Mitigating the Impact of False Negatives in Dense Retrieval with Contrastive Confidence Regularization

Shiqi Wang, Yeqin Zhang, Cam-Tu Nguyen*

National Key Laboratory for Novel Software Technology, Nanjing University, China
School of Artificial Intelligence, Nanjing University, China
{wangsky,zhangyeqin}@smail.nju.edu.cn, {ncamtu}@nju.edu.cn

Abstract

In open-domain Question Answering (QA), dense retrieval is crucial for finding relevant passages for answer generation. Typically, contrastive learning is used to train a retrieval model that maps passages and queries to the same semantic space. The objective is to make similar ones closer and dissimilar ones further apart. However, training such a system is challenging due to the false negative issue, where relevant passages may be missed during data annotation. Hard negative sampling, which is commonly used to improve contrastive learning, can introduce more noise in training. This is because hard negatives are those closer to a given query, and thus more likely to be false negatives. To address this issue, we propose a novel contrastive confidence regularizer for Noise Contrastive Estimation (NCE) loss, a commonly used loss for dense retrieval. Our analysis shows that the regularizer helps dense retrieval models be more robust against false negatives with a theoretical guarantee. Additionally, we propose a model-agnostic method to filter out noisy negative passages in the dataset, improving any downstream dense retrieval models. Through experiments on three datasets, we demonstrate that our method achieves better retrieval performance in comparison to existing state-of-the-art dense retrieval systems.

Introduction

Text retrieval involves searching for relevant information in vast text collections based on user queries. Efficient and effective methods for this task have revolutionized how we interact with information systems. Recently, there has been growing interest in augmenting large language models (LLMs) with text retrieval for question answering (QA) (Lewis et al. 2020a; Guu et al. 2020; Glass et al. 2022; Borgeaud et al. 2022; Fu et al. 2022; Zhang et al. 2023). These approaches harness retrieval models to obtain external knowledge and ground LLM outputs, reducing hallucinations and the need for frequent LLM updates. Interestingly, augmenting an LM with a retrieval helps reduce the number of parameters required to achieve similar performance as larger LMs (Mialon et al. 2023).

Text retrieval methods can be broadly categorized into two main approaches: sparse and dense retrievals. Sparse methods, such as BM25, exploit the frequency of words to measure the relevance between a passage and a query. While efficient, these methods often fall short of capturing intricate relationships and contextual nuances of language. In contrast, dense retrieval methods aim to learn meaningful representations from the semantic content of passages and queries effectively. These models can be trained based on a pretraining model (e.g. BERT, RoBERTa) as well as fine-tuned for downstream QA tasks (Lewis et al. 2020b), offering easy integration. In addition, it is possible to apply approximate nearest neighbors (ANN) with dense retrieval (Xiong et al. 2021) for efficient retrieval.

This paper focuses on dense retrieval, where contrastive learning is often employed to train passage and query encoders. The core principle of contrastive learning is to encode passages and queries such that relevant passages are closer to their corresponding query in the embedding space, while irrelevant passages are farther away. To train such encoders, we need a labeled dataset with queries annotated with relevant passages (positive samples). However, due to the vast number of candidate passages and the complexity of questions, it is common for annotators to miss relevant information (texts) during data preparation, leading to unlabeled positive examples (false negatives) in the training set. Recent studies support this assumption. For instance, Ni, Gardner, and Dasigi (2021) found that over half of 50 answerable questions from the IIRC dataset (Ferguson et al. 2020) had at least one missing piece of evidence. Similarly, Qu et al. (2021) manually reviewed top-retrieved passages not labeled as positives in MSMARCO (Nguyen et al. 2016) and detected a 70% false negative rate. On the other hand, it is essential to sample hard negatives for effective contrastive learning. Here, hard negatives refer to passages obtained from the top results of a pre-trained dense retrieval model or BM25. Unfortunately, hard negative sampling is susceptible to higher false negative rates in noisy datasets because such negative samples are more likely to be mislabeled ones. Therefore, mitigating the impact of false negatives can potentially improve the performance of dense retrieval.

Several strategies have recently emerged to address the problem of false negatives. Qu et al. (2021) use a highly effective but inefficient reranker based on a cross-encoder to identify high-confidence negatives as true negatives, which were then used to train the retrieval model. Ni, Gardner, and

*Corresponding authors.

Dasigi (2021) leverage answers in the downstream QA task and design several heuristics to detect valid contexts such as lexical overlapping (between a gold answer and a candidate passage). Recently, Zhou et al. (2022) suggests selecting samples that are highly similar to positive samples but not too close to the query. These samples are considered informative negatives and unlikely to be false negatives. Despite the recent progress, these current methods are primarily based on heuristics and lack a theoretical guarantee.

This paper formalizes the problem of training dense retrieval with false negatives into the peer loss framework (Liu and Guo 2020; Cheng et al. 2021), a theoretical sound approach to learning with label noise. We extend this framework by developing a confidence regularizer for NCE, a commonly used loss for training dense retrieval models. Our regularized loss function increases the model's confidence and proves to be robust against false negatives. By encouraging confident scoring, we prevent the model from overfitting to noise, resulting in a more robust retrieval model. We then propose a new passage sieve algorithm, that makes use of a confidence regularized retrieval model to select true hard negatives. The clean dataset after the passage sieve is then used to train a stronger retrieval model. We prove that our method can successfully filter out false negatives from hard negatives under mild assumptions. Through experiments on three datasets, we demonstrate that our method achieves better retrieval performance compared to existing state-of-the-art dense retrieval systems that rely on heuristic false negative filtering. Supplementary materials (the appendix, codes) can be found in our GitHub[1].

Related Works

Dense Retrieval

The dual-encoder (two-towers or biencoder) architecture (Huang et al. 2013; Reimers and Gurevych 2019) is the common choice for dense retrieval thanks to its high efficiency. However, vanilla dual encoders have several challenges such as the limited expressiveness compared to cross-encoders, and the suboptimal performance due to non-informative negative samples. As a result, various solutions have been introduced to improve vanilla dual encoders from different perspectives such as knowledge distillation (Ren et al. 2021b; Lu et al. 2022), lightweight interaction models (Khattab and Zaharia 2020; Humeau et al. 2019), sophisticated training procedure (Zhang et al. 2021; Qu et al. 2021; Ren et al. 2021b), negative sampling strategies (Karpukhin et al. 2020; Xiong et al. 2021; Qu et al. 2021; Zhou et al. 2022). In this paper, we focus mostly on negative sampling strategies, particularly targeting the false negative issue. Unlike the closely related works (Qu et al. 2021; Zhou et al. 2022; Ni, Gardner, and Dasigi 2021), which exploit heuristic strategies, our method leverages the peer-loss approach (Liu and Guo 2020), effectively combining practical application with a theoretically-informed perspective.

[1]https://github.com/wangskyGit/passage-sieve

Label-Noise Robust Machine Learning

Developing machine learning models that are robust against label noise is important for supervised learning. Existing methods tackle label noise based on the type of noise, such as random noise (Natarajan et al. 2013; Manwani and Sastry 2013), class-dependent noise (Liu and Tao 2015; Patrini et al. 2017; Yao et al. 2020), or instance-dependent noise (Zhu, Liu, and Liu 2021; Cheng et al. 2021; Xia et al. 2020; Yang et al. 2022; Hao et al. 2022). Unfortunately, these methods are mainly designed for multi-class classification, and thus cannot be directly applied to our task.

Also relevant to our work are (Chuang et al. 2020; Robinson et al. 2020) which consider the issue of bias in contrastive learning for multi-class classification. These studies rely on the assumption of a fixed uniform noise probability across different classes (i.e. queries in our case) to approximate the positive and negative distributions. In contrast, our work concerns noise in supervised contrastive learning for query-dependent ranking, where the distributions of positives and negatives are not the same for different queries. For example, it is more likely for queries with high recall to be associated with false negatives. In other words, the noise probability is not uniform across queries.

Preliminaries

Problem Formalization

Let C be a collection of textual passages, and $\tilde{\mathcal{D}} = \{(q_n, p_n^+, \tilde{\mathcal{N}}_{q_n})\}$ indicates the noisy set of tuples, each consists of a query q_n, an annotated positive $p_n^+ \in C$, and a set of sampled negatives $\mathcal{N}_{q_n} \subset C$ with possible false negatives. Our objective is to mitigate the impacts of such false negatives and learn a robust retrieval model that can closely match the model trained on the clean dataset $\mathcal{D} = \{(q_n, p_n^+, \mathcal{N}_{q_n})\}$ without false negatives.

Dual-Encoders for Dense Retrieval

A question or a passage is encoded as dense vectors separately, and the similarity is determined as the dot product of the encoder outputs.

$$sim(q, p) =< E_{qry}(q), E_{psg}(p) > \quad (1)$$

where E_{qry}, E_{psg} represent distinct encoders that map the query and the passage into dense vectors, and $< . >$ is the similarity function such as dot product, cosine or Euclidean distance. The encoders are often built based on a pre-trained language model such as BERT-based (Karpukhin et al. 2020; Luan et al. 2021; Xiong et al. 2021; Oguz et al. 2022; Zhang et al. 2022), ERNIE-based (Qu et al. 2021; Ren et al. 2021b; Zhang et al. 2021) or RoBERTa-based (Oguz et al. 2022) etc. Since passages and queries are encoded separately, passage embeddings can be precomputed and indexed using Faiss (Johnson, Douze, and Jégou 2021) for efficient search.

Contrastive learning is commonly used to train dual encoders (Karpukhin et al. 2020; Xiong et al. 2021; Qu et al. 2021; Zhou et al. 2022). Typically, it is assumed that we have access to the clean training set $\mathcal{D} = \{(q_n, p_n^+, \mathcal{N}_{q_n})\}$. Using

this training set, we measure the NCE (Noise-Contrastive Estimation) loss for a given query q_n as follows:

$$\ell_{NCE}(p_n^+, q_n) = -\ln(f(p_n^+, q_n)) = -\ln(\frac{e^{sim(p_n^+, q_n)}}{\sum_{p_i^- \in \mathcal{N}_{q_n}} e^{sim(p_i^-, q_n)} + e^{sim(p_n^+, q_n)}}) \quad (2)$$

The total loss function can be calculated as follows:

$$\frac{1}{N} \sum_{n \in [N]} \ell_{NCE}(p_n^+, q_n) \quad (3)$$

where N is the total number of queries in the dataset, and $[N] = \{1, 2, \ldots, N\}$.

Negative sampling aims to select samples for $\mathcal{N}_{q_n}$ and plays an important role in learning effective representations with contrastive learning. In the context of dense retrieval, two common strategies for negative sampling are in-batch negatives and hard negatives (Karpukhin et al. 2020; Xiong et al. 2021; Qu et al. 2021). In-batch negatives involve selecting positive passages from other queries in the same batch as negative samples. Generally, increasing the number of in-batch negatives improves dense retrieval performance. On the other hand, hard negatives are usually informative samples that receive a high similarity score from another retrieval model (e.g., BM25, a pre-trained DPR) (Karpukhin et al. 2020). Hard negatives can result in more effective training for DPR, yet including such samples may exaggerate the issue of false negatives.

Peer Loss and Confidence Regularization

The problem of learning with label noise has been extensively researched in the context of classification tasks (Liu and Guo 2020; Xia et al. 2020; Zhu, Liu, and Liu 2021; Cheng et al. 2021; Yang et al. 2022; Hao et al. 2022). Recently, two effective methods called Peer Loss (Liu and Guo 2020) and its inspired Confidence Regularizer (Cheng et al. 2021; Zhu, Liu, and Liu 2021) have been introduced to train robust machine learning models. One advantage of these methods is that they can work without requiring knowledge of the noise transition matrix or the probability of labels being flipped between classes.

The concept of peer loss is initially introduced for the binary classification problem. In this setting, we use x to indicate an instance (a feature vector), $y \in \{-1, 1\}$ and $\tilde{y} \in \{-1, 1\}$ to represent the clean label and noisy labels, respectively. For each sample $(x_n, \tilde{y}_n)$, the peer loss is then defined for cross-entropy loss as follows:

$$\ell_{PL}(g(x_n), \tilde{y}_n) = \ell_{CE}(g(x_n), \tilde{y}_n) - \ell_{CE}(g(x_{n1}), \tilde{y}_{n2}) \quad (4)$$

where g indicates the classification function, x_{n1} and $\tilde{y}_{n2}$ are derived from different, randomly selected peer sample for n. Here, the first term measures the loss of the classifier prediction, whereas the second term penalizes the model when it excessively agrees with the incorrect or noisy labels (Liu and Guo 2020).

Inspired by peer loss, Cheng et al. (2021) develops CORES^2, which extends the framework for multi-class classification and uses the first-order statistic instead of randomly selecting peer samples. Specifically, the new loss in CORES^2 is defined as follows:

$$\ell(g(x_n), \tilde{y}_n) = \ell_{CE}(g(x_n), \tilde{y}_n) - \beta \mathbb{E}_{\tilde{D}_{\tilde{Y}}}[\ell_{CE}(g(x_n), \tilde{Y})] \quad (5)$$

where β is a hyper-parameter, $\tilde{Y}$ denotes the random variable corresponding to the noisy label and $\tilde{D}_{\tilde{Y}}$ indicates the marginal distribution of $\tilde{Y}$. It has been shown that learning with an appropriate β will make the loss function robust to instance-dependent label noise with theoretical guarantee (Cheng et al. 2021).

Contrastive Confidence Regularizer

We can adopt the above confidence regularization by introducing a binary label y associated with each pair of queries and passages, where $y = +1$ indicates the positive pair and $y = -1$ vice versa. Subsequently, CORES^2 can be used with pairwise cross-entropy to mitigate the impacts of false negatives on training the retrieval model. Unfortunately, the cross-entropy loss is not as effective as the NCE loss (Eq. 3) for dense retrieval since the latter can learn a good ranking function by contrasting a positive sample over a list of negatives (Karpukhin et al. 2020). As a result, we aim to extend the framework of peer loss and tailor the confidence regularization to the NCE loss.

To adopt the peer-loss framework, we extend NCE loss to measure loss values associated with negative pairs (p_i^-, q_n) besides positive ones. It is noteworthy that the original NCE loss (Eq. 3) incorporates all positive and negative pairs as normalization factors but only calculates loss values for positive pairs (p_n^+, q_n) while disregarding the negative pairs. Formally, we use p without superscript as a generic term for positive passage p_n^+ and negative passage $p_i^- \in \mathcal{N}_{q_n}$, and define a general NCE loss as follows:

$$\ell_{NCE}(p, q_n) = -\ln(f(p, q_n)) = -\ln(\frac{e^{sim(p, q_n)}}{\sum_{p_i^- \in \mathcal{N}_{q_n}} e^{sim(p_i^-, q_n)} + e^{sim(p_n^+, q_n)}}) \quad (6)$$

where p could be either positive or negative passages for query q_n. By this convention, the peer-loss framework can be applied by introducing randomly selected pairs (p_{n_1}, q_{n_1}), (p_{n_2}, q_{n_2}) $(n_1 \neq n_2)$ as peer samples to regularize the NCE loss. We then obtain the contrastive peer loss as follows:

$$\ell_{PL}(p_n^+, q_n) = \ell_{NCE}(p_n^+, q_n) - \ell_{NCE}(p_{n_1}, q_{n_2}) \quad (7)$$

While our regularized loss appears similar to the original peer-loss outlined in Eq. 4, the major difference lies in the structure of the loss functions. Specifically, the first term of Eq. 7 is calculated only for positive pairs whereas the first term of Eq. 4 involves both positive ($\tilde{y} = +1$) and negative ($\tilde{y} = -1$) samples.

Similar to CORES^2, we introduce P_{n_1} and Q_{n_2} as the random variables corresponding to the peer passage p_{n_1} and the peer query q_{n_2}, respectively. Let $\tilde{\mathcal{D}}_P$ be the distribution

of P_{n_1} given distribution $\tilde{\mathcal{D}}$. Note that Q_{n_2} is a uniform random variable, or $\mathbb{P}(Q_{n_2} = q_{n'}|\tilde{\mathcal{D}}) = 1/N$, the contrastive peer loss has the following form in expectation:

$$
\begin{aligned}
&\frac{1}{N}\sum_{n\in[N]}\mathbb{E}_{\tilde{\mathcal{D}}_{P_{n_1},Q_{n_2}}}\left[\ell_{NCE}(p_n^+,q_n)-\ell_{NCE}(P_{n_1},Q_{n_2})\right]\\
=&\frac{1}{N}\sum_{n\in[N]}[\ell_{NCE}(p_n^+,q_n)-\\
&\sum_{n'\in[N]}\mathbb{P}(Q_{n_2}=q_{n'}|\tilde{\mathcal{D}})\mathbb{E}_{\tilde{\mathcal{D}}_P}\left[\ell_{NCE}(P_{n_1},q_{n'})\right]\\
=&\frac{1}{N}\sum_{n\in[N]}\left[\ell_{NCE}(p_n^+,q_n)-\mathbb{E}_{\tilde{\mathcal{D}}_P}\left[\ell_{NCE}(P,q_n)\right]\right]\\
\approx&\frac{1}{N}\sum_{n\in[N]}\left[\ell_{NCE}(p_n^+,q_n)-\mathbb{E}_{\tilde{\mathcal{D}}_{P|q_n}}\left[\ell_{NCE}(P,q_n)\right]\right]
\end{aligned}
$$

The last approximation equation is obtained because when considering the batch training in dense retrieval, P_{n_1} is drawn from in-batch passages. In addition, all the passages in batch form the conditional passage distribution of query q_n. From this derivation, we can get the new noise robust contrastive loss function (i.e. ℓ_{RCL}) with contrastive confidence regularizer denoted by ℓ_{CCR} as follows:

$$
\begin{aligned}
\ell_{CCR}(p_n^+,q_n)&=\mathbb{E}_{\tilde{\mathcal{D}}_{P|q_n}}\left[\ell_{NCE}(p,q_n)\right]\\
\ell_{RCL}(p_n^+,q_n)&=\ell_{NCE}(p_n^+,q_n)-\beta*\ell_{CCR}(p_n^+,q_n)
\end{aligned}
\tag{8}
$$

In the following, we first empirically prove that our regularized loss function makes the retrieval model more confident, hence being more robust against false negatives. We then present the main theories that guarantee the robustness of our regularized NCE loss.

Analysis on Simulated Data

The RCL loss (eq. 8) is minimized by making the first term smaller and the expectation term bigger. In other words, we pull the loss associated with positive passages further from the average (the expectation) loss, subsequently making the model more confident in predicting positive passages. Intuitively, as label noise distorts the learning signal in clean data, the model trained on noisy datasets often fits noises, hence becoming less confident (Cheng et al. 2021). By making the model more confident, we can partially reverse it and obtain a more robust retrieval model. We verify this intuition by conducting a simulation on the Natural Question (NQ) dataset (Kwiatkowski et al. 2019). Specifically, we randomly convert some positive passages to false negatives and observe the effects of contrastive confidence regularizer ℓ_{CCR} on the distributions of the similarity scores in different groups of passages including false negatives, hard negatives and in-batch negatives. The experimental results in Figure 1 indicate that if we do not include the ℓ_{CCR} term (i.e., $\beta = 0$), the distributions are close to each other. On the other hand, incorporating the ℓ_{CCR} term makes it easier to separate these distributions. It should be noted that a small overlapping between the false negative and hard negative distributions is expected. This is because while all samples in the simulated "false negatives" category are certainly positive passages, there exist (unknown) false negatives in the "hard negatives" category.

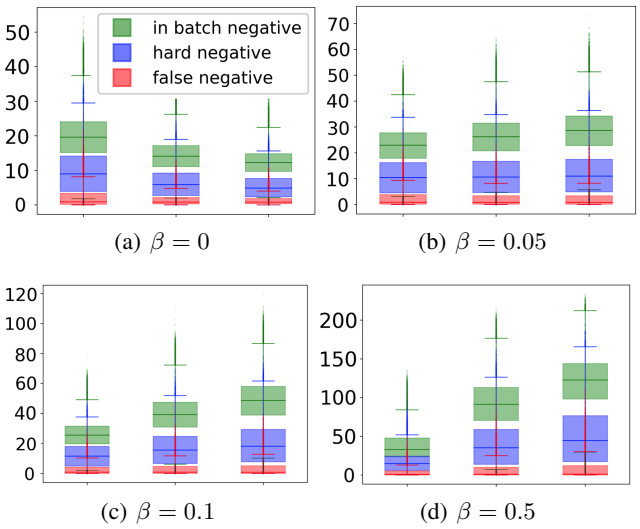

Figure 1: In the four figures, the horizontal axes, from left to right, represent the model trained through one to three epochs. The effect of the contrastive confidence regularizer. Four figures show analysis from four experiments with all the same settings but with different values of beta. All of them are continuously trained based on the same pre-trained vanilla DPR. The y-axis value represents the NCE loss value between queries and different types of passages. It shows that normal DPR training will cause different distributions to be squeezed together while ℓ_{CCR} has the potential to help distinguish false negatives from real negatives.

Theoretical Analysis

Theoretical analysis shows that our RCL loss enjoys similar properties with CORES^2. The detailed proof can be found in the Appendix, where the main idea is that we introduce pseudo-labels to bridge NCE loss and cross-entropy loss and follow a similar proof sketch with (Cheng et al. 2021). The main result is summarized in the following theorem.

Theorem 1. *With the assumption that the possibility of a noisy pair is smaller than a clean pair and a suitable selection of β, we have: minimizing $\mathbb{E}_{\tilde{\mathcal{D}}}[\ell_{NCE}(p^+,q) - \beta * \ell_{CCR}(p^+,q)]$ is equivalent to minimizing $\mathbb{E}_{\mathcal{D}}[\ell_{NCE}(p^+,q)]$*

Theorem 1 indicates that the contrastive confidence regularizer effectively addresses the adverse effect of false negatives on the loss function. Specifically, we can decompose the regularized loss so that the impact of false negatives is mostly captured by the product of ℓ_{CCR} and a scaling factor (the last term in Eq. (6), Appendix). Here, the scaling factor is intuitively related to the difference between the relative clean rate of a sample (compared to the average clean rate) and β. When β is large enough, this scaling factor is negative, thus reversing the (term) gradient associated with instances with labeled noise. Note that β should not be too large, otherwise it will affect learning from clean data. This

Algorithm 1: Passage Sieve Algorithm

Input: Noisy dataset $\tilde{\mathcal{D}}$ with false negatives
Parameter: Hyper-parameter β, learning rate α, training epochs T
Output: Sieved dataset $\mathcal{D}^*$

Get the pre-trained DPR model $\mathcal{M}$
Train $\mathcal{M}$ with robust contrastive loss for T epoch:
for $t \in [T]$ **do**
  for $(\mathbf{q}, \mathbf{p}^+, \mathcal{N}_\mathbf{q})$ **in** $\tilde{\mathcal{D}}$ **do**
    Calculate ℓ_{RCL} according to equation 8
    Compute gradients $\nabla \leftarrow \nabla_{\mathcal{M}} \ell_{RCL}$
    Update the model parameters: $\mathcal{M} \leftarrow \mathcal{M} - \alpha \cdot \nabla$
Evaluate all the hard negatives:
Initialize empty dataset $\mathcal{D}^*$
for $(\mathbf{q}, \mathbf{p}^+, \mathcal{N}_\mathbf{q})$ **in** $\tilde{\mathcal{D}}$ **do**
  $\mathcal{N}_q^* \leftarrow []$ {list of selected hard negatives}
  $th = \frac{1}{|D_{p|q}|} \sum_{p \in D_{p|q}} \ell_{NCE}(p, q; \mathcal{M})$ {Threshold}
  for p^- **in** $\mathcal{N}_q$ **do**
    $\ell_p^- \leftarrow \ell_{NCE}(p^-, q; \mathcal{M})$
    if $\ell_p^- \geq th$ **then**
      Append p^- to $\mathcal{N}_q^*$
  Append $(\mathbf{q}, \mathbf{p}^+, \mathcal{N}_\mathbf{q}^*)$ to $\mathcal{D}^*$
Return Sieved dataset $\mathcal{D}^*$ with confident false negatives

raises a crucial question regarding the existence of parameter β. Analyzing the existence of β becomes challenging when the nature of the noise is uncertain. Fortunately, in the context of false negatives in dense passage retrieval, we can derive the following theorem.

Theorem 2. *When the assumption in Theorem 1 holds, the similarity function sim is bounded (e.g. cosine similarity) and there is no false positive in the dataset $\tilde{\mathcal{D}}$, the β that satisfies Theorem 1 must exist and in the interval [0,1].*

Theorem 2 shows that under the setting of dense passage retrieval with false negatives, one can select a suitable β between the interval [0,1]. However, this theorem only applies when the similarity function is bounded and there is no other source of noise besides false negatives. In practical scenarios, many dense passage retrieval algorithms employ dot-product similarity instead of cosine similarity. When the loss function is not bounded, and there is an excessive number of negative passages, the confidence regularizer can cause the model to overly optimize the loss associated with negative cases. Our experiments indicate that in such situations, the contrastive confidence regularizer still helps as long as a sufficiently small value for β is selected. In general, as the batch size increases and in-batch negative sampling is utilized, we should decrease the value of β.

Passage Sieve Method

The previously proposed regularization is helpful towards dense retrieval models based on contrastive NCE loss, but not applicable to sophisticated methods such as AR2 (Zhang et al. 2021), which do not make use of NCE loss. As a result, we design a novel passage sieve algorithm, which can be used as pre-processing for the noisy datasets $\tilde{\mathcal{D}}$ to obtain relatively clean datasets for training.

Our method is presented in Algorithm 1. We first train DPR (Karpukhin et al. 2020), a dual-encoder based retrieval model on the noisy dataset. We then refine the retrieval model using the contrastive confidence regularizer. Afterward, we identify the hard negative passages with a loss function value $\ell_{NCE}(p, q_n)$ higher than the average loss value $\frac{1}{|N_{q_n}|+1} \sum_{p \in N_{q_n} \cup \{p_n^+\}} \ell_{NCE}(p, q_n)$, and set them as confident negatives. Finally, we discard all other passages except for the confident negatives to obtain a "clean" dataset. It is noteworthy that we employ cosine similarity to satisfy the assumptions outlined in Theorem 2.

Lemma 1. *Algorithm 1 ensures that a negative sample p^- in hard negatives will NOT be selected into the sieved dataset $\mathcal{D}^*$ if its score $f(p^-, q)$ given by the model f (eq. 2) is more than a random guess, i.e. its similarity score after softmax is bigger than the average value $1/(|\mathcal{N}_q| + 1)$.*

Lemma 1 provides the sufficient condition for accurately selecting true negatives in Algorithm 1. In this algorithm, the DPR model is trained with robust contrastive loss and cosine similarity. According to Theorems 1 and 2, by selecting a suitable value of $\beta \in [0, 1]$, minimizing $\mathbb{E}_{\tilde{D}}[\ell(p, q) + \ell_{CCR}(p, q)]$ is equivalent to minimizing $\mathbb{E}_{\mathcal{D}}[\ell(p, q)]$. This implies that false negatives are gradually given scores closer to the positive passage, and higher compared to hard negatives and in-batch negatives. Considering the abundance of true negatives, the scores assigned to false negatives will exceed random guesses at some point. Consequently, false negatives will be excluded from the sieved dataset in Algorithm 1, resulting in a clean dataset $\tilde{D}$.

Compared to vanilla DPR, DPR with CCR introduces additional computation associated with the expectation calculation, the second term in Eq. 8. Fortunately, this can be approximated by taking the mean value of the NCE losses of all passages given the query. In addition, the calculation of NCE losses is dominated by the calculation of the similarity matrix (Eq. 2), which is also needed in vanilla DPR. Consequently, the contrastive confidence regularizer only slightly increases the time complexity when implemented appropriately. It is worth noting that when using retrieval models like DPR, ANCE, or RocketQA that employ NCE loss, using the robust contrastive loss in section should be sufficient.

Experiment

Experimental Setup

Datasets We conduct experiments on three public QA datasets: Natural Question (NQ) (Kwiatkowski et al. 2019), Trivia QA (TQ) (Joshi et al. 2017) and MSMARCO Passage Ranking (MS-pas) (Nguyen et al. 2016). The detailed statistics are given in the supplementary material.

Evaluation Metrics Following previous works, we report R@k (k=5, 20, 100) for NQ and TQ, and MRR@10, R@k (k=50, 1K) for MS-pas. Here, MRR refers to the Mean Reciprocal Rank that calculates the reciprocal rank where the first relevant passage is achieved, and R@k measures the proportion of relevant passages (recall) in top-k results.

Method	NQ			TQ			MS-pas		
	R@5	R@20	R@100	R@5	R@20	R@100	MRR@10	R@50	R@1k
DPR (Karpukhin et al. 2020)	-	78.4	85.3	-	79.3	84.9	-	-	-
ANCE (Xiong et al. 2021)	71.8	81.9	87.5	-	80.3	85.3	33.0	81.1	95.9
COIL (Gao, Dai, and Callan 2021)	-	-	-	-	-	-	35.5	-	96.3
ME-BERT (Luan et al. 2021)	-	-	-	-	-	-	33.8	-	-
Individual top-k (Sachan et al. 2021)	75.0	84.0	89.2	76.8	83.1	87.0	-	-	-
RocketQA (Qu et al. 2021)	74.0	82.7	88.5	-	-	-	37.0	85.5	97.9
RDR (Yang and Seo 2020)	-	82.8	88.2	-	82.5	87.3	-	-	-
RocketQAv2 (Ren et al. 2021b)	75.1	83.7	89.0	-	-	-	38.8	86.2	98.1
PAIR (Ren et al. 2021a)	74.9	83.5	89.1	-	-	-	37.9	86.4	98.2
DPR-PAQ (Oguz et al. 2022)	74.2	84.0	89.2	-	-	-	31.1	-	-
Condenser (Gao and Callan 2021)	-	83.2	88.4	-	81.9	86.2	36.6	-	97.4
coCondenser (Gao and Callan 2022)	75.8	84.3	89.0	76.8	83.2	87.3	38.2	-	98.4
ERNIE-Search (Lu et al. 2022)	77.0	85.3	89.7	-	-	-	**40.1**	87.7	98.2
MVR (Zhang et al. 2022)	76.2	84.8	89.3	77.1	83.4	87.4	-	-	-
PROD (Lin et al. 2023)	75.6	84.7	89.6	-	-	-	39.3	87.0	98.4
COT-MAE (Wu et al. 2023)	75.5	84.3	89.3	-	-	-	39.4	87.0	**98.7**
AR2 (Zhang et al. 2021)	77.9	86.0	90.1	78.2	84.4	87.9	39.5	87.8	98.6
AR2+passage sieve	**79.2**	**86.4**	**90.7**	**78.7**	**84.7**	**88.2**	39.8	**88.3**	98.6

Table 1: Performance of passage sieve with the robust contrastive loss on NQ, TQ test sets, and MS-pas development set. The results of the baselines are from the original papers. All results with AR2 as the backbone are obtained from training AR2 with the same batch size and the same number of hard negatives as in (Zhang et al. 2021).

	Method	B	R@5	R@20	R@100
NQ	AR2	64	77.9	86.0	90.1
	AR2+SimANS	64	78.6	86.2	90.3
	AR2+passage sieve	32	78.6	86.1	90.5
	AR2+passage sieve	64	**79.2**	**86.4**	**90.6**
TQ	AR2	64	78.2	84.4	87.9
	AR2+SimANS	64	78.6	84.6	88.1
	AR2+passage sieve	32	78.6	**84.8**	**88.3**
	AR2+passage sieve	64	**78.7**	84.7	88.2

Table 2: Comparison of AR2, AR2+SimANS and AR2+passage sieve, where **B** stands for batch size. Here, the number of hard negatives is set to 15 which is the same as in Section .

Method	# hn	NQ		
		R@5	R@20	R@100
AR2	1	76.4	85.3	89.7
AR2+passage sieve	1	**77.6**	**86.1**	**90.3**
AR2	5	76.9	85.3	89.7
AR2+passage sieve	5	**78.0**	**85.9**	**90.6**
AR2	15	77.9	86.0	90.1
AR2+passage sieve	15	**78.6**	**86.1**	**90.5**

Table 3: Performance of AR2 + passage sieve with the robust contrastive loss on NQ test sets. Here, # hn stands for the number of hard negatives during training. We set the batch size to 32 for all compared methods. The results of the baselines are from (Zhang et al. 2021).

Effects of Passage Sieve Method

Experimental Design We implement our passage sieve method on top of AR2 (Zhang et al. 2021), which is referred to as AR2+passage sieve. To evaluate the effectiveness of our passage sieve method, we compare it with the original AR2 along with other contemporary baselines. The details on the baselines are given in the supplementary material.

During the passage sieve procedure in AR2+passage sieve, we set $\beta = 0.5$ and epochs $T = 1$, learning rate $\alpha = 1e - 7$. Additionally, we leverage the cosine similarity function and set the initial number of hard negative passages to be two times the number of hard negatives to be used in the downstream AR2 model. As for AR2 training, all settings are set the same way as in (Zhang et al. 2021). Specifically, the batch size is set to 64 and the number of hard negatives is 15.

Overall Results Table 1 presents the results of AR2+passage sieve on NQ and TQ test sets, as well as MS-pas development set. It can be observed that AR2 outperforms most of the baseline models on all three datasets across different evaluation metrics. Furthermore, the proposed passage sieve approach enhances the performance of AR2 on all three datasets and evaluation metrics. Notably, our approach utilizes only half of the batch size compared to AR2 on NQ and TQ datasets. The findings suggest that the sieve algorithm significantly contributes to improving the quality of hard negative samples in the dataset, enabling better results even with less GPU memory requirement. It is worth mentioning that our passage sieve procedure only accounts for one-tenth of the training time compared to AR2, the downstream retrieval model.

SimANS (Zhou et al. 2022) is the recently proposed method to address the issue of false negatives based on some

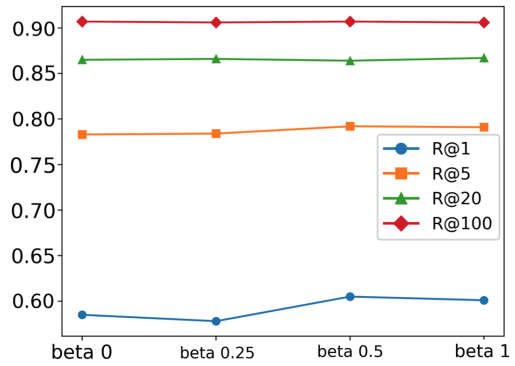

Figure 2: Performance of AR2+passage sieve with different hyper-parameter β on NQ test sets. Overall, the passage sieve algorithm is not sensitive to β.

useful heuristic assumptions. When combined with AR2, AR2+SimANS achieves state-of-the-art results in dense retrieval. In contrast, we start from theory in noise-robust machine learning algorithms and develop a new loss function. The results in Table 2 show that both SimANS and our method enhance the performance of AR2. However, compared to AR2+SimANS, the passage sieve achieves better results in all terms of TQ and NQ datasets. The advantage is more clearly seen on NQ dataset. It should be noted that we do not include the result on the MS-pas dataset of SimANS because they leverage a 4 times larger batch size than AR2 which makes the comparison unfair.

Detailed Analysis

Influence of limited batch size We investigate the impact of smaller batch size on the performance of AR2+passage sieve. Table 2 shows that AR2+passage sieve achieves better results in terms of R@5, R@20, and R@100 on TQ dataset, even with a smaller batch size. On NQ dataset, our method achieves comparable results with AR2+SimANS while having only a slight decrease in R@20.

Influence of the number of hard negatives To evaluate the impact of hard negatives, we experiment with varying numbers of hard negatives during training. To speed up the experiment, the batch size is set to 32 and we only work on NQ dataset. The results in Table 3 demonstrate that the passage sieve approach consistently enhances the performance of AR2 across all settings. Particularly, with only 5 hard negatives, our method can reach a performance near that of 15 hard negatives. Intuitively, when the quality of hard negative samples is higher, we do not need many negative samples, thus reducing computing resources for training.

Influence of hyper-parameter β To study the influence of the hyper-parameter β on the passage sieve algorithm, we conduct experiments on NQ dataset with different β and all other settings remain the same. Specifically, we vary the values of β to be 0, 0.25, 0.5. Figure 2 indicates the passage sieve method is not sensitive to β. This is because DPR in the passage sieve method is trained with cosine similarity, which satisfies our assumptions in Theorem 2.

	Method	Batch size	R@5	R@20	R@100
NQ	DPR	128	-	78.4	85.4
	DPR+CCR	64	65.9	77.6	85.4
	DPR+CCR	128	**68.4**	**79.5**	**86.1**
TQ	DPR	128	-	79.3	84.9
	DPR+CCR	64	70.5	78.7	84.9
	DPR+CCR	128	**71.5**	**79.8**	**85.1**

Table 4: Performance of DPR+Contrastive Confidence Regularizer on NQ test sets. The results of the baselines are from original papers.

Effects of Contrastive Confidence Regularizer

Exterimental Design We evaluate the performance of the Contrastive Confidence Regularizer on the DPR model (Karpukhin et al. 2020). We incorporate the regularizer into the existing NCE loss function during the final 5 epochs of training. Other parameters and configurations are kept consistent with those in the original DPR paper. Since the DPR model utilizes a dot-product similarity function, we select the values of β from the range of 0.001 to 0.0001. The selection of β is based on the performance observed on the validation set.

Experiment Results Table 4 shows results on NQ and TQ datasets. It is observable that the proposed contrastive confidence regularizer helps DPR get a better performance across all metrics on NQ and TQ datasets. The experimental results also show that the reduction of batch size has a greater impact on R@20 and R@5, but a smaller impact on R@100. On both datasets, the proposed CCR helps achieve the same R@100 value as DPR with only half the batch size.

Conclusion

This paper aims to mitigate the impact of false negatives on dense passage retrieval. Toward such a goal, we extend the peer-loss framework and develop a confidence regularization for training robust retrieval models. The proposed regularization is compatible with any base retrieval model that uses NCE loss, a widely used contrastive loss function in dense retrieval. Through empirical and theoretical analysis, it is demonstrated that contrastive confidence regularization leads to more robust retrieval models. Building on this regularization, a passage sieve algorithm is proposed. The algorithm leverages a dense retrieval model trained with confidence regularized NCE loss to filter out false negatives, thereby improving any downstream retrieval model including those that do not exploit NCE loss. The effectiveness of both the passage sieve algorithm and the confidence regularization method is validated through extensive experiments on three commonly used QA datasets. The results show that these methods can enhance base retrieval models, even when fewer negative samples are used.

Acknowledgments

We thank all the reviewers for their helpful comments. This work is supported by the National Key R&D Program of China (2022ZD0116600).

References

Borgeaud, S.; Mensch, A.; Hoffmann, J.; Cai, T.; Rutherford, E.; et al. 2022. Improving Language Models by Retrieving from Trillions of Tokens. In *International Conference on Machine Learning*.

Cheng, H.; Zhu, Z.; Li, X.; Gong, Y.; Sun, X.; and Liu, Y. 2021. Learning with Instance-Dependent Label Noise: A Sample Sieve Approach. In *International Conference on Learning Representations*.

Chuang, C.-Y.; Robinson, J.; Lin, Y.-C.; Torralba, A.; and Jegelka, S. 2020. Debiased contrastive learning. *Advances in Neural Information Processing Systems*.

Ferguson, J.; Gardner, M.; Hajishirzi, H.; Khot, T.; and Dasigi, P. 2020. IIRC: A Dataset of Incomplete Information Reading Comprehension Questions. In *Proceedings of the 2020 Conference on EMNLP*. Association for Computational Linguistics.

Fu, H.; Zhang, Y.; Yu, H.; Sun, J.; Huang, F.; Si, L.; Li, Y.; and Nguyen, C.-T. 2022. Doc2Bot: Accessing Heterogeneous Documents via Conversational Bots. In *Findings of the Association for Computational Linguistics: EMNLP*.

Gao, L.; and Callan, J. 2021. Is your language model ready for dense representation fine-tuning. *arXiv preprint arXiv:2104.08253*.

Gao, L.; and Callan, J. 2022. Unsupervised Corpus Aware Language Model Pre-training for Dense Passage Retrieval. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*.

Gao, L.; Dai, Z.; and Callan, J. 2021. COIL: Revisit Exact Lexical Match in Information Retrieval with Contextualized Inverted List. In *Proceedings of the 2021 Conference of NAACL: Human Language Technologies*.

Glass, M. R.; Rossiello, G.; Chowdhury, M. F. M.; Naik, A.; Cai, P.; and Gliozzo, A. 2022. Re2G: Retrieve, Rerank, Generate. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

Guu, K.; Lee, K.; Tung, Z.; Pasupat, P.; and Chang, M. 2020. Retrieval Augmented Language Model Pre-Training. In *Proceedings of the 37th International Conference on Machine Learning*.

Hao, S.; Li, P.; Wu, R.; and Chu, X. 2022. A Model-Agnostic approach for learning with noisy labels of arbitrary distributions. In *International Conference on Data Engineering*. IEEE.

Huang, P.-S.; He, X.; Gao, J.; et al. 2013. Learning Deep Structured Semantic Models for Web Search Using Clickthrough Data. In *CIKM*.

Humeau, S.; Shuster, K.; Lachaux, M.-A.; and Weston, J. 2019. Poly-encoders: Architectures and Pre-training Strategies for Fast and Accurate Multi-sentence Scoring. In *International Conference on Learning Representations*.

Johnson, J.; Douze, M.; and Jégou, H. 2021. Billion-Scale Similarity Search with GPUs. *IEEE Trans. Big Data*, 535–547.

Joshi, M.; Choi, E.; Weld, D. S.; and Zettlemoyer, L. 2017. TriviaQA: A Large Scale Distantly Supervised Challenge Dataset for Reading Comprehension. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*.

Karpukhin, V.; Oguz, B.; Min, S.; Lewis, P.; Wu, L.; Edunov, S.; Chen, D.; and Yih, W.-t. 2020. Dense Passage Retrieval for Open-Domain Question Answering. In *Proceedings of the 2020 Conference on EMNLP*. Association for Computational Linguistics.

Khattab, O.; and Zaharia, M. 2020. Colbert: Efficient and effective passage search via contextualized late interaction over bert. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval*.

Kwiatkowski, T.; Palomaki, J.; Redfield, O.; Collins, M.; Parikh, A.; Alberti, C.; Epstein, D.; Polosukhin, I.; Devlin, J.; Lee, K.; et al. 2019. Natural questions: a benchmark for question answering research. *Transactions of the Association for Computational Linguistics*.

Lewis, P. S. H.; Perez, E.; Piktus, A.; Petroni, F.; Karpukhin, V.; Goyal, N.; Küttler, H.; Lewis, M.; Yih, W.; Rocktäschel, T.; Riedel, S.; and Kiela, D. 2020a. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020*.

Lewis, P. S. H.; Perez, E.; Piktus, A.; Petroni, F.; Karpukhin, V.; Goyal, N.; Küttler, H.; Lewis, M.; Yih, W.; Rocktäschel, T.; Riedel, S.; and Kiela, D. 2020b. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In *Advances in Neural Information Processing Systems*.

Lin, Z.; Gong, Y.; Liu, X.; Zhang, H.; Lin, C.; Dong, A.; Jiao, J.; Lu, J.; Jiang, D.; Majumder, R.; et al. 2023. Prod: Progressive distillation for dense retrieval. In *Proceedings of the ACM Web Conference*.

Liu, T.; and Tao, D. 2015. Classification with noisy labels by importance reweighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Liu, Y.; and Guo, H. 2020. Peer loss functions: Learning from noisy labels without knowing noise rates. In *International Conference on Machine Learning*. PMLR.

Lu, Y.; Liu, Y.; Liu, J.; Shi, Y.; Huang, Z.; Sun, S. F. Y.; Tian, H.; Wu, H.; Wang, S.; Yin, D.; et al. 2022. Ernie-search: Bridging cross-encoder with dual-encoder via self on-the-fly distillation for dense passage retrieval. *arXiv preprint arXiv:2205.09153*.

Luan, Y.; Eisenstein, J.; Toutanova, K.; and Collins, M. 2021. Sparse, dense, and attentional representations for text retrieval. *Transactions of the Association for Computational Linguistics*.

Manwani, N.; and Sastry, P. 2013. Noise tolerance under risk minimization. *IEEE Transactions on Cybernetics*.

Mialon, G.; Dessì, R.; Lomeli, M.; Nalmpantis, C.; Pasunuru, R.; Raileanu, R.; Rozière, B.; Schick, T.; Dwivedi-Yu, J.; Celikyilmaz, A.; et al. 2023. Augmented language models: a survey. *arXiv preprint arXiv:2302.07842*.

Natarajan, N.; Dhillon, I. S.; Ravikumar, P. K.; and Tewari, A. 2013. Learning with noisy labels. *Advances in Neural Information Processing Systems*.

Nguyen, T.; Rosenberg, M.; Song, X.; Gao, J.; Tiwary, S.; Majumder, R.; and Deng, L. 2016. MS MARCO: A Human Generated MAchine Reading COmprehension Dataset. *choice*.

Ni, A.; Gardner, M.; and Dasigi, P. 2021. Mitigating False-Negative Contexts in Multi-document Question Answering with Retrieval Marginalization. In *Proceedings of the 2021 Conference on EMNLP*. Association for Computational Linguistics.

Oguz, B.; Lakhotia, K.; Gupta, A.; Lewis, P.; Karpukhin, V.; Piktus, A.; Chen, X.; Riedel, S.; Yih, W.; Gupta, S.; et al. 2022. Domain-matched Pre-training Tasks for Dense Retrieval. In *Findings of the Association for Computational Linguistics: NAACL 2022-Findings*. Association for Computational Linguistics.

Patrini, G.; Rozza, A.; Krishna Menon, A.; Nock, R.; and Qu, L. 2017. Making deep neural networks robust to label noise: A loss correction approach. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Qu, Y.; Ding, Y.; Liu, J.; Liu, K.; Ren, R.; Zhao, W. X.; Dong, D.; Wu, H.; and Wang, H. 2021. RocketQA: An Optimized Training Approach to Dense Passage Retrieval for Open-Domain Question Answering. In *Proceedings of the 2021 Conference of the NAACL: Human Language Technologies*.

Reimers, N.; and Gurevych, I. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on EMNLP-IJCNLP*. Association for Computational Linguistics.

Ren, R.; Lv, S.; Qu, Y.; Liu, J.; Zhao, W. X.; She, Q.; Wu, H.; Wang, H.; and Wen, J.-R. 2021a. PAIR: Leveraging Passage-Centric Similarity Relation for Improving Dense Passage Retrieval. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*.

Ren, R.; Qu, Y.; Liu, J.; Zhao, W. X.; She, Q.; Wu, H.; Wang, H.; and Wen, J.-R. 2021b. RocketQAv2: A Joint Training Method for Dense Passage Retrieval and Passage Reranking. In *Proceedings of the 2021 Conference on EMNLP*.

Robinson, J. D.; Chuang, C.-Y.; Sra, S.; and Jegelka, S. 2020. Contrastive Learning with Hard Negative Samples. In *International Conference on Learning Representations*.

Sachan, D.; Patwary, M.; Shoeybi, M.; Kant, N.; Ping, W.; Hamilton, W. L.; and Catanzaro, B. 2021. End-to-End Training of Neural Retrievers for Open-Domain Question Answering. In *Proceddings of the Association for Computational Linguistics: ACL-IJCNLP 2021*.

Wu, X.; Ma, G.; Lin, M.; Lin, Z.; Wang, Z.; and Hu, S. 2023. Contextual masked auto-encoder for dense passage retrieval. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Xia, X.; Liu, T.; Han, B.; Wang, N.; Gong, M.; Liu, H.; Niu, G.; Tao, D.; and Sugiyama, M. 2020. Part-dependent label noise: Towards instance-dependent label noise. *Advances in Neural Information Processing Systems*.

Xiong, L.; Xiong, C.; Li, Y.; Tang, K.; Liu, J.; Bennett, P. N.; Ahmed, J.; and Overwijk, A. 2021. Approximate Nearest Neighbor Negative Contrastive Learning for Dense Text Retrieval. In *9th International Conference on Learning Representations*.

Yang, S.; and Seo, M. 2020. Is retriever merely an approximator of reader? *arXiv preprint arXiv:2010.10999*.

Yang, S.; Yang, E.; Han, B.; Liu, Y.; Xu, M.; Niu, G.; and Liu, T. 2022. Estimating instance-dependent Bayes-label transition matrix using a deep neural network. In *International Conference on Machine Learning*. PMLR.

Yao, Y.; Liu, T.; Han, B.; Gong, M.; Deng, J.; Niu, G.; and Sugiyama, M. 2020. Dual t: Reducing estimation error for transition matrix in label-noise learning. *Advances in Neural Information Processing Systems*.

Zhang, H.; Gong, Y.; Shen, Y.; Lv, J.; Duan, N.; and Chen, W. 2021. Adversarial Retriever-Ranker for Dense Text Retrieval. In *International Conference on Learning Representations*.

Zhang, S.; Liang, Y.; Gong, M.; Jiang, D.; and Duan, N. 2022. Multi-View Document Representation Learning for Open-Domain Dense Retrieval. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics.

Zhang, Y.; Fu, H.; Fu, C.; Yu, H.; Li, Y.; and Nguyen, C.-T. 2023. Coarse-To-Fine Knowledge Selection for Document Grounded Dialogs. In *2023 IEEE International Conference on Acoustics, Speech and Signal Processing*.

Zhou, K.; Gong, Y.; Liu, X.; Zhao, W. X.; Shen, Y.; Dong, A.; Lu, J.; Majumder, R.; Wen, J.-R.; and Duan, N. 2022. SimANS: Simple Ambiguous Negatives Sampling for Dense Text Retrieval. In *Proceedings of the 2022 Conference on EMNLP: Industry Track*.

Zhu, Z.; Liu, T.; and Liu, Y. 2021. A second-order approach to learning with instance-dependent label noise. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.