

Asignación automática de dominios a definiciones

WordNet BabelDomains

Josu Barrutia

20 Diciembre, 2023

Facultad de Informática, UPV/EHU

1. Introducción y Conceptos

2. Datos

3. Modelos Preentrenados

Búsqueda de Hiperparámetros

distilBERT

BERT

roBERTa

4. Few-Shot

Introducción y Conceptos

WordNet es una base de datos léxica en la que los sustantivos, verbos, adjetivos y adverbios se agrupan en conjuntos de sinónimos (synsets), cada uno expresando un concepto distinto.

Ejemplo

```
00002730 00 a 01 acroscopic 0 002 ;c 06066555 n 0000 ! 00002843 a  
0101 | facing or on the side toward the apex
```

BabelDomain es un recurso que incluye información de dominios de con un total de 34 etiquetas. Aunque la asignación se anota de forma semiautomática, se proporciona un conjunto estándar de oro para WordNet.

Dada una glosa clasificar el concepto en un dominio de
BabelDomains

- Clasificación de texto
- Multiclase

Para evaluar el rendimiento de nuestro modelo utilizaremos:

- Accuracy
- F1
- Recall
- Precision

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (3)$$

Maneras de hacer el average

- Macro → sensitivo a dataset desbalanceado
- Micro → $\text{recall} = \text{precision} = \text{accuracy} = f1$
- Weighted

Datos

Los ficheros han sido descargados de *BabelDomains*₁

babeldomains_wordnet.txt

n04789689 Philosophy and psychology
1.0

v00891734 Business, economics, and
finance 1.0

n06610332 Language and linguistics 1.0

n05110583 Mathematics 1.0

n07038615 Music 1.0

wordnet_dataset_gold.txt

n05875723 Physics and astronomy

n13424183 Physics and astronomy

n13549311 Physics and astronomy

n13587763 Physics and astronomy

n08498580 Physics and astronomy

[1] <http://lcl.uniroma1.it/babeldomains/>

Problema de Desbalance

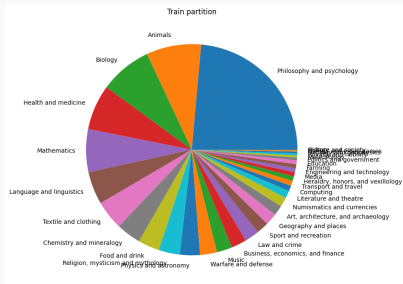


Figure 1: Dataset de entrenamiento

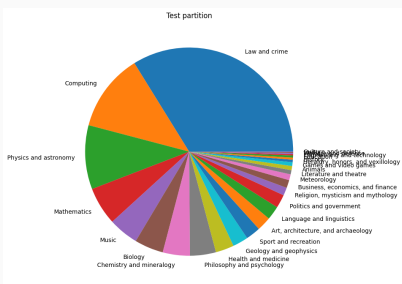


Figure 2: Dataset de test

Modelos Preentrenados

Esta tarea de clasificación de texto es ideal para la arquitectura
Transformers

- distilBERT (distilbert-base-uncased)
- BERT (bert-base-uncased)
- roBERTa (roberta-base)

Tasa de aprendizaje

Una tasa de aprendizaje adecuada es clave para evitar el "Catastrophic Forgetting" (Olvido Catastrófico).

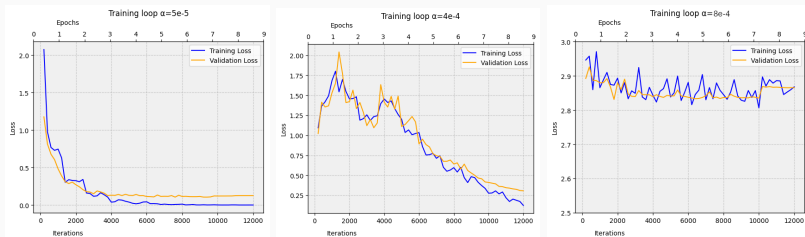


Figure 3: Evolución de la pérdida con diferentes Tasas de Aprendizaje

Usaremos tasas de aprendizaje entre $2e-5$ y $5e-5$

Tamaño Batch de Entrenamiento

El tamaño de batch de entrenamiento para el fine-tuning no es relevante, dependerá del tamaño de la secuencia máxima¹.

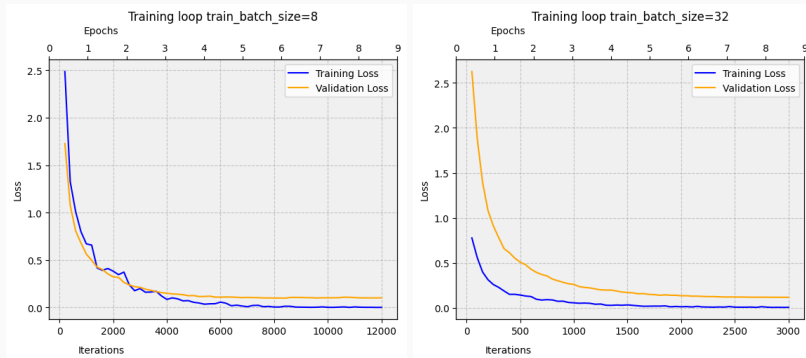


Figure 4: Evolución de la pérdida con diferente Tamaño de Batch

Número de Épocas

El número de épocas que necesita el modelo para aprender la tarea de clasificación es bastante pequeño.

3 épocas para las tareas GLUE₁.

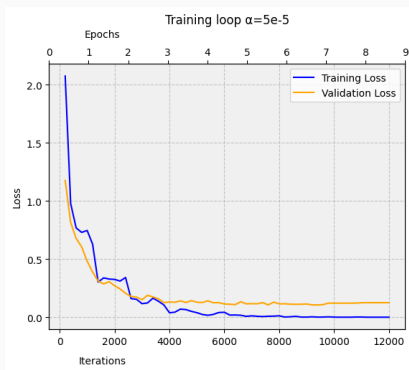


Figure 5: Evolución de la pérdida en entrenamiento y validación

Capas de clasificación

Añadir dos capas lineales 768x768 y 768x34 y un dropout de 0.2 a la salida

```
DistilBertForSequenceClassification(  
  (distilbert): DistilBertModel(  
    (embeddings): Embeddings(  
      (word_embeddings): Embedding(30522, 768, padding_idx=0)  
      (position_embeddings): Embedding(512, 768)  
      (LayerNorm): LayerNorm((768,), eps=1e-12, elementwise_affine=True)  
      (dropout): Dropout(p=0.1, inplace=False)  
    )  
    (transformer): Transformer(  
      (layer): ModuleList(  
        (0-5): 6 x TransformerBlock(  
          (attention): MultiHeadSelfAttention(  
            (dropout): Dropout(p=0.1, inplace=False)  
            (q_lin): Linear(in_features=768, out_features=768, bias=True)  
            (k_lin): Linear(in_features=768, out_features=768, bias=True)  
            (v_lin): Linear(in_features=768, out_features=768, bias=True)  
            (out_lin): Linear(in_features=768, out_features=768, bias=True)  
          )  
          (sa_layer_norm): LayerNorm((768,), eps=1e-12, elementwise_affine=True)  
          (ffn): FFN(  
            (dropout): Dropout(p=0.1, inplace=False)  
            (lin1): Linear(in_features=768, out_features=3072, bias=True)  
            (lin2): Linear(in_features=3072, out_features=768, bias=True)  
            (activation): GELUActivation()  
          )  
          (output_layer_norm): LayerNorm((768,), eps=1e-12, elementwise_affine=True)  
        )  
      )  
    )  
    (pre_classifier): Linear(in_features=768, out_features=768, bias=True)  
    (classifier): Linear(in_features=768, out_features=34, bias=True)  
    (dropout): Dropout(p=0.2, inplace=False)  
  )  
)
```

Tasa Aprendizaje	Planificador	Épocas	Accuracy validación
2e-5	Lineal sin Warm-Up	4	0.9352
2e-5	Lineal sin Warm-Up	10	0.9774
4e-4	Lineal sin Warm-Up	10	0.9335
5e-5	Lineal sin Warm-Up	4	0.9602
5e-5	Lineal sin Warm-Up	6	0.9723
5e-5	Lineal Warm-Up 100	4	0.9628
5e-5	Coseno Warm-Up 100	4	0.9606
5e-5	Coseno Warm-Up 100	6	0.9746

Table 1: Búsqueda hiperparámetros distilBERT

	Accuracy	F1	Precision	Recall
Partición Train	0.8792	0.8765	0.8910	0.8792
Partición Train + Dev	0.8643	0.8608	0.8788	0.8643

Table 2: Rendimiento en Test.

Tasa Aprendizaje	Planificador	Épocas	Accuracy validación
2e-5	Lineal sin Warm-Up	4	0.9369
2e-5	Lineal sin Warm-Up	10	0.9763
5e-5	Lineal sin Warm-Up	4	0.9634
5e-5	Lineal sin Warm-Up	10	0.9819
3e-5	Lineal Warm-Up 100	4	0.9583
5e-5	Coseno Warm-Up 100	4	0.9769
5e-5	Coseno Warm-Up 100	6	0.9758

Table 3: Búsqueda hiperparámetros BERT

	Accuracy	F1	Precision	Recall
Partición Train	0.8526	0.8555	0.8800	0.8525
Partición Train + Dev	0.8662	0.8638	0.8822	0.8662

Table 4: Rendimiento en Test

Tasa Aprendizaje	Planificador	Épocas	Accuracy validación
2e-5	Lineal sin Warm-Up	4	0.9583
2e-5	Lineal sin Warm-Up	6	0.9651
5e-5	Lineal sin Warm-Up	4	0.9561
5e-5	Lineal sin Warm-Up	6	0.9645
2e-5	Cosene Warm-Up 100	4	0.9639
2e-5	Coseno Warm-Up 100	6	0.9707
5e-5	Coseno Warm-Up 100	4	0.9696
5e-5	Coseno Warm-Up 100	6	0.9702

Table 5: Búsqueda hiperparámetros roBERTa

	Accuracy	F1	Precision	Recall
Partición Train	0.8467	0.8482	0.8690	0.8467
Partición Train + Dev	0.8597	0.8593	0.8766	0.8597

Table 6: Rendimiento en Test

Input pattern	Top-1	Top-3	Top-5
Topic: [label]	59.61	69.48	74.02
Domain: [label]	58.50	67.40	72.27
Theme: [label]	59.67	73.96	81.36
Subject: [label]	60.58	69.74	74.35
Is about [label]	73.37	87.72	91.94
Topic or domain about [label]	78.44	87.46	89.74
The topic of the sentence is about [label]	80.71	92.92	95.77
The domain of the sentence is about [label]	81.62	93.96	96.42
The topic or domain of the sentence is about [label]	76.62	88.63	91.23

Table 3: Some of the explored *input patterns* for the MNLI approach and their Top-1, Top-3 and Top-5 accuracy.

Few-Shot

- Disminuir el tamaño del batch
- Aumentar el número de épocas hasta 20

Número Instancias/Clase	Accuracy	F1	Precision	Recall
5	0.7766	0.7969	0.8479	0.7766
10	0.7597	0.7756	0.8363	0.7597
20	0.8279	0.8433	0.8786	0.8279

Table 7: Rendimiento en Test

Preguntas?