

Deep contextualized word representations

IIC3692-Alvaro Soto

30/08/2023

Presentado por: Daniel Gómez



Hay limitaciones no superadas

- No toda la información de una palabra en el contexto de la frase se logra guardar (sintaxis y semántica).
- Puede haber palabras con más de un significado (polisemia).

Veamos un ejemplo simple

- La planta debe recibir luz
- Se clavó una astilla en la planta del pie.

¿Qué soluciones se han propuesto?

- Enriquecer los vectores con subpalabras (e.g., Wieting et al., 2016; Bojanowski et al., 2017)
- Aprender distintos vectores de una misma palabra para sus distintos contextos/significados (e.g., Neelakantan et al., 2014).
- Aprender representaciones dependientes del contexto a traves de biLSTM (Belinkov et al. (2017)).

Embeddings from Language Models(ELMo)



- Las representaciones de ELMo son una función de una red bidireccional
- Los embeddings son una función lineal de los estados internos de la red (es un embedding contextual)
- Podemos incorporar ELMo a distintas arquitecturas de redes neuronales de NLP

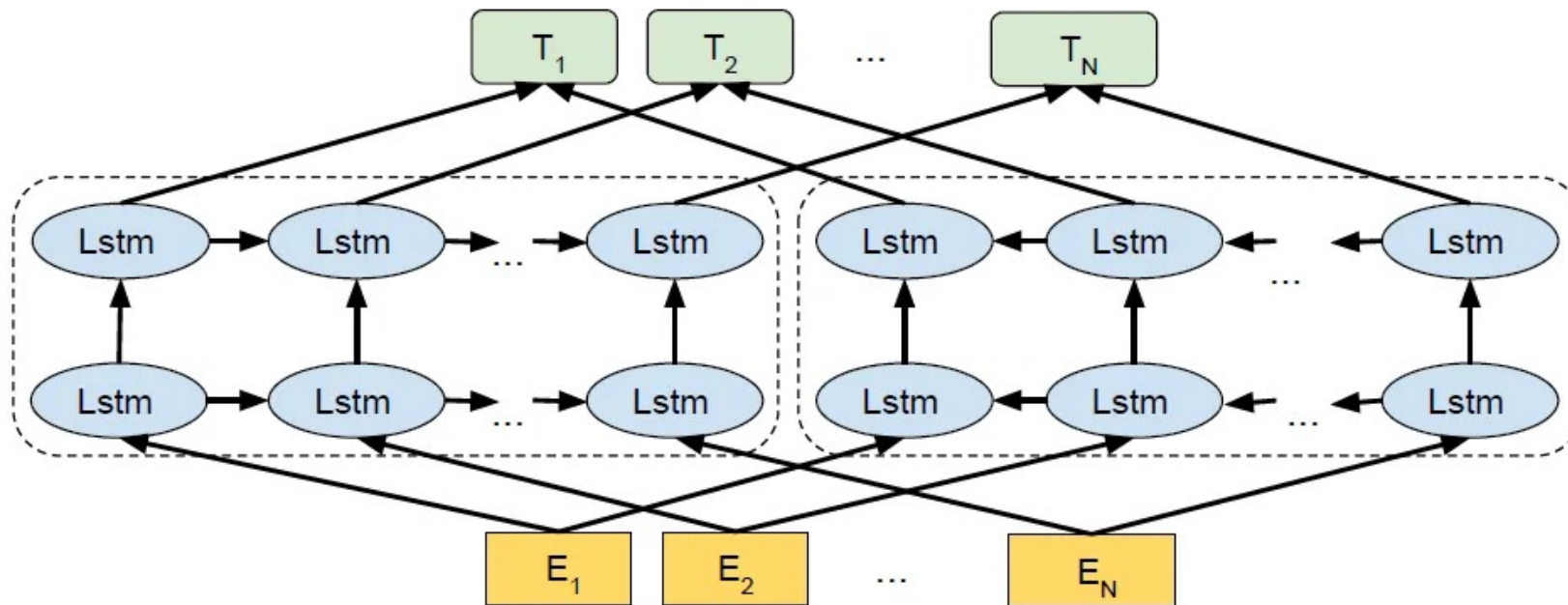
Como funciona ELMo

Empezamos con dos redes neuronales

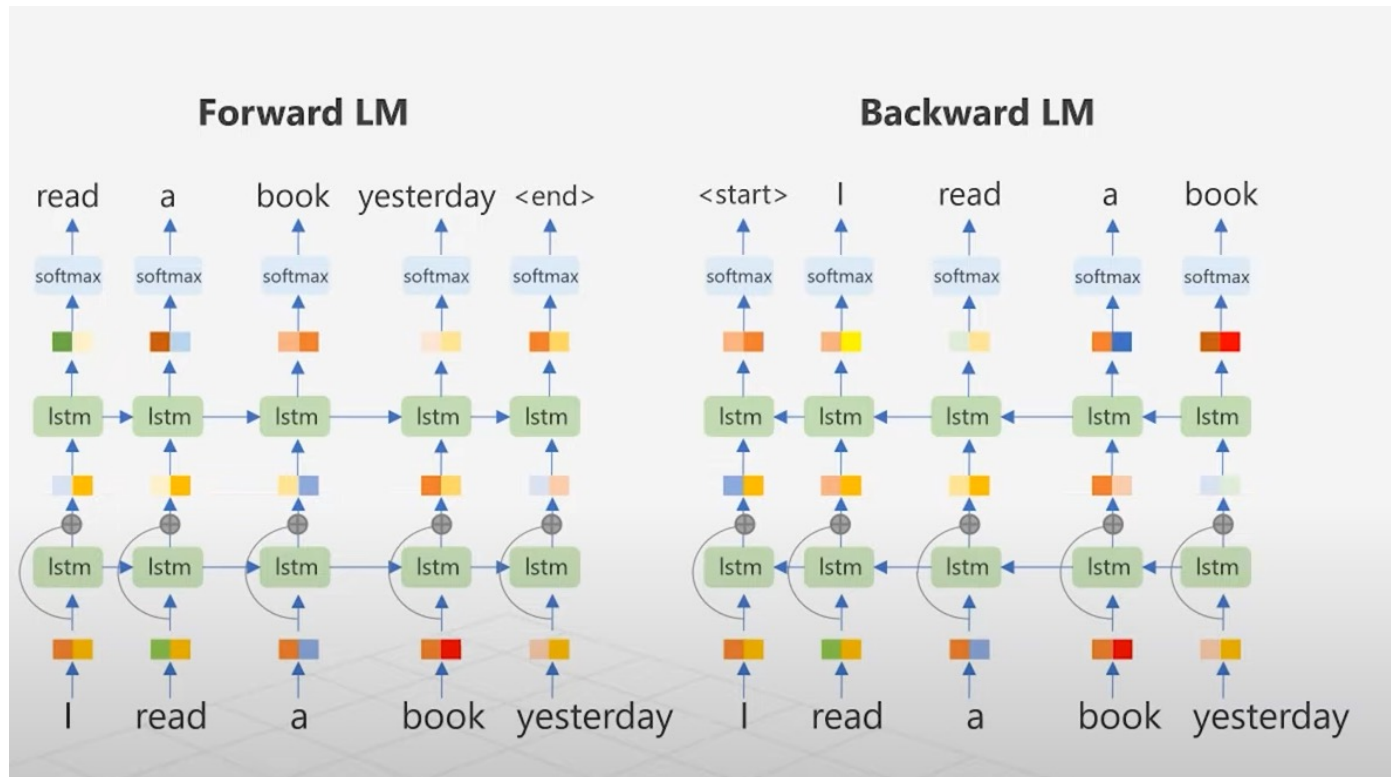


ELMo

Las unimos



Como funciona ELMo



$$R_k = \{\mathbf{x}_k^{LM}, \vec{\mathbf{h}}_{k,j}^{LM}, \overleftarrow{\mathbf{h}}_{k,j}^{LM} \mid j = 1, \dots, L\}$$
$$= \{\mathbf{h}_{k,j}^{LM} \mid j = 0, \dots, L\},$$

$$\text{ELMo}_k = E(R_k; \Theta_e).$$

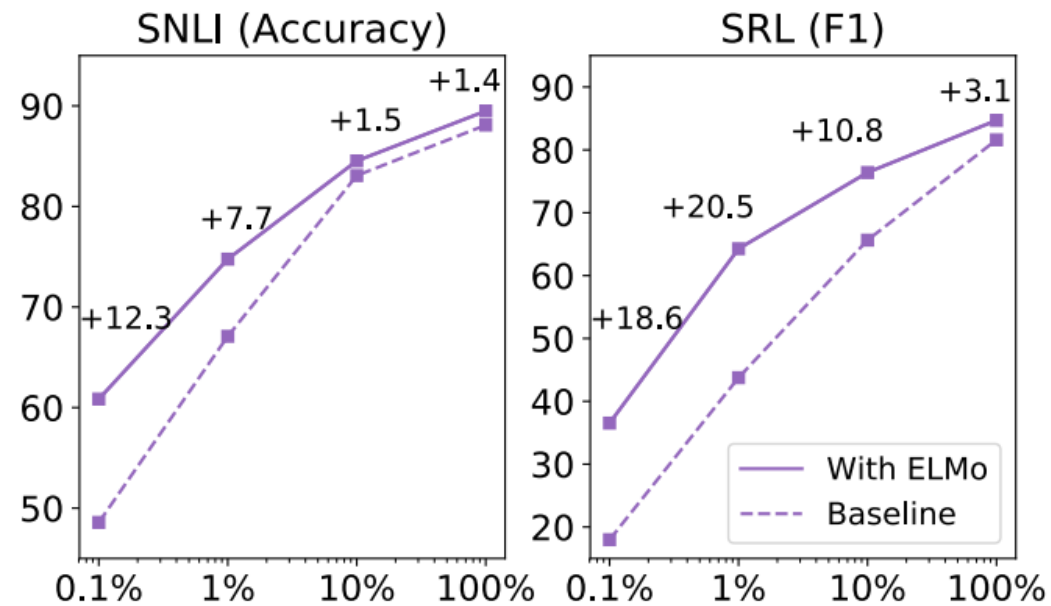
$$\text{ELMo}_k^{\text{task}} = E(R_k; \Theta^{\text{task}}) = \gamma^{\text{task}} \sum_{j=0}^L s_j^{\text{task}} \mathbf{h}_{k,j}^{LM}.$$

Resultados

	TASK	PREVIOUS SOTA		OUR BASELINE	ELMo + BASELINE	INCREASE (ABSOLUTE/ RELATIVE)
Question answering	SQuAD	Liu et al. (2017)	84.4	81.1	85.8	4.7 / 24.9%
Textual entailment	SNLI	Chen et al. (2017)	88.6	88.0	88.7 \pm 0.17	0.7 / 5.8%
Semantic role labeling	SRL	He et al. (2017)	81.7	81.4	84.6	3.2 / 17.2%
Coreference resolution	Coref	Lee et al. (2017)	67.2	67.2	70.4	3.2 / 9.8%
Named entity extraction	NER	Peters et al. (2017)	91.93 \pm 0.19	90.15	92.22 \pm 0.10	2.06 / 21%
Sentiment analysis	SST-5	McCann et al. (2017)	53.7	51.4	54.7 \pm 0.5	3.3 / 6.8%

Resultados

Modelos con ELMo usan
sets de entrenamiento
más eficientemente





Conclusiones

- ELMo permite guardar mas informacion contextual en los embeddings
- Usando multiples layers del biLM se mejora el aprendizaje de características sintacticas y semanticas de las frases
- ELMo permite tener una mejora en multiples pruebas de NLP



Análisis Crítico

- No se busca ver cuál es el futuro en investigación para ELMo
- Falta analizar más falencias del modelo