
Sentence-BERT

Information Retrieval - Qatent

Pablo Mollá – 2024/2025

Agenda

- Paper Research
 - Motivation + Challenges of BERT
 - SBERT vs BERT Architecture
-

Sentence-BERT: Sentence Embeddings using Siamese BERT- Networks

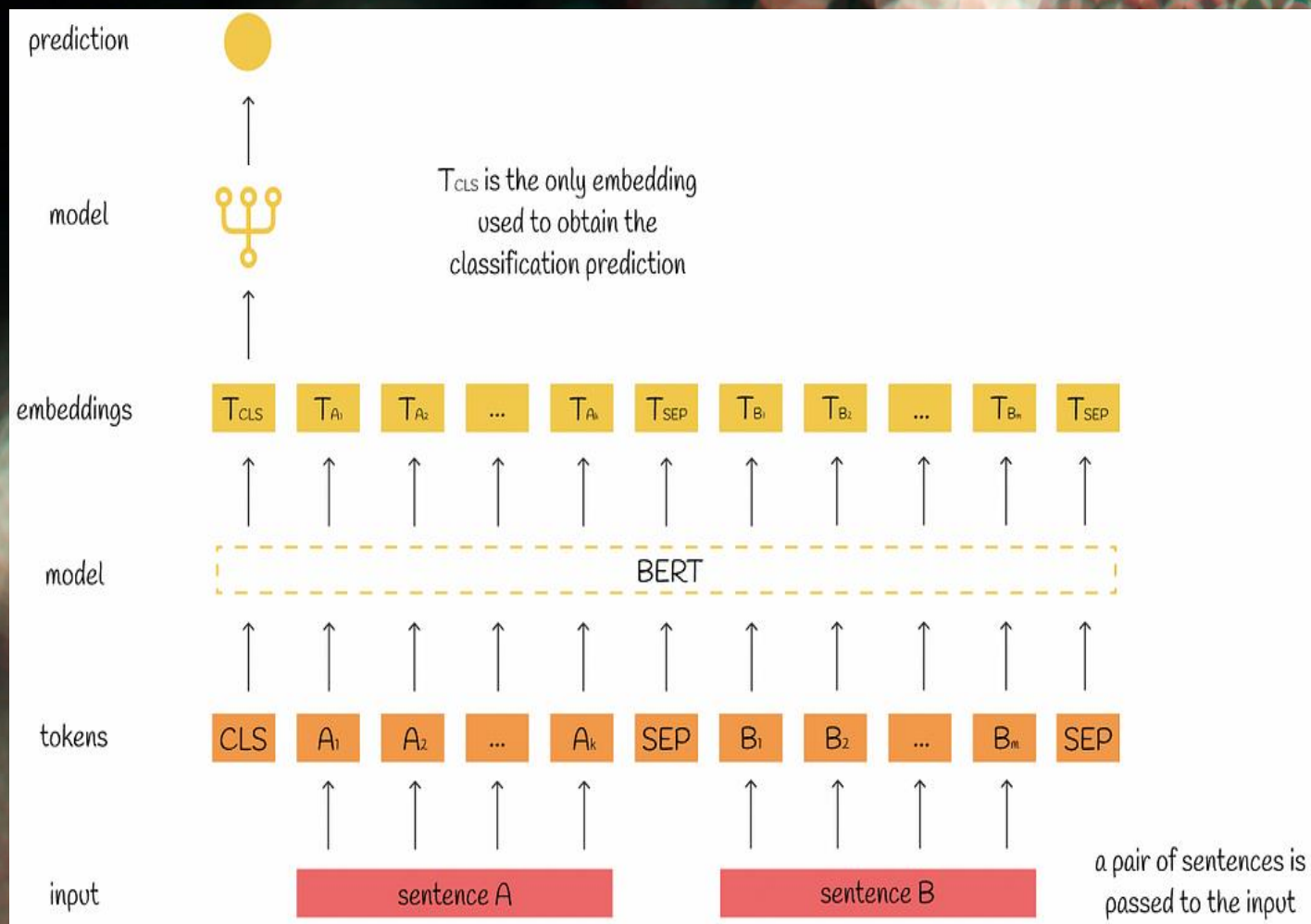
- Date: 27 Aug 2019
 - Authors: Nils Reimers and Iryna Gurevych – Technische Universität Darmstadt
 - Content: Fine-tuned version of the pre-trained BERT transform model designed to improve efficiency + accuracy of sentence embeddings
-

-
- Computational inefficiency of BERT for sentence-pair regression (e.g Semantic Textual Similarity) – Not reusable

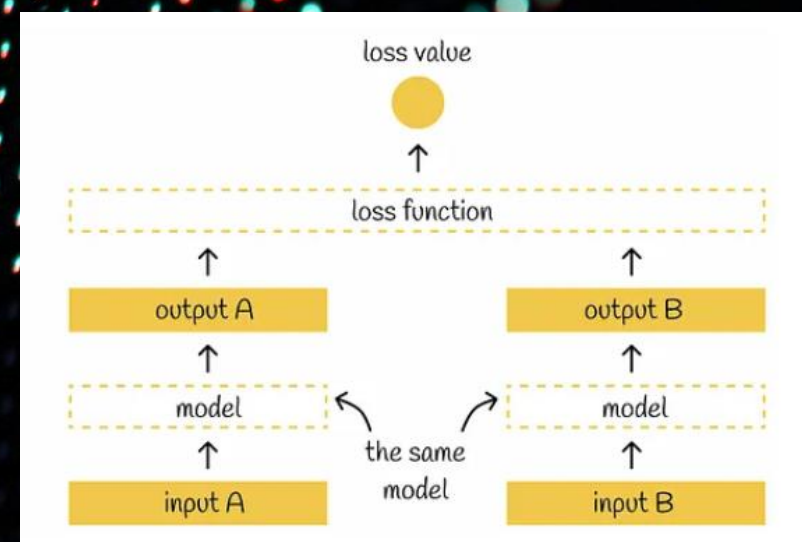
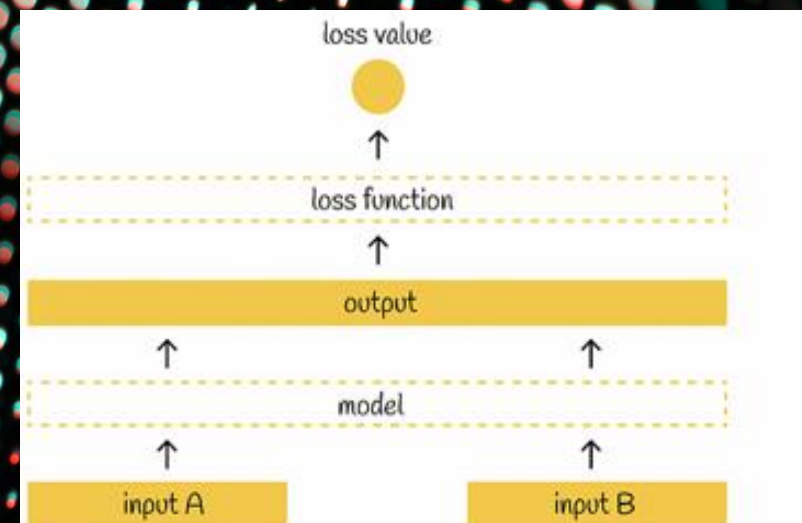
- Contextual Embeddings Issue – Not comparable

Motivation

BERT's Architecture

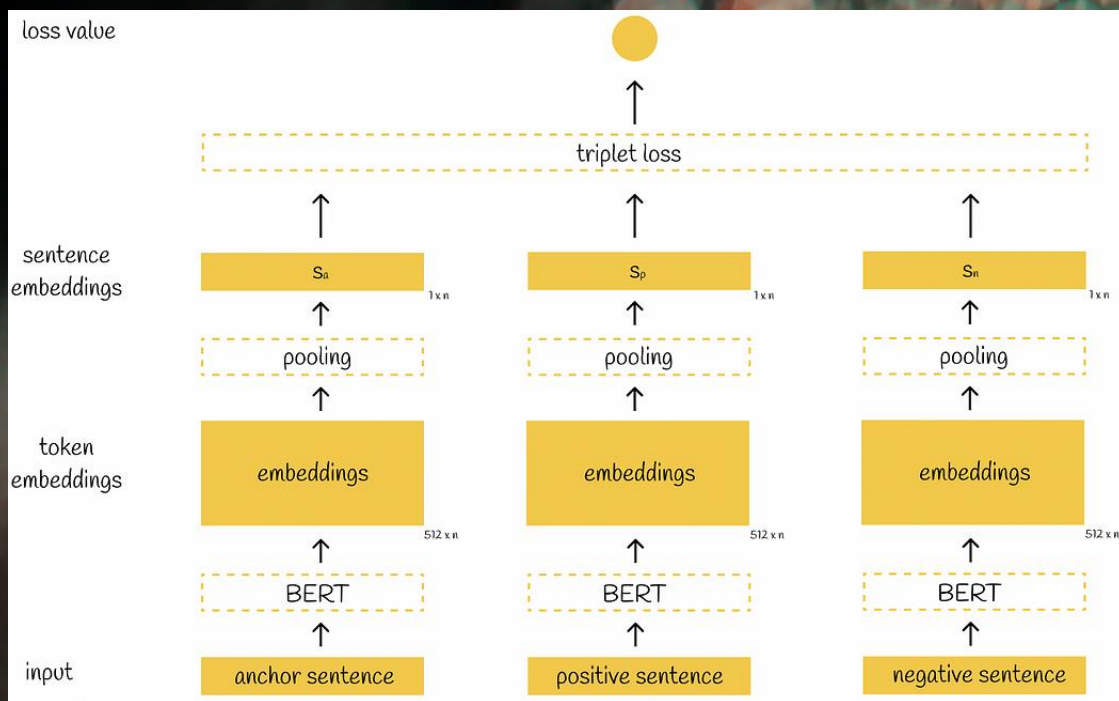


Cross-Encoder

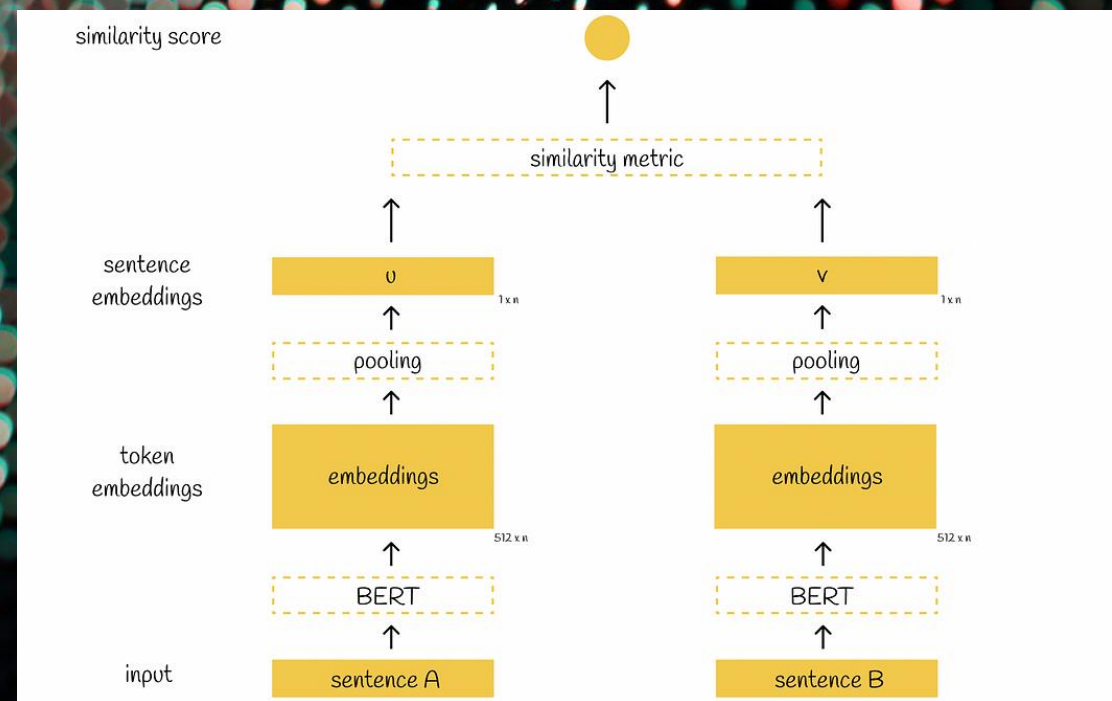


Siamese/Bi-Encoder

SBERT Architecture



Triple Network



Siamese Network

SBERT's Sentence Embedding

- Whole sentence representation (pooling strategy)
- Contextual Richness (core content words & function words, syntactic information)
- Semantic Density (pooling compresses into dense vector + reduced noises)
- Efficiency in Similarity Computation (independent sentence embeddings)
- Flexibility (types of pooling)

BERT's CLS

- Individual tokens + semantic meaning
- Relative + absolute positions in the pair sequence
- Contextual relationships between all tokens within each sentence + across sentence pair
- Aggregated info from deeper layers (syntactic, grammatical structure)

Questions & Possible Answers

