

Resumen

¿Modelos de lenguaje como bases de conocimiento?

Los modelos de lenguaje de alta capacidad que son previamente entrenados como ELMo y BERT se han vuelto cada vez más importantes para el procesamiento del Lenguaje Neuronal, estos están optimizados para predecir la siguiente palabra enmascarada en cualquier parte que se encuentre, los parámetros de estos modelos almacenan una gran cantidad de conocimientos lingüísticos que serán útiles para posteriores tareas. Por otra parte tenemos también a las bases de conocimiento, las cuales son una solución efectiva para acceder a datos relacionales estandarizados, lo cual nos permite hacer consultas como la siguiente: Dante, nacido en, X, pero en ocasiones es necesario extraer datos de otras modalidades o textos para poder poblar las bases de conocimiento, este requiere grandes procesamiento de lenguaje neuronal que involucren extracción de entidades, resolución de coreferencias entre otras componentes que necesiten asistencia humana, supervisión o esquemas fijos, en lugar de esto podemos utilizar modelos de lenguaje neuronal en busca de datos relacionales. El texto también menciona a LAMA (Language Model Analysis) que es un conjunto de fuentes de conocimiento, cada una de ellas compuesta por hechos, este es utilizado para evaluar la capacidad de los distintos modelos de lenguajes para responder preguntas de conocimiento fácticos y de sentido común, como lo es BERT-large quien supera considerablemente a otros modelos lingüísticos para estos trabajos. En la sección del texto llamada Background se nos da información estadística sobre los modelos de lenguaje unidireccionales y

Como se utilizan para estimar la probabilidad de una secuencia de palabras que se representa como $w = [w_1, w_2, \dots, w_n]$, modelos de lenguaje unidireccionales comúnmente asignan una probabilidad $p(w)$ a la secuencia factorizándola como se mostrará a continuación:
$$p(w) = \prod_t p(w_t | w_{t-1}, \dots, w_1).$$

Una forma común de estimar esta probabilidad es utilizando modelos de lenguaje neuronal. Con:

$$p(w_t) = w_t | w_{t-1}, \dots, w_1 = \text{softmax}(W h_t + b)$$

Donde $h_t \in \mathbb{R}^{V \times K}$ es una matriz de parámetros aprendidos para h_t .

Este texto nos habla sobre los modelos de lenguaje pre-entrenados y como estos tienen un gran potencial como bases de conocimiento para mejorar la comprensión del lenguaje neuronal, lo que los convierte en una alternativa prometedora a las bases de conocimiento tradicionales, sin embargo también se habla sobre las limitaciones con las que cuentan estos modelos en la comprensión de ciertos tipos de conocimiento, estos desafíos se tienen que abordar para poder ser aprovechados al máximo.