SimSCR: A Simple Supervised Contrastive Learning Framework for Response Selection of Dialogue Systems

Anonymous ACL submission

Abstract

Supervised contrastive learning has shown impressive performance across multiple NLP tasks (Gunel et al., 2020; Li et al., 2021; Gao et al., 2021), enhancing model generalization by shortening the distance between semantic representations of samples in the same category and increasing the distance between those of different categories. For the task of response selection, directly calculating the similarity between context and response may lead to suboptimal model performance due to insufficient attention mechanism interaction, as compared to traditional full attention methods. To address this issue, we propose an innovative interactive supervised contrastive learning framework that transforms the problem of response selection from classification into a matching issue by introducing a special response named anchor response during training, effectively applying contrastive learning to this task. This framework not only combines the advantages of deep context interaction found in traditional methods but also leverages the strong generalization capability of contrastive learning. Additionally, we introduce a heuristic method for hard negative responses sampling, which significantly reduces the need for large numbers of negative samples in contrastive learning. Applying our method, the results obtained on three publicly available response selection datasets have reached the current state-of-the-art level [1].

1 Introduction

Multi-turn dialogue systems are crucial in NLP, aiming to enable AI to converse with humans using natural language. These systems must comprehend context and generate fitting responses using two primary approaches: retrieval-based and generative methods. Retrieval-based systems select the best response from a scored set of candidates, while generative systems create new replies from contextual clues, offering more versatility but risking contextually inaccurate responses. To mitigate this, developers generate multiple responses for evaluation or use a Reward Model with PPO algorithm (Bai et al., 2022) for fine-tuning. Regardless of the approach, ranking candidate responses and selecting the appropriate effectively remains a key challenge.

In addressing the issue of candidate response selection, the research community has proposed a variety of approaches. Gu et al. (2019) and Chen and Wang (2019) proposed matching networks based on Bidirectional Long Short-Term Memory networks (Bi-LSTM). Tao et al. (2019) and Yuan et al. (2019) combined the attention mechanism with multi-hop interaction models using Gated Recurrent Unit (GRU) networks. With the rise of pretrained models such as BERT(Devlin et al., 2019), researchers have begun utilizing these models to improve the task of response selection, while also making customized enhancements considering the task's specificity. For instance, Gu et al. (2020) introduced a strategy to integrate role information of multi-turn dialogues into the BERT model. Humeau et al. (2019) and Song et al. (2023) presented methods capable of handling multiple candidate responses simultaneously and designed various attention strategies for the interaction between context and responses. Han et al. (2021) and Xu et al. (2021) proposed continual pretraining methods with self-supervised training objectives customized for the response selection task.

Previous research on the task of response selection has emphasized the interaction between context and responses and the design of pretraining objectives, with less focus on the limitations of the cross-entropy loss function. Liu et al. (2016); Cao et al. (2019) have shown that this loss function can hinder model generalization and is vulnerable to noise and attacks. Supervised contrastive

[1] Codes will be released when the manuscript is accepted.

learning is suggested (Gunel et al., 2020; Liu et al., 2016) to overcome these issues by clustering semantically similar samples and separating dissimilar ones. The temperature parameter in contrastive loss also affects the model's emphasis on harder samples (Wang and Liu, 2021) and can be tuned for optimal performance. This paper aims to develop an effective supervised contrastive learning framework for response selection.

In applying supervised contrastive learning to response selection, the challenges include selecting appropriate anchors and negative responses, as they affect model outcomes and computational efficiency. Unlike Li et al. (2021) that rely on complex data augmentation such as synonym replacement or dropout, this study introduces a simpler approach that avoids these operations. Our innovations have led to state-of-the-art results on various public datasets. The main contributions of this paper can be summarized as follows:

1. We propose a supervised contrastive learning framework tailored for the task of response selection, which can train models with better generalization than previous approaches based on cross-entropy classification.

2. On the foundation of the aforementioned framework, we introduce a heuristic method for hard negative responses sampling, which, based on similarity measures, further enhances the training efficiency and performance of the model.

3. Experiments conducted on three publicly available response selection datasets confirm the effectiveness of our method, with the results currently representing the best-known performance.

We will make the source code and models publicly available for other researchers to reproduce our results or for future studies.

2 Releated Work

Uni-Encoder In response selection task, binary classification based on cross-entropy is standard (Gu et al., 2020; Whang et al., 2019; Humeau et al., 2019; Han et al., 2021), labeling context-response pairs as 0 or 1 and processing them through binary classifiers. This approach can lead to redundant context encoding for multiple responses. To address this, UniEncoder, a BERT-based encoder, was introduced by Song et al. (2023), which pairs a single context segment with concatenated candidate responses, modifying BERT's attention mask to isolate responses from each other while aligning them with the context. This encoder uses a multi-class classification network for response selection, proving more efficient and effective than binary methods. The dialogue encoder in our study builds on UniEncoder by incorporating a unidirectional attention mechanism to reduce confusion in contexts with many candidate responses.

Supervised Contrastive Learning Contrastive learning has recently excelled in various NLP tasks, beneficial in both unsupervised and supervised contexts. It works by aligning closer the representations of anchors and positives, distancing them from negatives, and promoting uniform vector distribution. Gunel et al. (2020) shows that supervised contrastive learning fine-tuning enhances generalization and robustness, even in data-scarce situations. Additionally, Li et al. (2021) applied it to response selection, combining it with cross-entropy loss to improve generalization. This paper introduces a new contrastive learning-based response selection framework, distinct from prior work as it eschews cross-entropy, focusing exclusively on contrastive learning for model training and better generalization.

Data Augmentation Data augmentation in NLP improves task performance, including response selection, where the correlation between responses and contexts is leveraged. A common technique is reusing positive responses from other samples as negative ones for a given context (Li et al., 2019; Humeau et al., 2019; Song et al., 2023). Li et al. (2019) suggests that selecting challenging negative responses, rather than random ones, is more effective. This involves randomly picking multiple negative responses, ranking them by the model's scores, and choosing the lower-scoring ones as difficult distractors. To avoid high computational costs, the initial pool of negative responses should be limited. This paper suggests a heuristic approach using cosine similarity to simplify the selection of challenging negative responses and reduce computational demands.

3 Methodology

3.1 Problem Formalization

In this research, we consider a dataset $D = (c_i, r_i, y_i)_{i=1}^{N}$ composed of dialogues, where $c_i =$

$u_1, u_2, u_3 \ldots u_m$ represents the multi-turn dialogue context, r_i denotes the corresponding response, and y_i is the label with 1 indicating the correct response and 0 indicating the incorrect response. Our task is to learn a matching function f, which can accept any context c_i and response r_i as input, and output their matching score $f(c_i, r_i)$. During the testing phase, for a given context c_i, n candidate responses $(r_j)_{j=1}^n$ are provided, and the model needs to compute the score for each response $f(c_i, r_j)$, subsequently ranking the candidate responses based on their scores.

3.2 Model Structure

In this paper, we propose a method based on the BERT architecture, as shown in Figure 1. This method inputs the context and multiple candidate responses into the encoding network together, and encodes them through ample attention interaction. At the end of the candidate responses, we introduce a special response, composed of k special tokens, to serve as a anchor response for the contrastive learning distance measurement. Through our specially designed encoder, we obtain semantic representations for the anchor response as well as the positive and negative samples, which are then processed through a nonlinear projection layer.

Through contrastive learning, our model strives to minimize the distance between the anchor response and positive responses while maximizing the distance between the anchor response and negative responses. To enhance the effects of contrastive learning, we utilize a memory bank to record and update the semantic representations of samples and employ a straightforward yet efficient metric function to select more challenging negative responses. During the testing phase, it is only necessary to encode the context and candidate responses with the anchor response through the trained model, calculate the similarity between the anchor response and the semantic representations of each candidate response, and then rank them based on similarity.

3.2.1 Dialogue Encoding

In this work, we developed a BERT-based encoder with specific enhancements. The input sequence includes dialogue context, a positive response, several negative responses, and a special anchor response, detailed in Section 3.2.2. We concatenated these components, inserting role identifiers [SPK1] and [SPK2] between sentences of the context, a [CLS] token at the beginning of the context and responses, and [SEP] tokens as separators. To distinguish between the context and the responses, we used different segment ids, and for the positional encoding, each response's positional identifier was continuous with the context but discontinuous between different responses to minimize the influence of the positive and negative sample concatenation order on the model.

We modified the attention mechanism to prevent information leakage between positive and negative samples, as suggested by Song et al. (2023). Our custom attention allows unidirectional flow from responses to context, blocking the opposite direction and inter-response perception, while preserving bidirectional attention within the context and individual responses. This is illustrated in the attention mask matrix in Figure 1. The encoder facilitates interaction between each response and the context. After the encoding process, we average pool the token hidden states to obtain response representations $h_{r_0}, h_{r_1} \ldots h_{r_{k+n}}, h_{anc}$, which are further processed via a two-layer projector to produce semantic vectors $h'_{r_0}, h'_{r_1} \ldots h'_{r_{k+n}}, h'_{anc}$, in line with Chen et al. (2020).

3.2.2 Anchor Response

In the Masked Language Model (MLM), we mask some of the tokens in a sentence and then use the context to predict these tokens, thereby achieving self-supervised learning of token semantic representations. Inspired by this approach, we introduced a special sentence composed of several special tokens (such as [unused]) into the candidate responses, which we refer to as the anchor response. The purpose of this anchor response is to utilize context information to learn the semantic representation of appropriate responses. In a specific dialogue, there may be multiple correct responses, and the same response can be expressed in various ways; hence we cannot directly predict the composition of the correct response in terms of tokens as we do in the MLM task. However, multiple correct responses should be semantically similar, or in other words, their distance in the representation space should be relatively close. Therefore, we consider employing a contrastive learning approach. By minimizing the distance between the anchor response and positive responses in the semantic space, and maximizing the distance from negative responses, we aim to learn the semantic representation of the anchor response.

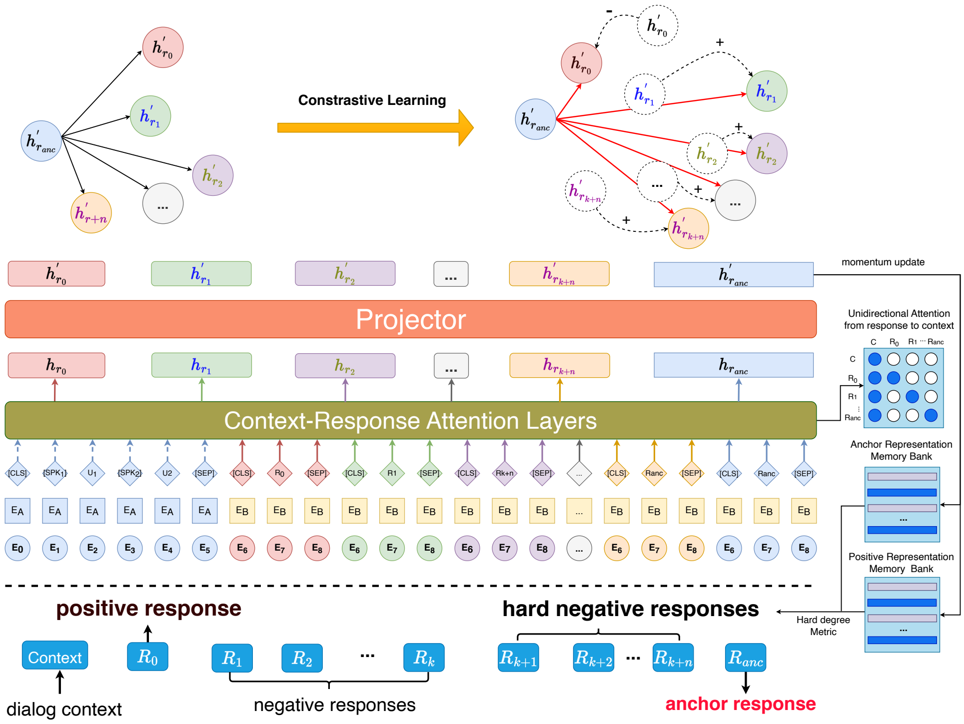

Figure 1: SimSCR Model Overview: a) Dialogue context, positive/negative responses, and an anchor response are encoded, with attention facilitating hidden state acquisition and contrastive learning refining semantic proximities. b) Responses use repeated position encoding with unidirectional attention to context. c) Memory Banks archive representation vectors and a metric function identifies challenging negatives.

3.2.3 Training Objective

Gunel et al. (2020) suggests that models trained with contrastive loss demonstrate stronger generalization capabilities and higher robustness to noise. Additionally, Wang and Liu (2021) also mentions that contrastive loss functions can automatically identify challenging negative samples. Based on these insights, we adopt contrastive loss as the optimization objective in this paper. Our objective function is defined as shown in Equation 1, where r_0 denotes the positive sample response, and sim represents the similarity measure function; in our experiments, we use cosine similarity. The τ is a temperature parameter, which ranges from 0 to 1, and can be adjusted during the training process to control the model's sensitivity to difficult negative responses.

$$\ell_{sim} = -log\left[\frac{e^{sim(h'_{R_{anc}}, h'_{R_0})/\tau}}{\sum_{j=0}^{N} e^{sim(h'_{R_{anc}}, h'_{R_j})/\tau}}\right] \quad (1)$$

3.3 Hard Negative Responses Sampling

Adopting the methodology from Humeau et al. (2019), we use positive responses from the same batch as negative responses for the current dialogue. Though more negatives could theoretically improve model performance, their computational cost is quadratic due to the Transformer encoder's complexity. Consequently, we have to limit the negative responses, which restricts the contrastive loss's efficiency in identifying hard negatives. To mitigate this, we propose a heuristic method for mining hard negative responses that selectively adds a small number of more challenging negative responses to the model input, thereby aiding the model in learning better semantic representations.

We utilize two memory banks M_{anc} and M_{pos}, each with a capacity of N (the dataset's size), to store the encoded and projected hidden states h'_{anc} and h'_{pos} of anchors and positives. For each sample i, we fetch its h'_{anc} from the anchor memory bank, select K h'_{pos} from the positive memory bank at random, and compute their cosine distances. This measurement helps discern the difficulty of negatives, allowing us to choose the hardest up to a limit of n, since our training objective is to optimize the cosine distance between these responses.

The two memory banks, initially seeded with random vectors, are updated synchronously post each forward model pass. For updates, we follow Wu et al. (2018b) 's momentum-based strategy, governed by hyperparameter m (0 to 1), dictating update smoothness as per Equation 2.

$$h_t = mh' + (1-m)h_{t-1} \quad (2)$$

The pseudocode for the hard negative responses sampling algorithm is as Algorithm 1.

Algorithm 1: Hard negative responses sampling algorithm

Data: M_{anc}, M_{pos}, K, difficulty coefficient β, sample index i, sampling number n

Result: hard negative responses of the i_{th} sample $S_i = \{r_j\}_{j=1}^{n}$

begin
- $h'_{anc} \leftarrow M_{anc}[i]$
- $C = \{h'_{pos_j}\}_{j=1}^{K} \leftarrow random(M_{pos}, K)$
- $G = \{d_j\}_{j=1}^{K} \leftarrow cosine(h'_{anc}, C)$
- $G' = \{(d_j, index_j)\}_{j=1}^{K} \leftarrow sort(G, reverse = true)$
- $r \leftarrow max(n+1, K * \beta)$
- $I = \{(d_k, index_k)\}_{k=1}^{n} \leftarrow random(G'[0:r], n+1)$
- $S \leftarrow []$
- **foreach** *element* e *in* I **do**
  - $index \leftarrow e[1]$
  - **if** $index \neq i$ **then**
    - $r^+_{index} \leftarrow$ positive response of the $index_{th}$ sample
    - $S.append(r^+_{index})$
  - **if** $len(S) >= n$ **then**
    - **return** S

4 Experiments

4.1 Datasets

This study conducted experiments on three public response selection datasets to evaluate the effectiveness of our method. These datasets include:

- **Douban Conversation** (Wu et al., 2017), which is a conversation dataset crawled from the Douban forum, a popular social media website in China.
- **E-commerce**(Wu et al., 2018a), comprising multi-turn dialogue data between customer service representatives and consumers collected from Taobao, the largest e-commerce platform in China.
- **Ubuntu Dialogue v1** (Lowe et al., 2015), an English multi-turn dialogue dataset widely used in technical support scenarios, particularly regarding the Ubuntu system.

The size and number of turns for each dataset are summarized in Table 1.

Table 1: Dataset Statistics

Dataset	Metric	Train	Valid	Test
Douban	Volume	1M	50K	6670
	Turns	6.69	6.75	5.95
	Pos:Neg	1:1	1:1	1.2:8.8
E-commerce	Volume	1M	10K	10K
	Turns	5.51	5.48	5.64
	Pos:Neg	1:1	1:1	1:9
Ubuntu v1	Volume	1M	500K	500K
	Turns	10.13	10.11	10.11
	Pos:Neg	1:1	1:9	1:9

Following the practice in Song et al. (2023), we have transformed the dataset D to ensure that each sample contains both a positive response r_i^+ and at least one negative response r_i^-, resulting in the updated dataset $D' = (c_i, r_i^+, r_i^-)_{i=1}^{N}$. This transformation includes the following steps:

- Aggregating samples with the same context c_i so that each sample contains only one positive response and at least one negative response. If multiple positive responses exist, they can be split into multiple independent samples.
- For samples that only have positive responses, we select responses from other samples to serve as negative responses.
- For samples that only have negative responses, we take the last turn of the original context as positive response and the other turns as new context, while keeping the negative responses unchanged.

4.2 Metrics

Similar to previous studies (Song et al., 2023; Han et al., 2021), we adopt recall as the primary evaluation metric, where recall is defined as $R_n@k$, indicating the proportion of correct responses that are ranked within the top k out of all n candidate replies by the model. Specifically, we use $R_{10}@1$, $R_{10}@2$, and $R_{10}@5$ as evaluation metrics for the Ubuntu v1 and E-commerce datasets. For the Douban dataset, since some dialogues in the test set contain multiple positive responses, we introduce additional evaluation metrics including top 1 accuracy $P@1$, Mean Average Precision (MAP), and Mean Reciprocal Rank (MRR) as a supplement.

4.3 Baselines

In the field of response selection tasks, the performance of pre-trained models has significantly surpassed that of traditional matching algorithms. Consequently, the baseline methods selected in this study are all based on pre-trained models, including BERT(Gu et al., 2020), RoBERTa-SS-DA(Lu et al., 2020), SA-BERT(Gu et al., 2020), SA-BERT+HCL(Su et al., 2021), UMSBERT and UMSBERT+(Whang et al., 2021), MDFN(Liu et al., 2021), BERT-SL(Xu et al., 2021), BERT-UMS+FGC(Li et al., 2021), BERT-FP (Han et al., 2021), Uni-Enc+BERT-FP(Song et al., 2023). Among them, **Uni-Enc+BERT-FP** was the previous state-of-the-art method.

4.4 Setup

Our model, built using the transformers library[2], initialized with pre-trained weights from HuggingFace[3] and fine-tuned weights by Han et al. (2021), was trained on an NVIDIA A100-SXM4-80GB GPU. We used the Adam optimizer with a cosine learning rate scheduler at a rate of 5e-5 and employed deepspeed[4] for efficient mixed-precision training. The training settings included a batch size of 16, a default contrastive learning temperature of 0.07, and 8 hard negative responses. The difficulty coefficient β was dynamically adjusted; it was 0.05 for the Douban and e-commerce datasets and 0.75 for Ubuntu-v1. Consistent with Song et al. (2023), we incorporated the MLM loss into the model's final loss function to enhance training stability and effectiveness.

4.5 Results

Table 2 presents the performance of the proposed SimSCR method on three different datasets in this study. The comparative analysis of the results shows that SimSCR has surpassed existing comparative methods across all three datasets. Specifically, SimSCR has achieved improvements of 4.8%, 1.9%, and 0.3% on the $R_{10}@1$ metric for the E-commerce, Douban, and Ubuntu v1 datasets, respectively. Notably, the performance gains of SimSCR are more significant on the Douban and E-commerce datasets, which have relatively lower benchmarks. This reflects the suitability of the proposed method in handling more challenging datasets. Additionally, the contrastive loss function adopted in this study is characterized by its ability to identify difficult negative samples (Wang and Liu, 2021), which may be a key factor contributing to the performance improvement.

Although BERT-UMS+FGC (Li et al., 2021) also applied a contrastive learning mechanism, the design of its loss function suggests that contrastive learning only serves as an auxiliary to the cross-entropy loss, thus limiting the performance gains it can provide. The experimental results indicate that, even without the adoption of post-training weights, SimSCR outperforms BERT-UMS+FGC on most evaluation metrics on the Douban and Ubuntu v1 datasets. In summary, these experimental results fully validate the effectiveness of the proposed method in the task of response selection.

5 Further Analysis

5.1 Supervised Contrastive Loss vs. Cross-Entropy Loss

For a candidate response R_i, let its hidden state representation obtained after encoding by the Encoder be denoted as h'_{r_i}. The cross-entropy loss function is defined as shown in Equation 3, where R_0 is the positive response, and the function f is a classifier composed of one or more linear layers, with the final output dimension being 1. Our objective function can also be expressed in the form of Equation 4, where f' is the projector in the model, consisting of two linear layers, with an output dimension of d. When $d = 1$, f and f' can be considered equivalent. Moreover, if we assume $f'(h_{r_{anc}}) \equiv 1$, and set the temperature parameter τ to 1 as well, then the expression $cos(f'(h_{r_{anc}}), f'(h_{r_j}))/\tau$ approximates $f(h_{r_j})$. In this case, Equation 4 simplifies to Equation 3. In other words, the cross-entropy loss function can be regarded as a special form of the supervised contrastive learning scheme.

$$\ell_{ce} = -log[\frac{e^{f(h_{r_0})}}{\sum_{j=0}^{N} e^{f(h_{r_j})}}] \tag{3}$$

$$\ell_{sim} = -log[\frac{e^{cos(f'(h_{r_{anc}}),f'(h_{r_0}))/\tau}}{\sum_{j=0}^{N} e^{cos(f'(h_{r_{anc}}),f'(h_{r_j}))/\tau}}] \tag{4}$$

[2]https://github.com/huggingface/transformers
[3]https://huggingface.co/models
[4]https://github.com/microsoft/DeepSpeed

Table 2: Main results on three benchmarks. † denotes post-training, other entries use naive BERT weights. Results acquired using SimSCR outperforms the original results with a significance level p-value < 0.05.

Models	E-commerce			Douban						Ubuntu v1		
	$R_{10}@1$	$R_{10}@2$	$R_{10}@5$	MAP	MRR	$P@1$	$R_{10}@1$	$R_{10}@2$	$R_{10}@5$	$R_{10}@1$	$R_{10}@2$	$R_{10}@5$
BERT(Gu et al., 2020)	0.610	0.814	0.973	0.591	0.633	0.454	0.280	0.470	0.828	0.808	0.897	0.975
RoBERTa-SS-DA(Lu et al., 2020)	0.627	0.835	0.980	0.602	0.646	0.460	0.280	0.495	0.847	0.826	0.909	0.978
SA-BERT(Gu et al., 2020)	0.704	0.879	0.985	0.619	0.659	0.496	0.313	0.481	0.847	0.855	0.928	0.983
SA-BERT+HCL(Su et al., 2021)	0.721	0.896	0.993	0.639	0.681	0.514	0.330	0.531	0.858	0.867	0.940	0.992
UMSBERT(Whang et al., 2021)	0.674	0.861	0.980	0.597	0.639	0.466	0.285	0.471	0.829	0.843	0.920	0.982
UMSBERT+(Whang et al., 2021)	0.762	0.905	0.986	0.625	0.664	0.499	0.318	0.482	0.858	0.875	0.942	0.988
MDFN(Liu et al., 2021)	0.639	0.829	0.971	0.624	0.663	0.498	0.325	0.511	0.855	0.866	0.932	0.984
BERT-SL(Xu et al., 2021)	0.776	0.919	0.991	-	-	-	-	-	-	0.884	0.946	0.990
† BERT-UMS+FGC(Li et al., 2021)	-	-	-	0.627	0.670	0.500	0.326	0.512	0.869	0.886	0.948	0.990
† BERT-FP(Han et al., 2021)	0.870	0.956	0.993	0.644	0.680	0.512	0.324	0.542	0.870	0.911	0.962	**0.994**
† Uni-Enc+BERT-FP(Song et al., 2023)	-	-	-	0.648	0.688	0.518	0.327	0.557	0.865	0.916	**0.965**	**0.994**
SimSCR(Ours)	0.888	0.959	0.993	0.651	0.692	0.519	0.329	0.553	0.868	0.890	0.947	0.989
† SimSCR+BERT-FP(Ours)	**0.918**	**0.972**	**0.998**	**0.669**	**0.706**	**0.532**	**0.346**	**0.584**	**0.887**	**0.919**	**0.965**	**0.994**

Table 3: Ablation study on Douban Conversation dataset. UARC - unidirectional attention from response to context, SCL - supervised contrastive learning loss, HNRS - hard negative responses sampling. UniEncoder + FP is the baseline.

Models	Full Douban Conversation Training Set						Douban Conversation Small-Sample Training Set (20k)					
	MAP	MRR	$P@1$	$R_{10}@1$	$R_{10}@2$	$R_{10}@5$	MAP	MRR	$P@1$	$R_{10}@1$	$R_{10}@2$	$R_{10}@5$
UniEncoder + FP(Song et al., 2023)	0.648	0.688	0.518	0.327	0.557	0.865	0.622	0.664	0.493	0.304	0.518	0.837
+ UARC (Ours)	0.649 (+0.1%)	0.690	0.523	0.333	0.546	0.867	0.625 (+0.3%)	0.665	0.489	0.304	0.524	0.846
+ UARC + SCL (Ours)	0.651 (+0.2%)	0.694	0.527	0.339	0.545	0.870	0.635 (+1.0%)	0.677	0.505	0.317	0.530	0.862
+ UARC + SCL + HNRS (SimSCR, Ours)	0.669 (+1.8%)	0.706	0.532	0.346	0.584	0.887	0.653 (+1.3%)	0.693	0.526	0.337	0.551	0.854

5.2 Ablation Study

To evaluate the contributions of individual components in the proposed method to the final performance, we conducted two sets of ablation experiments on the Douban Conversation dataset: one using the complete dataset and the other being a small-sample experiment, for which 20,000 data entries were extracted from the complete training set as training samples. We chose UniEncoder+FP as the comparison baseline and separately examined the effects of three key components in our method, with the specific results displayed in Table 3. From the results, for the complete training set, the HNRS (hard negative responses sampling) contributed most significantly, bringing a 1.8% improvement in MAP; followed by SCL (supervised contrastive learning), which yielded a 0.2% advancement; and compared to the bidirectional attention mechanism, the application of UARC (unidirectional attention from response to context) achieved better results, with an improvement of 0.1%. The reason is that unidirectional attention can prevent the indirect information leakage caused by bidirectional attention across multiple negative responses.

The small-sample experiment revealed that SCL led to a greater performance enhancement on the small-sample dataset (1.0% vs. 0.2%), indicating that models obtained through contrastive learning possess stronger generalization capabilities. Similarly, on the small-sample dataset, the UARC and HNRS also showed stable performance improvements. Notably, with only 20,000 training samples (1/25 of the full dataset), the application of the method proposed in this study has already surpassed the best results under the full dataset (0.653 vs. 0.648), a finding that has practical application potential in terms of reducing annotation costs.

5.3 Impact of Parameters

In the self-supervised contrastive learning framework proposed in this study, three key parameters have a significant impact on model performance: the temperature τ in the contrastive loss, the sampling number of hard negative responses n , and the difficulty coefficient h. To investigate the effects of these parameters, we conducted thorough experimental analyses on the Douban Conversation dataset, with the results displayed in Figure 2, 3 and 4.

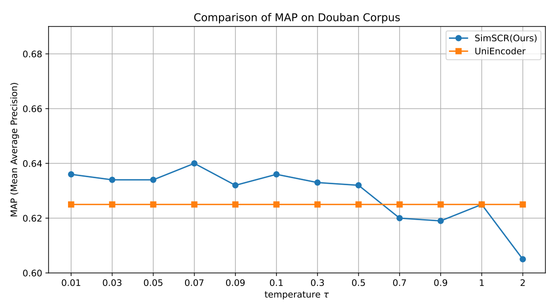

Figure 2: The Impact of Temperature τ

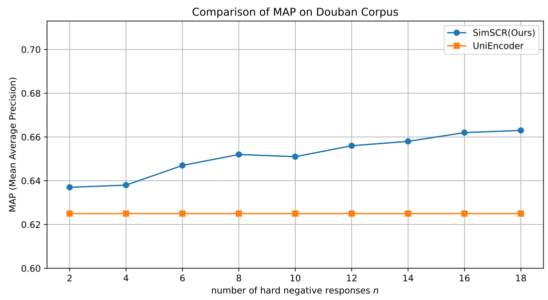

Figure 3: The Impact of Hard-Neg Responses Number n

Observing the experimental outcomes in Figure 2, it is evident that an increase in the temperature τ results in a performance trend that first rises and then declines, with the model achieving optimal performance at $\tau = 0.07$.

As the number of hard negative responses increases, the MAP value shows an upward trend (as presented in Figure 3); however, considering that a large number of hard negative responses would increase computational complexity, the experiment defaults to sampling 8 hard negative responses based on performance considerations.

As shown in Figure 4, a higher difficulty coefficient of negative responses has a positive impact on training effectiveness. When the difficulty coefficient $h = 1.0$, the model degenerates into a random sampling strategy, and the results indirectly confirms the superiority of the sampling approach proposed in this study compared to random sampling.

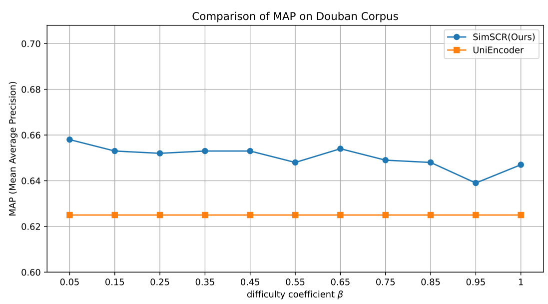

Figure 4: The Impact of Difficulty Coefficient β

6 Discussion

This chapter will explore the possibility of applying our method to other Transformer models, as well as the application extensions.

6.1 Extending to Other Transformer Models

Despite the remarkable achievements of the BERT model in the field of natural language processing, emerging models such as GPT(Brown et al., 2020) and GLM(Du et al., 2022) have also demonstrated outstanding performance and application potential. The supervised contrastive learning framework proposed in this paper is also applicable to these models. During application, since these models do not support segment IDs, different markers can be added before the context responses to distinguish them, thereby achieving a similar effect to segment IDs.

6.2 As a Reward Model for Training LLM

The successful application of the self-supervised contrastive learning method in dialogue response ranking tasks also inspires us to use it as a reward model during the training process of large-scale language models. Especially when using PPO for reinforcement learning training, an efficient reward model is crucial. Our method can serve as a means to train reward models, providing more accurate reward signals, thereby helping to guide the model's training and optimize the final performance.

7 Conclusion

In this paper, we introduce an innovative supervised contrastive learning framework to enhance the performance of response selection tasks. This approach not only draws on the advantages of deep context interaction found in tradition methods but also enhances the model's generalization ability through contrastive learning. To further improve model efficiency, we devised a heuristic sampling method for hard negative responses to reduce the dependence of contrastive learning on large negative sample sets. In experiments, our framework achieved state-of-the-art results on three public response selection datasets, proving the efficiency and practicality of the supervised contrastive learning framework and the negative sampling strategy. These achievements provide a powerful new tool for response selection tasks and offer valuable references for the future application of supervised contrastive learning in other NLP tasks. We look forward to extending our framework to a broader range of tasks in future research and exploring new ways to improve model effectiveness.

8 Limitations

Despite the remarkable achievements of this study in the task of response selection, we must acknowledge some limitations. First, although the interactive supervised contrastive learning framework performs well on the current datasets, these datasets may not fully cover all types of conversational scenarios. Therefore, the universality and robustness of the framework under different dialogue systems and diverse contexts still require further validation. Our interaction mechanism, while effective in facilitating attention interactions between context and responses, may not be optimized in its design, potentially leading to issues with computational efficiency or model complexity. Future research could explore more efficient interactive architectures to reduce the computational burden of the model while enhancing performance. Lastly, as a training strategy, the generalizability of contrastive learning across different NLP tasks still needs further research and validation. In particular, whether the approach presented in this paper remains effective in tasks substantially different in nature from response selection is a question that awaits future exploration. In summary, the method proposed in this study brings a new perspective and significant performance improvements to the task of response selection, but it is important to consider its limitations regarding data coverage, optimization of interaction mechanisms, and generalizability across different tasks. Future work will be devoted to deeper exploration and improvement in these areas.

References

Yuntao Bai, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn Drain, Stanislav Fort, Deep Ganguli, Tom Henighan, et al. 2022. Training a helpful and harmless assistant with reinforcement learning from human feedback. *arXiv preprint arXiv:2204.05862*.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

Kaidi Cao, Colin Wei, Adrien Gaidon, Nikos Arechiga, and Tengyu Ma. 2019. Learning imbalanced datasets with label-distribution-aware margin loss. *Advances in neural information processing systems*, 32.

Qian Chen and Wen Wang. 2019. Sequential attention-based network for noetic end-to-end response selection. *arXiv preprint arXiv:1901.02609*.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Zhengxiao Du, Yujie Qian, Xiao Liu, Ming Ding, Jiezhong Qiu, Zhilin Yang, and Jie Tang. 2022. Glm: General language model pretraining with autoregressive blank infilling. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 320–335.

Tianyu Gao, Xingcheng Yao, and Danqi Chen. 2021. SimCSE: Simple contrastive learning of sentence embeddings. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6894–6910, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Jia-Chen Gu, Tianda Li, Quan Liu, Zhen-Hua Ling, Zhiming Su, Si Wei, and Xiaodan Zhu. 2020. Speaker-aware bert for multi-turn response selection in retrieval-based chatbots. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 2041–2044.

Jia-Chen Gu, Zhen-Hua Ling, and Quan Liu. 2019. Interactive matching network for multi-turn response selection in retrieval-based chatbots. In *proceedings of the 28th ACM international conference on information and knowledge management*, pages 2321–2324.

Beliz Gunel, Jingfei Du, Alexis Conneau, and Veselin Stoyanov. 2020. Supervised contrastive learning for pre-trained language model fine-tuning. In *International Conference on Learning Representations*.

Janghoon Han, Taesuk Hong, Byoungjae Kim, Youngjoong Ko, and Jungyun Seo. 2021. Fine-grained post-training for improving retrieval-based dialogue systems. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1549–1558, Online. Association for Computational Linguistics.

Samuel Humeau, Kurt Shuster, Marie-Anne Lachaux, and Jason Weston. 2019. Poly-encoders: Architectures and pre-training strategies for fast and accurate multi-sentence scoring. In *International Conference on Learning Representations*.

Jia Li, Chongyang Tao, Wei Wu, Yansong Feng, Dongyan Zhao, and Rui Yan. 2019. Sampling matters! an empirical study of negative sampling strategies for learning of matching models in retrieval-based dialogue systems. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1291–1296, Hong Kong, China. Association for Computational Linguistics.

Yuntao Li, Can Xu, Huang Hu, Lei Sha, Yan Zhang, and Daxin Jiang. 2021. Small changes make big differences: Improving multi-turn response selection in dialogue systems via fine-grained contrastive learning. *arXiv preprint arXiv:2111.10154*.

Longxiang Liu, Zhuosheng Zhang, Hai Zhao, Xi Zhou, and Xiang Zhou. 2021. Filling the gap of utterance-aware and speaker-aware representation for multi-turn dialogue. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 13406–13414.

Weiyang Liu, Yandong Wen, Zhiding Yu, and Meng Yang. 2016. Large-margin softmax loss for convolutional neural networks. In *Proceedings of the 33rd International Conference on International Conference on Machine Learning-Volume 48*, pages 507–516.

Ryan Lowe, Nissan Pow, Iulian Serban, and Joelle Pineau. 2015. The Ubuntu dialogue corpus: A large dataset for research in unstructured multi-turn dialogue systems. In *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 285–294, Prague, Czech Republic. Association for Computational Linguistics.

Junyu Lu, Xiancong Ren, Yazhou Ren, Ao Liu, and Zenglin Xu. 2020. Improving contextual language models for response retrieval in multi-turn conversation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1805–1808.

Chiyu Song, Hongliang He, Haofei Yu, Pengfei Fang, Leyang Cui, and Zhenzhong Lan. 2023. Uni-encoder: A fast and accurate response selection paradigm for generation-based dialogue systems. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 6231–6244.

Yixuan Su, Deng Cai, Qingyu Zhou, Zibo Lin, Simon Baker, Yunbo Cao, Shuming Shi, Nigel Collier, and Yan Wang. 2021. Dialogue response selection with hierarchical curriculum learning. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1740–1751, Online. Association for Computational Linguistics.

Chongyang Tao, Wei Wu, Can Xu, Wenpeng Hu, Dongyan Zhao, and Rui Yan. 2019. One time of interaction may not be enough: Go deep with an interaction-over-interaction network for response selection in dialogues. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1–11, Florence, Italy. Association for Computational Linguistics.

Feng Wang and Huaping Liu. 2021. Understanding the behaviour of contrastive loss. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2495–2504.

Taesun Whang, Dongyub Lee, Chanhee Lee, Kisu Yang, Dongsuk Oh, and Heuiseok Lim. 2019. An effective domain adaptive post-training method for bert in response selection. *arXiv preprint arXiv:1908.04812*.

Taesun Whang, Dongyub Lee, Dongsuk Oh, Chanhee Lee, Kijong Han, Dong-hun Lee, and Saebyeok Lee. 2021. Do response selection models really know what's next? utterance manipulation strategies for multi-turn response selection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 14041–14049.

Xianchao Wu, Ander Martínez, and Momo Klyen. 2018a. Dialog generation using multi-turn reasoning neural networks. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2049–2059, New Orleans, Louisiana. Association for Computational Linguistics.

Yu Wu, Wei Wu, Chen Xing, Ming Zhou, and Zhoujun Li. 2017. Sequential matching network: A new architecture for multi-turn response selection in retrieval-based chatbots. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 496–505, Vancouver, Canada. Association for Computational Linguistics.

Zhirong Wu, Yuanjun Xiong, Stella X Yu, and Dahua Lin. 2018b. Unsupervised feature learning via non-parametric instance discrimination. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3733–3742.

Ruijian Xu, Chongyang Tao, Daxin Jiang, Xueliang Zhao, Dongyan Zhao, and Rui Yan. 2021. Learning an effective context-response matching model with self-supervised tasks for retrieval-based dialogues. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 14158–14166.

Chunyuan Yuan, Wei Zhou, Mingming Li, Shangwen Lv, Fuqing Zhu, Jizhong Han, and Songlin Hu. 2019.

Multi-hop selector network for multi-turn response selection in retrieval-based chatbots. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 111–120, Hong Kong, China. Association for Computational Linguistics.