Lab Class NLP:VIII

Exercise 1 : Word2Vec

Word2Vec is widely used to produce word representations, but it has some shortcomings.

(a) Word2Vec suffers from out-of-vocabulary problems. Explain a technique that you might use to mitigate them.

(b) Suppose that you are tasked with building a job recommendation system. You could use pre-trained word embeddings to match applicants to job postings based on a personal profile. What ethical problems might arise if you did this?

(c) Explain a technique that you might use to mitigate the problems identified in (b).

Exercise 2 : Skip-Gram with Negative Sampling

The skip-gram model learns embeddings for individual words. Consider the function $y(\mathbf{t}, \mathbf{u})$ below [NLP:III-113].

$$p(c = 1 \mid t, u; W_V, W_U) \coloneqq y(\mathbf{t}, \mathbf{u}) = \sigma\left(\left(\underbrace{W_U^\top \mathbf{u}}_{①}\right)^\top \left(\underbrace{W_V^\top \mathbf{t}}_{①}\right)\right) = \underbrace{\sigma}_{③}\left(\underbrace{\mathbf{w}_u^\top \mathbf{w}_t}_{②}\right)$$

(a) Write down the parameters of the skip-gram model.

(b) Explain what each numbered component does.

Exercise 3 : BERT Pre-Training and Fine-Tuning

(a) Consider the following masked input sequence used during BERT pre-training:

```
[CLS] The first [MASK] became the [MASK]-watched [MASK] in
Northern [MASK] since modern records.  [SEP] The series was
[MASK] shortly after the pilot [MASK] aired.  [PAD] [PAD] ...
```

Explain the purpose of each of the following special tokens and describe how it is used during BERT pre-training:

(i) [CLS]:

(ii) [MASK]:

(iii) [SEP]:

(iv) [PAD]:

(b) BERT training consists of two phases: pre-training and fine-tuning. Explain the purpose of each phase and describe the main differences between them.

(c) Figure 1 shows the architecture of BERT.

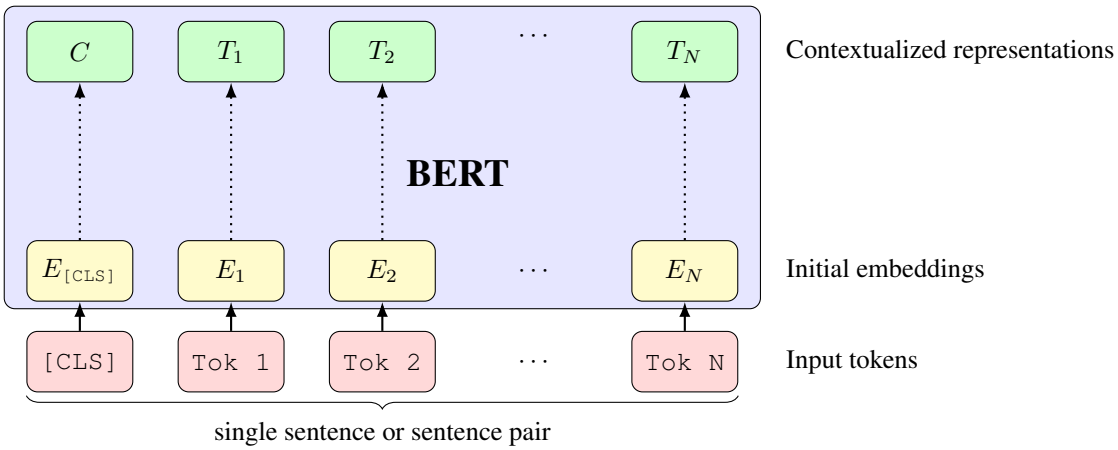

Figure 1: BERT architecture.

Explain the role of each of the following components:

(i) Input tokens (Tok1, Tok2, ...):

(ii) Initial embeddings ($E_1, E_2, \ldots$)

(iii) The contextualized representations ($T_1, T_2, \ldots$):

(iv) The representation C: