Curso Data Science

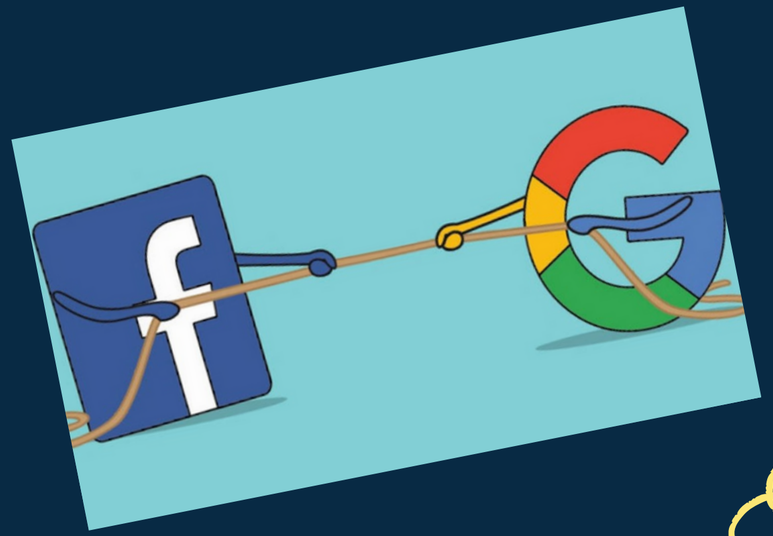
RoBERTa

NICOLL CRESPO RONCALLO

DEFINICION

RoBERTa es un modelo de PLN desarrollado por Facebook

Está basado en el modelo BERT. Pero se re-entrenó con mejoras en su metodología, más datos y mayor poder computacional, obteniendo mejores resultados en varias métricas.



The language model family

ELMO

Transformer

ULMFit

BERT

Bi-directional LM

GPT

+Data
+Model size

GPT-2

Defense

Grover

Permutation LM
TransformerXL
+Data

XLNet

XLM

Crosslingual

MT-DNN

Multitask

+Knowledge graph

ERNIE (Tsinghua)

Text2Text
+Data

T5

Masked spans

SpanBERT

+Data
+Train time
+Batchsize
-NSP

RoBERTa

Shared weights
NSP → Sentence order
Factorized Embedding

ALBERT

-Model size

Q-BERT

DistilBERT

deepset.ai

METODOLOGIA

- **IMPLEMENTADO EN PYTORCH**
  vs. TensorFlow
- **ELIMINACION DEL ENTRENAMIENTO NSP**
  (Next Sentence Prediction)
- **USO DE LOTES MÁS GRANDES**
  (8000 VS. 256)
- **ENTRENAMIENTO CON 160 GB DE TEXTO**
  vs. 16 GB
- **USO DE ENMARCARAMIENTO DINAMICO**
  VS. Estatico
- **MAYOR TIEMPO DE ENTRENAMIENTO**
  4-5 veces mas

RESULTADOS

- SUPERÓ A BERT Y XLNET EN LAS TAREAS DE RACE Y DE GLUE
- PUNTUACIÓN PROMEDIO MÁS ALTA EN LA TABLA DE CLASIFICACION GENERAL DE GLUE

  (88,5)
- NUEVO PUNTAJE MAS ALTO EN LA TABLA DE CLASIFICACIÓN PÚBLICA DE SQUAD

	MNLI	QNLI	QQP	RTE	SST	MRPC	CoLA	STS	WNLI	Avg
Single-task single models on dev										
$\text{BERT}_{\text{LARGE}}$	86.6/-	92.3	91.3	70.4	93.2	88.0	60.6	90.0	-	-
$\text{XLNet}_{\text{LARGE}}$	89.8/-	93.9	91.8	83.8	95.6	89.2	63.6	91.8	-	-
RoBERTa	**90.2/90.2**	**94.7**	**92.2**	**86.6**	**96.4**	**90.9**	**68.0**	**92.4**	**91.3**	-
Ensembles on test (from leaderboard as of July 25, 2019)										
ALICE	88.2/87.9	95.7	**90.7**	83.5	95.2	92.6	**68.6**	91.1	80.8	86.3
MT-DNN	87.9/87.4	96.0	89.9	86.3	96.5	92.7	68.4	91.1	89.0	87.6
XLNet	90.2/89.8	98.6	90.3	86.3	**96.8**	**93.0**	67.8	91.6	**90.4**	88.4
RoBERTa	**90.8/90.2**	**98.9**	90.2	**88.2**	96.7	92.3	67.8	**92.2**	89.0	**88.5**

Model	SQuAD 1.1		SQuAD 2.0	
	EM	F1	EM	F1
Single models on dev, w/o data augmentation				
$\text{BERT}_{\text{LARGE}}$	84.1	90.9	79.0	81.8
$\text{XLNet}_{\text{LARGE}}$	**89.0**	94.5	86.1	88.8
RoBERTa	88.9	**94.6**	**86.5**	**89.4**
Single models on test (as of July 25, 2019)				
$\text{XLNet}_{\text{LARGE}}$			86.3†	89.1†
RoBERTa			86.8	89.8
XLNet + SG-Net Verifier			**87.0**†	**89.9**†

Model	Accuracy	Middle	High
Single models on test (as of July 25, 2019)			
$\text{BERT}_{\text{LARGE}}$	72.0	76.6	70.1
$\text{XLNet}_{\text{LARGE}}$	81.7	85.4	80.2
RoBERTa	**83.2**	**86.5**	**81.3**

APLICACION

Wang, Y., Sun, Y., Ma, Z., Gao, L., Xu, Y., & Sun, T. (2020, August). Application of pre-training models in named entity recognition.

CONSIDERACIONES

- Problema de NER
- Uso del MSRA-2006 dataset
- Uso de 2 epocas
- Razon de aprendizaje de 5e-5
- Tamaño de lote de 16

MODELOS

- BERT
- ERNIE
- ERNIE2.0-tiny
- RoBERT

RESULTADOS

THE RESULTS OF NER USING DIFFERENT PRE-TRAINING MODELS

Models	Precision/%	Recall/%	F1/%
Baseline	92.54	88.20	90.32
BERT-base	92.68	94.18	93.30
ERNIE	92.92	94.07	93.37
ERNIE-tiny	83.89	89.88	86.52
RoBERTa	**93.64**	**94.93**	**94.17**

REFERENCIAS

1. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.
2. Wang, Y., Sun, Y., Ma, Z., Gao, L., Xu, Y., & Sun, T. (2020, August). Application of pre-training models in named entity recognition. In 2020 12th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC) (Vol. 1, pp. 23-26). IEEE.