



# Information Retrieval


Luís Filipe Cunha

lfc@di.uminho.pt

José João Almeida

jj@di.uminho.pt

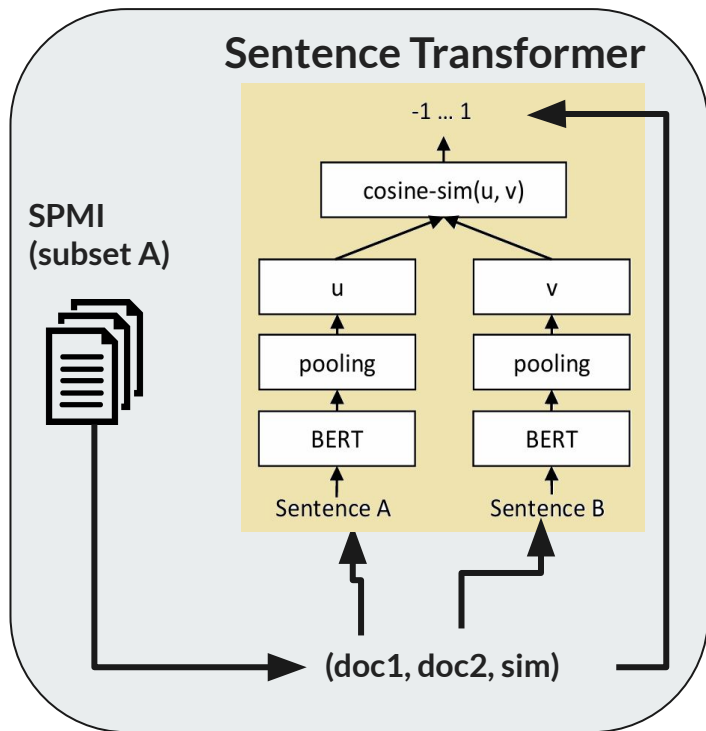




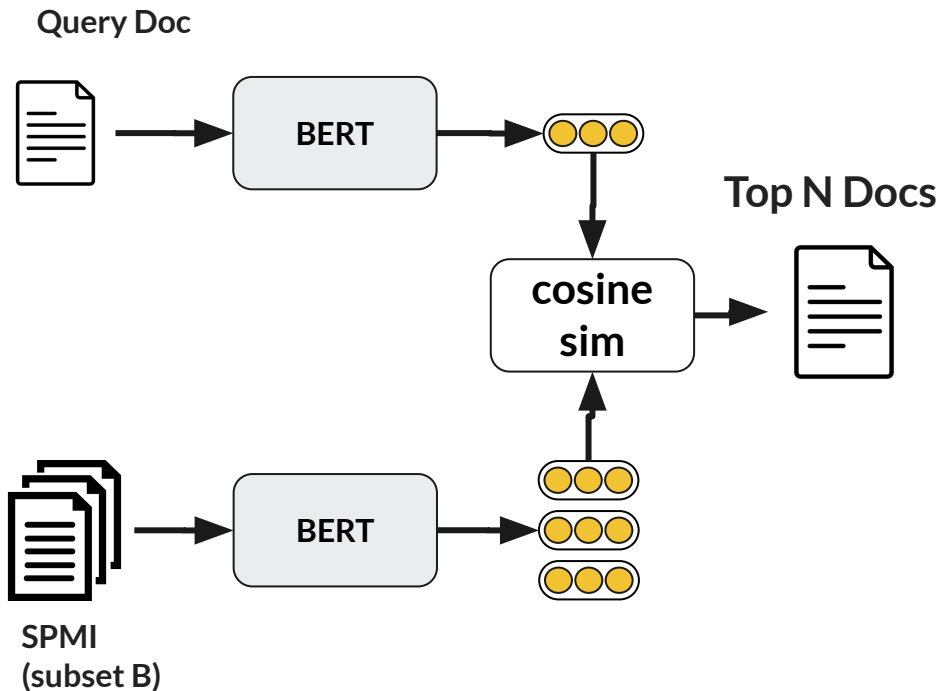
Model	STS12	STS13	STS14	STS15	STS16	STSb	SICK-R	Avg.
Avg. GloVe embeddings	55.14	70.66	59.73	68.25	63.66	58.02	53.76	61.32
Avg. BERT embeddings	38.78	57.98	57.98	63.15	61.06	46.35	58.40	54.81
BERT CLS-vector	20.16	30.01	20.09	36.88	38.08	16.50	42.63	29.19
InferSent - Glove	52.86	66.75	62.15	72.77	66.87	68.03	65.65	65.01
Universal Sentence Encoder	64.49	67.80	64.61	76.83	73.18	74.92	<b>76.69</b>	71.22
SBERT-NLI-base	70.97	76.53	73.19	79.09	74.30	77.03	72.91	74.89
SBERT-NLI-large	72.27	<b>78.46</b>	<b>74.90</b>	80.99	76.25	<b>79.23</b>	73.75	76.55
SROBERTa-NLI-base	71.54	72.49	70.80	78.74	73.69	77.77	74.46	74.21
SROBERTa-NLI-large	<b>74.53</b>	77.00	73.18	<b>81.85</b>	<b>76.82</b>	79.10	74.29	<b>76.68</b>

# RoadMap

## Model Training



## Model Inference



ARTICLE



# MedLink: Retrieval and Ranking of Case Reports to Assist Clinical Decision Making

**Authors:**  [Luís Filipe Cunha](#),  [Nuno Guimarães](#),  [Alexandra Mendes](#),  [Ricardo Campos](#),  [Alípio Jorge](#) | [Authors Info & Claims](#)

Advances in Information Retrieval: 47th European Conference on Information Retrieval, ECIR 2025, Lucca, Italy, April 6–10, 2025, Proceedings, Part V

Pages 72 - 77 • [https://doi.org/10.1007/978-3-031-88720-8\\_13](https://doi.org/10.1007/978-3-031-88720-8_13)

**Published:** 07 April 2025 [Publication History](#)

 0  0



# Toolkits

```
#Calculo de combinações
from itertools import combinations
combinations(lista, 2) # pares

#Balancear dados
from sklearn.utils import resample
undersampled_majority_class = resample(majority_class,
                                       replace=False,      # Don't duplicate samples
                                       n_samples= len(minority_class), # Match minority
                                       random_state=42)

#Criar train, test split estratificados
from sklearn.model_selection import train_test_split
scores = [score for _, _, score, *_ in balanced_abstract_pairs]
train_data, test_data = train_test_split(
    balanced_abstract_pairs,
    test_size=0.2,
    random_state=42,
    stratify=scores
)
```

# Model training

```
from sentence_transformers import SentenceTransformerTrainer, SentenceTransformerTrainingArguments
from sentence_transformers.similarity_functions import SimilarityFunction
from sentence_transformers.evaluation import EmbeddingSimilarityEvaluator
```

```
args = SentenceTransformerTrainingArguments(
    # Required parameter:
    output_dir="models/sentence_transformers/bert-base-portuguese",
    report_to="none",
    # Optional training parameters:
    num_train_epochs=5,
    per_device_train_batch_size=16,
    per_device_eval_batch_size=16,
    fp16=True, # Set to False if you get an error that your GPU can't run on FP16
    bf16=False, # Set to True if you have a GPU that supports BF16
    # Optional tracking/debugging parameters:
    eval_strategy="epoch",
    save_strategy="epoch",
    save_total_limit=2,
    load_best_model_at_end=True,
)
```

# Model training

```
# Create the evaluator
dev_evaluator = EmbeddingSimilarityEvaluator(
    test_dataset['abstract1'], # Assuming these are the sentence pairs for evaluation
    test_dataset['abstract2'],
    test_dataset['score'], # Assuming this contains the similarity scores
    main_similarity=SimilarityFunction.COSINE,
)

# 6. Create the trainer & start training
trainer = SentenceTransformerTrainer(
    model=model,
    args=args,
    train_dataset=train_dataset,
    eval_dataset=test_dataset,
    loss=loss,
    evaluator=dev_evaluator,
)
```

# Inference

```
from sentence_transformers import util
import torch

embeddings = model.encode(abstracts, convert_to_tensor=True)
query_embedding = model.encode(query_text, convert_to_tensor=True)

# Calculate the similarity between the query and the abstracts
cosine_scores = util.pytorch_cos_sim(query_embedding, embeddings)
retrieval_results = torch.topk(cosine_scores, k=15)

retrieval_results.values -> lista de pares (score, index)
```



S/

Identificação: género feminino, 24 anos.

AP:

# sem antecedentes conhecidos.

MH:

# contraceptivo oral

HDA: Recorre ao serviço de urgência por **nódulos violáceos dolorosos na região pré-tibial bilateral**, que estenderam a toda a perna e coxa, com 1 mês de evolução. Três semanas antes do início do quadro, realizou a 2ª dose da vacina contra **SARS-CoV-2** Comirnaty® (Pfizer-BioNTech), a mesma que realizou um mês antes. Sem outras queixas, nomeadamente sugestivas de patologia auto-imune ou síndrome onstitucional.

O/

Exame físico- apirética, sem alterações da orofaringe ou adenopatias, observando-se lesões nodulares em ambas as coxas e pernas, de cor violácea, eritematosas, com 2 cm de maior eixo, dolorosas à palpação.

A/

# Analiticamente- **leucocitose de 12,6 x 10<sup>9</sup>/L com neutrofilia, proteína C reactiva 210 mg/L e velocidade de sedimentação 42 mm/h.**

# Anticorpo anti-estreptolisina O ne-gativo, serologias para sífilis e vírus Epstein-Barr, hepatite B e C e VIH negativas. Pannel imunológico, com pesquisa de anticorpos anti-nucleares (ANA), anticorpos anti-double-stranded DNA (dsDNA), anticorpos itoplasmáticos anti-neutrófilos (ANCA), complemento, factor reumatóide e anticorpos anti-péptido citrulinado cíclico, negativos.

# Marcadores de lesão hepática e função renal sem alterações.

# Hemoculturas negativas.

# TC TAP sem alterações.

Assim, assumido **eritema nodoso secundário à administração da vacina contra SARS-CoV-2.**

Iniciada **prednisolona 60 mg/dia.**

Evolução:

# Uma semana depois houve melhoria clínica significativa e diminuição acentuada dos parâmetros inflamatórios, pelo que se iniciou redução gradual de corticoterapia, com resolução completa do quadro ao fim de duas semanas.

# Passados seis meses, a doente teve doença ligeira por SARS-CoV-2 e uma semana depois surgiram lesões com características idênticas às do episódio anterior. Iniciou novo ciclo de corticoterapia, com resolução do quadro. Pela exuberância das lesões, a doente foi aconselhada a não repetir imunização com vacina de mRNA contra SARS-CoV-2.

# Até à data, não voltou ter infecção por SARS-CoV-2 nem teve ressurgimento das lesões cutâneas.

S/

Identificação: Género feminino, 31 anos

AP:

# Puérpera IGIP

## parto eutócico às 40 semanas e 4 dias de gestação

## História de 3 episódios de bacteriúria assintomática, com uroculturas positivas para *Escherichia coli* durante a gravidez tratadas com fosfomicina.

## fez suplementação oral com ácido fólico e ferro durante a gravidez. Restantes antecedentes pessoais sem relevo no presente contexto.

HDA: Ao 3º dia de pós-parto foi realizada visita de Ginecologia para decisão de alta clínica. À avaliação clínica por Ginecologia e Obstícia- "hemodinamicamente estável e apirética, **pálida com escleróticas ligeiramente ictéricas**. Lóquios normais, mamas tensas e períneo com cicatriz de episiorrafia sem sinais inflamatórios. Foram realizadas análises de controlo que revelaram **anemia normocítica normocrômica e trombocitopénia grave**"

Neste contexto, pedida avaliação urgente por Medicina Interna.

À minha avaliação, refere cansaço e cefaleia holocraniana que atribui ao pós-parto e à privação de sono. Nega hematúria, rectorragia, gengivorragia ou outras perdas hemorrágicas.

O/

Exame físico sem hematomas, petéquias ou equimoses. Sem outras manifestações neurológicas.

A/

# Analiticamente com sinais evidentes de **anemia hemolítica com elevação da LDH, haptoglobina indoseável e elevação da bilirrubina à custa da indireta, com Coombs direto e indireto negativos**. Sem consumo de complemento.

# esfregaço de sangue periférico com **anisocitose, policromatofilia, 1,2% de esquizócitos, 1 eritroblasto por cada 100 leucocitos, raras plaquetas grandes**.

# ecografia abdominal sem esplenomegália ou outras alterações.

# serologias virais foram negativas, bem como a pesquisa de marcadores de autoimunidade. # **anticorpo anti-ADAMTS13 positivo fraco com um valor de 17 UI/mL (negativo < 13 UI/mL) e a atividade ADAMTS13 foi de 6% (positivo < 10%)**.

Assim, assumida púrpura trombocitopénica trombótica

Iniciada plasmaférese cerca de 6 horas após o início da suspeita clínica e completou 5 sessões, em dias consecutivos. Realizou 3 dias de metilprednisolona 1g/dia.

Com melhoria analítica progressiva e estabilização dos valores de hemograma e resolução dos sintomas neurológicos.

Até à atualidade não apresentou recidiva.



# Information Retrieval

## RepositoriUM

Luís Filipe Cunha  
lfc@di.uminho.pt

José João Almeida  
jj@di.uminho.pt

