1 Problem Definition

The primary goal of document-level dense passage retrieval is to retrieve the most relevant passage p^* from a given document D_i for a specific query q. Formally, we aim to find:

$$p^* = \arg\max_{p \in P_i} \mathrm{sim}(q, p)$$

where P_i is the set of all passages in D_i, and $\mathrm{sim}(q, p)$ is the similarity function defined between query q and passage p. The similarity function $\mathrm{sim}(q, p)$ between a query q and a passage p is defined as:

$$\mathrm{sim}(q, p) = E_Q(q)^\top E_P(p)$$

where the query and passage encoders are denoted as E_Q and E_P respectively, where $E_Q : Q \to \mathbb{R}^d$ and $E_P : P \to \mathbb{R}^d$.

2 Training Objective

Let $D = \{D_1, D_2, \ldots, D_n\}$ be the set of training documents, where each D_i consists of passages P_i and associated queries Q_i. During training, we optimize over one document in a single batch. Given a document D, we extract a batch of n unique query-passage pairs $(q_1, p_1), \ldots, (q_n, p_n)$.

The embeddings for the queries and the passages are defined as :

$$Q_e = [E_Q(q_1), E_Q(q_2), \ldots, E_Q(q_n)]$$
$$P_e = [E_P(p_1), E_P(p_2), \ldots, E_P(p_n)]$$

We calculate the logits as the scaled dot-product of the embeddings:

$$\mathrm{logits} = Q_e \cdot P_e^\top \times \exp(t)$$

where t is a learned temperature parameter.

The cross-entropy loss for queries loss_q and passages loss_p are computed as follows:

$$\mathrm{loss}_q = \mathrm{CrossEntropy}(\mathrm{logits}, \mathrm{labels})$$

$$\mathrm{loss}_p = \mathrm{CrossEntropy}(\mathrm{logits}^\top, \mathrm{labels})$$

where $labels = [1, 2, \ldots, n]$.

Finally, the symmetric contrastive loss $\mathcal{L}$ is defined as the average of loss_q and loss_p:

$$\mathcal{L} = \frac{\mathrm{loss}_q + \mathrm{loss}_p}{2}$$