2017), “*ATTENTION IS ALL YOU NEED*”, Paper consultado de la Universidad de Cornell. Nueva York.

Un paper que propone una nueva arquitectura de red simple, el Transformer, basada únicamente en mecanismos de atención, prescindiendo completamente de recurrencias y convoluciones. Los experimentos en dos tareas de traducción automática muestran que estos modelos son superiores en calidad, mientras que son más paralelizables y requieren mucho menos tiempo para entrenar. El modelo logra 28.4 BLEU en la tarea de traducción de inglés a alemán de WMT 2014, superando los resultados existentes, incluidos los ensambles, en más de 2 BLEU. En la tarea de traducción de inglés a francés de WMT 2014, el modelo establece un nuevo puntaje BLEU de modelo único de 41.8 después de un entrenamiento de 3,5 días en ocho GPUs, a penas una pequeña fracción de los costos de entrenamiento de los mejores modelos de la literatura. También se muestra que transformer se generaliza bien a otras tareas, tanto con datos de entrenamiento grandes como limitados.

Comentario: Este paper, es el más importante para nuestra información bibliográfica, porque presenta por primera vez a transformer basada exclusivamente en mecanismos de atención. Además, explica, por qué esta arquitectura de red neuronal es superior a las demás en problemas de traducción.