CL4DIV: A Contrastive Learning Framework for Search Result Diversification

Zhirui Deng
Zhicheng Dou
zrdeng@ruc.edu.cn
dou@ruc.edu.cn
Gaoling School of Artificial Intelligence
Renmin University of China
Beijing, China

Yutao Zhu
Ji-Rong Wen
yutaozhu94@gmail.com
jrwen@ruc.edu.cn
Gaoling School of Artificial Intelligence
Renmin University of China
Beijing, China

ABSTRACT

Search result diversification aims to provide a diversified document ranking list so as to cover as many intents as possible and satisfy the various information needs of different users. Existing approaches usually represented documents by pretrained embeddings (such as doc2vec and Glove). These document representations cannot adequately represent the document's content and are hard to capture the intrinsic user's intent coverage of the given query. Moreover, the limited number of labeled data for search result diversification exacerbates the difficulty of obtaining more efficient document representations. To alleviate these problems and learn more effective document representations, we propose a **C**ontrastive **L**earning framework for search result **DIV**ersification (CL4DIV). Specifically, we design three contrastive learning tasks from the perspective of subtopics, documents, and candidate document sequences, which correspond to three essential elements in search result diversification. These training tasks are employed to pretrain the document encoder and the document sequence encoder, which are used in the diversified ranking model. Experimental results show that CL4DIV significantly outperforms all existing diversification models. Further analysis demonstrates that our method has wide applicability and can also be used to improve several existing methods.

CCS CONCEPTS

• **Information systems** → **Information retrieval diversity**.

KEYWORDS

Search Result Diversification, Contrastive Learning, Self-Supervised Learning

ACM Reference Format:
Zhirui Deng, Zhicheng Dou, Yutao Zhu, and Ji-Rong Wen. 2024. CL4DIV: A Contrastive Learning Framework for Search Result Diversification. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining (WSDM '24), March 4–8, 2024, Merida, Mexico.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3616855.3635851

WSDM '24, March 4–8, 2024, Merida, Mexico.

ACM ISBN 979-8-4007-0371-3/24/03…$15.00
https://doi.org/10.1145/3616855.3635851

1 INTRODUCTION

Search result diversification focuses on tackling the ambiguity of short queries and meeting the diverse information needs of different users. For example, users issuing "Starbucks" may expect results about "Starbucks beverage menu" or "the nearest Starbucks store". To cope with this problem, search result diversification models return relevant and diverse documents to cover more subtopics and better satisfy users' search intents with "ten blue links".

Pioneering work for diversifying search results dates back to MMR [3], which employed a hyperparameter λ to balance documents' relevance and diversity. The diversity was measured by the documents' dissimilarity. Following MMR, some studies [10, 28] measured documents' diversity by modeling their coverage of user intents in an unsupervised manner. These methods required extensive hyperparameter tuning and heavily relied on manually designed functions. To tackle these problems, researchers [36, 42] switched to supervised learning and constructed approximate ideal rankings as ground-truth rankings. This enabled automatic learning of diverse ranking functions and direct optimization of the loss function, resulting in significant performance improvements. With the development of deep learning, recent works [25, 38] further leveraged deep neural networks to train advanced models. Yet, most previous studies focus on designing elaborate network architectures or effective loss functions. A fundamental factor—*the quality of documents' initial representations*—has been neglected for a long time.

Previous works adopt pretrained initial document representations such as doc2vec [16] and Glove [23], which can help distinguish documents at a coarse-grained semantic level, but may fail to identify the subtle difference of subtopic coverage between documents for a query that is necessary for search result diversification. Even if the interaction between features is carefully designed, the unpolished initial document representation will inevitably impact the quality of the final ranking. Further complicating the case is the limited labeled training data available for this task. With insufficient training data, it is hard to train superior supervised models with rough initial document representations. Although most existing models [30, 38] assume that the training data is sufficient to train a diversification model, the fact is disappointing: the Web Track dataset contains only about 200 annotated queries in total. This problem becomes extremely prominent in neural-based models [25, 30], as they usually require a large amount of training data.

The above problems motivate us to reconsider more effective models of document representation targeted for search result diversification. Recently, contrastive learning has shown its effectiveness in many information retrieval (IR) tasks, such as dense retrieval [17] and video search [41]. With sufficient training samples automatically generated from raw data and carefully designed pretraining tasks, the model can learn the intrinsic correlation of data and generate more robust data representations. Nevertheless, introducing contrastive learning to search result diversification is challenging since search result diversification models need to figure out the subtle subtopic differences and inherent semantic correlation between documents under a specific query. Moreover, search result diversification models are required to re-rank a document sequence. Therefore, it is important to consider the differences between documents at the document sequence level.

Inspired by this, we propose a **C**ontrastive **L**earning framework for search result **DIV**ersification (**CL4DIV**). We devise three contrastive learning tasks, denoted as subtopic-based (SUB), document-based (DOC), and sequence-based (SEQ) contrastive learning, to build training pairs from different perspectives and pretrain document representation models. Specifically, for the **subtopic-based contrastive learning** task, our target is to compare the similarity between documents from a fine-grained subtopic coverage perspective, which simulates the general process of (explicit) diversification models. We leverage T5 [27] to generate subtopics for each document automatically, which avoids introducing additional annotation costs. Next, to model the intrinsic semantic correlation within a document and learn more robust document representations, we propose **document-based contrastive learning**, which comprises random passage deletion, random passage exchange, and dropout. They can reflect the potential variations of documents. Moreover, to measure the relationship between documents from the overall document sequence perspective, we design **sequence-based contrastive learning**. The document replacement and document reorder operations are conducted to generate augmented document sequence pairs. By comparing similar/dissimilar document sequences, the model can identify critical information for search result diversification.

We adopt the three tasks to pretrain the document encoder and document sequence encoder. The two pretrained encoders are further combined with an attention-based implicit diversified ranking model to score documents simultaneously. Experimental results show that, even with a simple diversified ranking network architecture, CL4DIV can still significantly outperform all existing methods. This clearly demonstrates the superiority of applying contrastive learning for search result diversification. Our further experiments show that the proposed method is also compatible with several existing models, and can greatly improve their performance. This validates the wide applicability of our method.

Our main contributions are three-fold:

(1) We propose a contrastive learning framework to enhance data representations in search result diversification. This is the first time that contrastive learning is applied to this task.

(2) We propose three contrastive learning tasks from different perspectives. They are all designed based on the characteristics of search result diversification and contribute to learning better representations for documents and document sequences.

(3) Our comprehensive experiments validate that our pre-training method can be easily combined with various existing models, which demonstrates the applicability and robustness of our approach.

2 RELATED WORK

2.1 Search Result Diversification

Search result diversification methods can be empirically divided into implicit methods and explicit methods.

Implicit Methods. Implicit methods [30, 40] compare the similarity between documents and do not require subtopics during the test stage of the diversified ranking. MMR [3] is the fundamental of most implicit methods which combines the ad-hoc query-document relevance and the dissimilarity between documents with a parameter λ. Depending on MMR, R-LTR [42] addressed diversification as a learning problem and built supervised training signals. PAMM [36] used a maximal marginal relevance model for ranking and directly optimized evaluation metrics. NTN [37] automatically learned a nonlinear novelty function. Due to the advancement of deep neural networks, Graph4DIV [30] measured documents' similarity based on their intent coverage and used a GNN to compute document's diversity features. DALETOR [38] devised diversification-aware losses to approach the optimal ranking. MO4SRD [39] scored documents with a probability distribution and directly optimized evaluation metrics. KEDIV [31] introduced entities and their relationships from an external knowledge base and leveraged them to model the diversity of documents. Among the models that do not introduce external knowledge, Graph4DIV is the state-of-the-art method.

Explicit Methods. xQuAD [28] introduced sub-queries to estimate the satisfaction of the document to the uncovered aspects. PM2 [10] optimized proportionality by iteratively determining the topic that best maintains the overall proportionality. Based on xQuAD and PM2, plenty of unsupervised explicit approaches [11, 14] were proposed. HxQuAD/HPM2 [14] adopted hierarchical subtopics to model users' intents while TxQuAD/TPM2 [11] directly modeled term-level subtopics to relieve the difficulties in subtopic mining. Subsequently, researchers [15] incorporated supervision signals and proposed a list-pairwise loss which can significantly improve the performance. DESA [25] adopted the self-attention mechanism to learn the implicit and explicit diversity features and estimate documents' novelty simultaneously. Based on DESA [25], GDESA [26] incorporated global interaction and document selection to approach global optimal ranking results. To tackle the challenge of lack of high-quality training samples, DVGAN [18] leveraged GAN to generate more training samples effectively.

Subtopics in previous models were either manually annotated or provided by existing search engines (*e.g.*, Google suggestions). In this paper, we propose an implicit method to generate subtopics with a T5 model automatically.

2.2 Pretraining for IR

Since pretrained language models [12, 24] have been proposed, they have been widely implemented and tailored for many IR tasks [20, 21]. PROP [20] adopted the query likelihood model to build fake query-document pairs for ad-hoc retrieval. By constructing such pairs, the model can measure the relevance between query and document more accurately and generate better ranking. Recently, contrastive learning [5] has been verified as a promising pretraining method that generates self-supervised signals from unsupervised data. Through training data representations with self-supervised auxiliary tasks, the model can achieve better performance on the

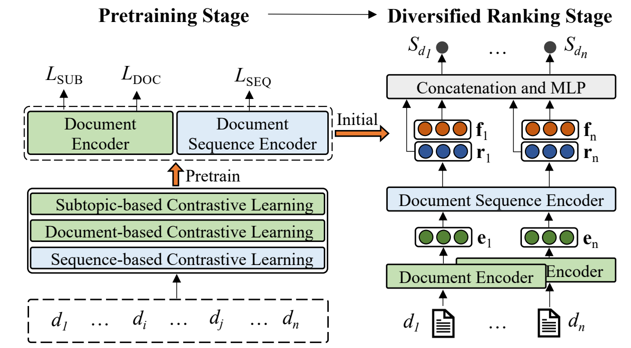


Figure 1: The architecture of CL4DIV containing two stages. The pretraining stage includes three contrastive learning tasks to pretrain the document encoder and the document sequence encoder. At the diversified ranking stage, the two encoders are initialized with the pretrained parameters to produce the final document ranking scores.

downstream tasks. COCA [43] designed three data augmentation strategies to enhance the representation of documents and user behavior sequences for session search.

In this paper, we focus on search result diversification and design contrastive learning tasks to build training pairs from unsupervised data to enhance the quality of data representations.

3 METHODOLOGY

Search result diversification aims to provide a document ranking covering more users' intents. Most existing methods utilize rough initial document representation (*e.g.*, doc2vec [16] and Glove [23]) which cannot capture the subtle subtopic differences between documents and suffer from the data sparsity problem. In this paper, we devise three contrastive learning tasks to optimize data representation. These tasks consider three essential elements of diversification, *i.e.*, subtopics, documents, and document sequences. Our CL4DIV is illustrated in Figure 1. It consists of a pretraining stage and a diversified ranking stage. The former leverages three contrastive learning tasks to pretrain the document encoder and the document sequence encoder, while the latter initializes the two encoders with pretrained parameters to diversify document ranking.

3.1 Problem Formulation.

Given the current query q and its initial ad-hoc ranking list $\mathcal{D} = \{d_1, \ldots, d_n\}$ that contains n candidate documents, search result diversification models re-rank these documents and generate a diversified document ranking list $\mathcal{R}$, in which novel documents are ranked higher and redundant ones are ranked lower.

Not only should search result diversification models consider modeling the query-document relevance like ad-hoc retrieval, but also focus on the diversity of documents. Since enumerating all possible document lists is an NP-hard problem, we follow previous studies [25, 38] and adopt an attention-based diversified ranking model to score all candidate documents simultaneously rather than greedy selection [30], *i.e.*, iteratively selecting the most novel and relevant document, to reduce the computational complexity.

3.2 Architecture of Diversified Ranking Model

Before introducing how to apply contrastive learning to search result diversification, we first describe the network architecture of our diversified ranking model, which is shown on the right side of Figure 1. First, each candidate document $d_i \in \mathcal{D}$ $(i \in [1, n])$ is encoded by a *document encoder* to obtain a document representation $\mathbf{e}_i$. Then, a *document sequence encoder* is used to model the relationship between documents and compute a contextualized document representation $\mathbf{r}_i$. Next, similar to existing methods [25], we compute the document ranking score S_{d_i} based on the concatenation of $\mathbf{r}_i$ and the relevance features $\mathbf{f}_i$ through a multi-layer perceptron (MLP). Finally, we generate the diversified document ranking by re-ranking the candidate documents according to their scores.

In this architecture, as shown in the left side of Figure 1, the document encoder is pretrained by the subtopic-based and document-based contrastive learning tasks (Section 3.3.1 and 3.3.2), while the document sequence encoder undergoes pretraining by the sequence-based contrastive learning tasks (Section 3.3.3). It is worth noting that our contrastive learning framework is flexible and compatible with several existing methods (*e.g.*, Graph4DIV [30], DESA [25] and DALETOR [38]) by replacing their document encoder and document sequence encoder by our pretrained ones. Experiments in Section 4.5.2 show the improvement provided by our framework.

The details of each component are introduced as follows.

(1) **Document Encoder.** The document encoder encodes each document $d_i \in \mathcal{D}$ into a vector. To learn semantic representations that can more accurately reflect the novelty of documents, we apply a Transformer [32] encoder DocE as the document encoder. Specifically, we compute document d_i's representation as follows:

$$d_i' = \texttt{[CLS]}d_i\texttt{[SEP]}, \quad \mathbf{e}_i = \text{DocE}(d_i')_{\texttt{[CLS]}}. \tag{1}$$

We use the [CLS] token as the document's representation.

(2) **Document Sequence Encoder.** In search result diversification, it is crucial to compare the documents' similarities and determine which one is more novel and should be ranked higher. Therefore, we apply another Transformer encoder SeqE as the document sequence encoder to enhance the interaction between documents. Concretely, for the representation $[\mathbf{e}_1, \ldots, \mathbf{e}_n]$ of documents $[d_1, \ldots, d_n]$, we update their representation as:

$$\mathbf{r}_1, \ldots, \mathbf{r}_n = \text{SeqE}(\mathbf{e}_1, \ldots, \mathbf{e}_n). \tag{2}$$

With the self-attention mechanism in the Transformer, $\mathbf{r}_i$ can consider its relationship with other documents. Therefore, we can treat $\mathbf{r}_i$ as the contextualized diversified document representation.

(3) **Relevance Features.** We use the same 18-dimension relevance features $\mathbf{f}_i$ as previous works [15, 18, 30], and details can be found in [15, 42]. The relevance features represent the query-document relevance, which is critical for diversification task [26].

(4) **Ranking Score Calculation.** After obtaining the diversified document representation and relevance features, we can calculate the final document ranking score through an MLP as:

$$S_{d_i} = \text{MLP}([\mathbf{r}_i; \mathbf{f}_i]), \tag{3}$$

where [;] is the concatenation operation.

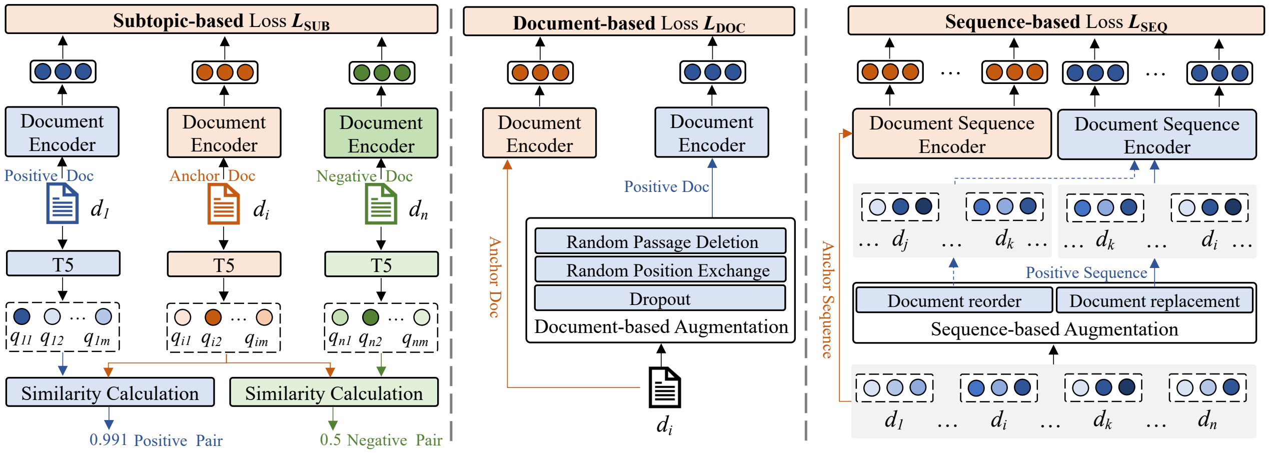


Figure 2: Three contrastive learning tasks in CL4DIV: Subtopic-based, Document-based, and Sequence-based. We devise different contrastive losses to optimize the parameters of the document encoder and the document sequence encoder.

3.3 Contrastive Learning Tasks

Current diversification methods are limited by their reliance on coarse document representation. Moreover, the data sparsity problem has long plagued the diversified ranking task, which directly impacts the training of document representation. Contrastive learning can mine the intrinsic content correlations in and between data samples and construct training samples in an unsupervised manner to learn high-quality data representation, which naturally fits our purpose. In this section, we will introduce three contrastive learning strategies tailored for search result diversification and explain how we use them to pretrain the document encoder and the document sequence encoder and apply them to the diversified ranking stage.

3.3.1 Subtopic-based Contrastive Learning. Comparing the similarity between documents and determining their novelty are critical steps for search result diversification. If two documents cover the same subtopics, their contents are intuitively similar. In other words, a document is more novel if it contains a greater number of subtopics not covered by other documents. Motivated by this observation, we first design subtopic-based contrastive learning, as shown in the left side of Figure 2, to reduce the representations' distance between two documents covering identical subtopics.

Subtopic-based Document Similarity. In search result diversification, the subtopic coverage information of each document is often manually labeled. However, due to the high cost, it is impractical to perform human annotation on a large number of documents. To tackle this problem, inspired by docTTTTTquery [22], we propose to use a generative method to obtain the subtopic of documents automatically. Specifically, we initialize a T5 [27] model by the released checkpoint of docTTTTTquery.[1] Then, we use this T5 model to generate m queries $q_{i,1}, \dots, q_{i,m}$ as subtopics for document d_i:

$$q_{i,1}, \dots, q_{i,m} = \text{T5}(d_i). \tag{4}$$

By this means, the long document is represented by several short subtopics. This method has two advantages: (1) the short queries can summarize the main content of the document, which greatly reduces the noise in the subsequent similarity calculation; and (2) the documents are viewed from the perspective of a subtopic, which naturally aligns with the goal of search result diversification.

To enhance the document encoder's capability of identifying similar and dissimilar documents in terms of subtopics, we design a contrastive learning method based on document pairs. In particular, considering two documents d_i and d_j, both of which have m subtopics $[q_{i,1}, \cdots, q_{i,m}]$ and $[q_{j,1}, \cdots, q_{j,m}]$, we represent them by a pretrained BERT model as:[2]

$$\mathbf{y}_i = \text{BERT}_{\texttt{[CLS]}}(\texttt{[CLS]}q_{i,1}\,\texttt{[SEP]}\,q_{i,2}, \dots, \texttt{[SEP]}\,q_{i,m}), \tag{5}$$

$$\mathbf{y}_j = \text{BERT}_{\texttt{[CLS]}}(\texttt{[CLS]}q_{j,1}\,\texttt{[SEP]}\,q_{j,2}, \dots, \texttt{[SEP]}\,q_{j,m}), \tag{6}$$

where the [SEP] tokens are added to separate each subtopic. Again, we use the representation of the [CLS] token as the subtopic-based document representation. Then, we compute the similarity between the documents d_i and d_j by a cosine function as follows:

$$y_{i,j} = \cos(\mathbf{y}_i, \mathbf{y}_j) = \frac{\mathbf{y}_i^\top \cdot \mathbf{y}_j}{|\mathbf{y}_i||\mathbf{y}_j|}. \tag{7}$$

Note that this similarity measures the documents' relationship from the aspect of subtopics. Other similarity calculation methods such as inner product are also practicable.

Contrastive Learning Objective. Based on the subtopic-based similarity scores, we can sample positive document pairs as:

$$R(d_i, d_j) = 1, \quad \text{if } y_{i,j} > \alpha, \tag{8}$$

where d_i and d_j construct a positive document pair for contrastive learning, and α is a threshold hyperparameter.

Given N positive pairs of documents in a mini-batch, for a positive pair (d_i, d_j), we treat the other $2(N-1)$ documents within the same mini-batch as their negative samples. Following previous studies [6, 43], the contrastive learning loss $\mathcal{L}_{\text{SUB}}$ is defined as:

$$\mathcal{L}_{\text{SUB}} = -\log \frac{\exp(\text{sim}(\mathbf{e}_i, \mathbf{e}_j)/\tau)}{\exp(\text{sim}(\mathbf{e}_i, \mathbf{e}_j)/\tau) + \sum\limits_{\mathbf{e}^- \in D^-} \exp(\text{sim}(\mathbf{e}_i, \mathbf{e}^-)/\tau)},$$

[1] https://huggingface.co/castorini/doc2query-t5-base-msmarco

[2] Checkpoint of pretrained BERT: https://huggingface.co/bert-base-uncased

where the similarity function $\text{sim}(\cdot)$ is also implemented by the cosine function as Equation (7). τ is a temperature hyperparameter. $\mathbf{e}_i$ and $\mathbf{e}_j$ are the representations of d_i and d_j computed by the document encoder denoted in Equation (1).

3.3.2 Document-based Contrastive Learning. In addition to subtopics, the document's actual content is crucial for determining its novelty. However, existing diversification methods often represent documents using pretrained embeddings (*e.g.*, doc2vec [16] or Glove [23]). Such representations are trained on fixed document content, making it not robust for potential variations. Inspired by recent data augmentation studies [1, 9, 17, 41], as shown in the middle of Figure 2, we design three augmentation strategies to handle minor content variations and employ contrastive learning to enable the model to focus on the overall content thereby learning more robust document representation.

Document-based Augmentation Strategies. We devise three augmentation strategies to modify document content.

(1) Random passage deletion. This method randomly deletes a passage in a document. The remaining part can be treated as a partial view of the original content, forcing the model to learn a more robust representation without relying on complete information.

(2) Passage position exchange. This approach randomly exchanges the position of two passages in a document. Some documents do not have "strict" orders, so this operation helps the model focus on learning content representation rather than its passage orders.

(3) Random dropout masking. Following SimCSE [13], we utilize the standard dropout [29] in the document encoder and pass a document to the encoder twice with random dropout masks. This operation can improve uniformity and keep steady alignment [33] which further enhances the robustness of representations.

Contrastive Learning Objective. We treat the original document and its augmentation as a positive pair while using other documents in the same mini-batch as negative samples D^-. Specifically, for a document d, we randomly select an augmentation approach and obtain a document pair (d, d'), where d' is the augmented document. We encode the two documents with the document encoder $\text{DocE}(\cdot)$ to obtain their representation $\mathbf{e}$ and $\mathbf{e}'$. The loss function $\mathcal{L}_{\text{DOC}}$ is defined as:

$$\mathcal{L}_{\text{DOC}} = -\log \frac{\exp(\text{sim}(\mathbf{e}, \mathbf{e}')/\tau)}{\exp(\text{sim}(\mathbf{e}, \mathbf{e}')/\tau) + \sum\limits_{\mathbf{e}^- \in D^-} \exp(\text{sim}(\mathbf{e}, \mathbf{e}^-)/\tau)}.$$

3.3.3 Sequence-based Contrastive Learning. Search result diversification focuses on returning a ranked document list and improving its subtopic richness. In addition to comparing the relationship of documents, the model should also consider the overall novelty of the document list (sequence). Therefore, we propose sequence-based contrastive learning to enhance the document sequence representation. Given the initial document sequence $\mathcal{D}$, a straightforward idea is to directly add some variations on $\mathcal{D}$. However, the limited number of training sequences in existing diversification datasets renders solely relying on these sequences insufficient for effective training. To cope with this problem, we perform k times random shuffling operation on the initial sequences, resulting in k different sequences. The premise that this design can be implemented is that our diversified ranking model is a position-invariant attention-based model. It can score all candidate documents simultaneously, so it is robust to the initial positions of documents.

Sequence-based Augmentation Strategies. Based on the shuffled document sequences, we propose two sequence augmentation strategies to generate contrastive learning training samples.

(1) Document replacement. Our first strategy is to replace a document in the sequence with another one. To guarantee query-document relevance and better measure the relationship between documents, we choose to select a document that is similar to the original one for replacement rather than using a random document. Similar to Section 3.3.1, we use subtopic-based document similarity to select similar documents. For a specific document d_i in the candidate document sequence $S = [d_1, \ldots, d_i, \ldots, d_n]$, we select another document d_j where $y_{i,j} > \alpha$ to replace it and generate a new sequence $S' = [d_1, \ldots, d_j, \ldots, d_n]$. We believe that $\mathcal{D}'$ is similar to $\mathcal{D}$ regarding subtopic coverage.

(2) Document reorder. We also randomly swap the position of two documents in the document sequence. This strategy is similar to *passage position exchange* in document-based contrastive learning. We consider that the model should pay more attention to the novelty of each document rather than their positions in the sequence.

Contrastive Learning Objective. We treat the original sequence and its augmentation as a positive pair and leverage contrastive learning to pull close their representation. In particular, for a document sequence $S = [d_1, \ldots, d_n]$, we randomly choose a strategy and apply it to obtain an augmented document sequence $S' = [d'_1, \ldots, d'_n]$. We first encode each document in the document sequence pair (S, S') with a document encoder $\text{DocE}(\cdot)$ and acquire their representation $[\mathbf{e}_1, \ldots, \mathbf{e}_n]$ and $[\mathbf{e}'_1, \ldots, \mathbf{e}'_n]$. Then, the two representations are encoded by the document sequence encoder $\text{SeqE}(\cdot)$, defined in Equation (2), to acquire the updated representation $[\mathbf{r}_1, \ldots, \mathbf{r}_n]$ and $[\mathbf{r}'_1, \ldots, \mathbf{r}'_n]$. We concatenate the updated representations and produce the document sequence representations of (S, S') by $\mathbf{s} = [\mathbf{r}_1; \ldots; \mathbf{r}_n]$ and $\mathbf{s}' = [\mathbf{r}'_1; \ldots; \mathbf{r}'_n]$. We define the loss function $\mathcal{L}_{\text{SEQ}}$ as follows:

$$\mathcal{L}_{\text{SEQ}} = -\log \frac{\exp(\text{sim}(\mathbf{s}, \mathbf{s}')/\tau)}{\exp(\text{sim}(\mathbf{s}, \mathbf{s}')/\tau) + \sum\limits_{\mathbf{s}^- \in S^-} \exp(\text{sim}(\mathbf{s}, \mathbf{s}^-)/\tau)}. \quad (9)$$

Similarly, the document sequence pair (S, S') is regarded as a positive pair and other document sequences in the same mini-batch are treated as the negative sample $\mathbf{s}^-$ in the negative sample set S^-.

3.4 Training and Optimization

The training of our model has two stages.

Pretraining. We optimize the loss of three pretraining tasks including $\mathcal{L}_{\text{SUB}}$, $\mathcal{L}_{\text{DOC}}$, and $\mathcal{L}_{\text{SEQ}}$ to pretrain the document encoder and the document sequence encoder.

Diversified Ranking. Following previous works [15, 25, 30], we leverage a list-pairwise loss to generate more training samples and optimize the model parameters of the diversified ranking part. Specifically, for a sample pair (r_1, r_2), it can be represented as (C, d_1, d_2), where C is the same previous document sequence context of document d_1 and d_2. (C, d_1) and (C, d_2) are the document sequence

r_1 and r_2, respectively. The loss function can be defined as follows:

$$\mathcal{L} = \sum_{q \in Q} \sum_{o \in O_q} |\Delta M| (y_o \log(P_{12}) + (1 - y_o) \log(1 - P_{12})), \quad (10)$$

where o is the sample pair in all sample pairs O_q of query q, $y_o = 1$ is positive and 0 is negative, $P_{12} = \frac{1}{1+\exp(S_{d_1} - S_{d_2})}$ is the probability of being positive and S_{d_1} and S_{d_2} is calculated by Equation (3). $\Delta M = M(r_1) - M(r_2)$ is the weight of this sample, and the larger the positive-negative rankings' gap, the more important the sample.

4 EXPERIMENTS

4.1 Datasets and Evaluation Metrics

Following previous studies [15, 18, 25, 30], we conduct experiments on the ClueWeb09 dataset [2], which consists of 200 queries of Web Track dataset from 2009 to 2012. Among the 200 queries, queries #95 and #100 are discarded as they have no diversity judgment. Each of the remaining 198 queries contains 3 to 8 subtopics, which are manually annotated users' intents. The binary relevance rating is given at the subtopic level. We leverage subtopics generated by T5 rather than labeled intents to build more self-supervised training data. Consistent with TREC Web Track and existing methods [15, 18, 30], we use the top 50 results of Lemur as the initial document ranking and compute all evaluation metrics based on the top 20 results of the diversification model.[3] For significance testing, we employ a two-tailed paired t-test with p-value < 0.05, which is the same as that used in existing works [18, 25, 30, 42].

To align with previous work [18], we use the official diversity evaluation metrics of Web Track, including ERR-IA [4], α-nDCG [7], and NRBP [8]. These metrics measure the diversity of document ranking by explicitly rewarding novelty and penalizing redundancy.

4.2 Baselines

To verify the effectiveness of our model CL4DIV, we compare it with several baseline methods including:

(1) Lemur and ListMLE [35] are representative non-diversified models. For a fair comparison with previous studies [15, 30], we use the same results of Lemur produced by the Indri engine. ListMLE is a typical learning-to-rank model without considering diversity.

(2) xQuAD [28], PM2 [10], HxQuAD, HPM2 [14], TxQuAD, and TPM2 [11] are typical unsupervised explicit diversification methods. They all use a parameter λ to combine the relevance score and the diversity score of a document. HxQuAD/HPM2 adopts hierarchical subtopics with an additional parameter α. TxQuAD/TPM2 leverages terms to model the query intent. Following [42], we use ListMLE as the prior relevance ranking for these models.

(3) DSSA [15], DVGAN [18], DESA [25] and GDESA [26] are supervised explicit methods. We use list-pairwise loss [15] to train the DSSA. GDESA was an extension model of DESA with greedy selection.

(4) MO4SRD [39], R-LTR [42], PAMM [36], NTN [37], DALETOR [38], and Graph4DIV [30] are typical implicit methods. We reproduce MO4SRD with Lemur results based on their released code.[4] We do not add relevance features because the $erf(\cdot)$ in MO4SRD will cause the loss to become NaN after adding them. For R-LTR and PAMM, we tune $h_s(R)$ from minimal, maximal, and average to achieve optimal performance. Moreover, for PAMM, we tune the number of positive rankings τ^+ and negative rankings τ^- per query. R-LTR-NTN and PAMM-NTN are NTN based on R-LTR and PAMM, respectively. It is worth noting that the vanilla DALETOR does not use relevance features. To align with previous works [25, 30, 42], we add the relevance features and train DALETOR based on the top 50 results of Lemur with the diversification-aware loss. Graph4DIV is currently the state-of-the-art method, and we use their released code to implement it.[5]

4.3 Implementation Details

We use the HuggingFace's Transformers [34] for implementation. For the document encoder, we use the same architecture as $\text{BERT}_{\text{base}}$ and initialize it with BERT parameters. The heads' number, hidden size, and the layer number of the document sequence encoder are set as 8, 400, and 2. The threshold α, subtopic number m, and shuffling times k are set as 0.99, 20 and 50. We use AdamW [19] optimizer in both stages. The SUB and DOC tasks are trained with a learning rate of 7e-5 and a batch size of 16 for 3 epochs. For the SEQ task, we train it with the learning rate of 1e-3 and the batch size of 1,048 for 1 epoch. For tasks SUB, DOC, and SEQ, we sample about 88k, 88k, and 1.4m contrastive pairs. The temperature τ in $\mathcal{L}_{\text{SUB}}$, $\mathcal{L}_{\text{DOC}}$, and $\mathcal{L}_{\text{SEQ}}$ is set as 0.4, 0.1, and 0.8 and their weights are set as 0.3, 0.7 and 1.0. To process 50 candidate documents at once efficiently, we set the maximum document length for the two encoders as 256 and divide each document into 8 passages for the DOC task. In the ranking stage, we set batch size and learning rate as 4 and 0.03 and tune learning rate from 0.03 to 1e-6. We use five-fold cross-validation based on α-nDCG to select the best model. Our code and more implementation details are released at Github.[6]

4.4 Overall Experimental Results

The overall results are shown in Table 1. We find that **CL4DIV significantly outperforms all existing models** which demonstrates its superiority. We further have the following observations.

(1) CL4DIV outperforms all other implicit methods by a significant margin across all evaluation metrics. In terms of α-nDCG, the improvement over the state-of-the-art method Graph4DIV is 1.8%. Graph4DIV is a greedy framework that constructs an intent graph to depict the relationship between different documents and employs graph neural networks for diversity features learning. In contrast, CL4DIV only uses two encoders and has a much simpler structure and faster ranking speed. The improvement obtained by CL4DIV demonstrates that our self-supervised pretraining framework for document representation learning is highly effective. It can capture the relationship between documents at a more granular level, resulting in better diversified document ranking.

(2) Intriguingly, CL4DIV also significantly outperforms all explicit methods. Concretely, the absolute value of α-nDCG is improved by 1.7% over the strong baseline GDESA. Remember that CL4DIV is a purely implicit method, meaning it does not rely on any explicit feature (*e.g.*, subtopics provided by Google suggestions or humans or T5) during the inference stage. This substantial improvement can be attributed to the proposed three contrastive learning tasks. By

[3] Lemur Service: http://boston.lti.cs.cmu.edu/Services/clueweb09_batch/

[4] MO4SRD: https://github.com/wildltr/ptranking

[5] Graph4DIV: https://github.com/su-zhan/Graph4DIV

[6] Our open source code: https://github.com/DengZhirui/CL4DIV/

Table 1: Performance of all methods. The best result is in bold. † indicates CL4DIV significantly outperforms all other methods in two-tailed paired t-test with p-value < 0.05.

Category	Method	ERR-IA	α-nDCG	NRBP
Ad-hoc	Lemur	.271	.369	.232
	ListMLE	.287	.387	.249
Explicit	xQuAD	.317	.413	.284
	TxQuAD	.308	.410	.272
	HxQuAD	.326	.421	.294
	PM2	.306	.411	.267
	TPM2	.291	.399	.250
	HPM2	.317	.420	.279
	DSSA	.356	.456	.326
	DESA	.363	.464	.332
	DVGAN	.367	.465	.334
	GDESA	.369	.469	.337
Implicit	MO4SRD	.283	.367	.252
	R-LTR	.303	.403	.267
	PAMM	.309	.411	.271
	R-LTR-NTN	.312	.415	.275
	PAMM-NTN	.311	.417	.272
	DALETOR	.364	.461	.333
	Graph4DIV	.370	.468	.338
	CL4DIV (Our)	**.393**†	**.486**†	**.364**†

Table 2: Results of ablation studies. We remove (1) three contrastive learning tasks, and (2) relevance features. We also report the performance of Graph4DIV as a comparison.

Method	Variant	ERR-IA	α-nDCG	NRBP
CL4DIV	Full	**.393**	**.486**	**.364**
	(1) w/o SUB	.381	.477	.352
	(1) w/o DOC	.385	.480	.357
	(1) w/o SEQ	.381	.478	.352
	(2) w/o REL	.303	.397	.268
Graph4DIV	Full	.370	.468	.338
	(2) w/o REL	.291	.387	.254

comparing similar documents/document sequences, the document encoder's and document sequence encoder's ability to model the novelty of documents can be greatly enhanced.

4.5 Discussion

We conduct a series of additional experiments to investigate the various aspects of CL4DIV in depth.

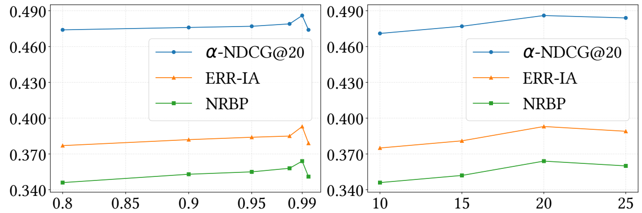
4.5.1 Ablation Studies. We conduct ablation studies to explore the impact of different modules in CL4DIV. First, we study the influence of our proposed three contrastive learning tasks, *i.e.*, subtopic-based (w/o SUB), document-based (w/o DOC), and sequence-based (w/o SEQ). Then, we remove the relevance features (w/o REL) to investigate their effect. Since the relevance features are commonly used in existing methods [15, 18, 25, 30, 42], we also report their influence on the best baseline method Graph4DIV as a comparison. The results are shown in Table 2, and we can see:

(1) Removing any contrastive learning task leads to performance degradation. This demonstrates that all three proposed tasks are beneficial for search result diversification. (2) Both contrastive learning of subtopic-based (SUB) and sequence-based (SEQ) contribute a lot to the final performance. Eliminating either of them results in a considerable drop in all metrics (*e.g.*, α-nDCG: 0.486 → 0.477 and NRBP: 0.364 → 0.352). This is consistent with our assumption, as the two tasks correspond to two essential components in search result diversification: (a) measuring the documents' novelty via subtopic coverage, and (b) evaluating the diversity of documents from the entire candidate document list perspective. (3) When the relevance features are discarded, the performance of both Graph4DIV and CL4DIV decreases. This demonstrates the importance of relevance modeling in search result diversification. Essentially, the relevance of documents is the basis for diversification. Existing studies have also reported similar findings [26]. Nevertheless, even without relevance features, CL4DIV is still superior to Graph4DIV. This validates again the benefit of our contrastive learning tasks.

Table 3: The results of equipping other baselines with our proposed contrastive learning tasks. Percentages in (·) are improvements by our contrastive learning tasks.

Method	ERR-IA	α-nDCG	NRBP
DALETOR	.364	.461	.333
+BERT	.373	.471	.342
+BERT+CL	**.393** (+2.0%)	.485 (+1.4%)	**.366** (+2.4%)
Graph4DIV	.370	.468	.338
+BERT	.376	.473	.346
+BERT+CL	.392 (+1.6%)	**.487** (+1.4%)	.364 (+1.8%)
DESA	.363	.464	.332
+BERT	.373	.468	.344
+BERT+CL	.387 (+1.4%)	.482 (+1.4%)	.359 (+1.5%)
CL4DIV	**.393** (+2.0%)	.486 (+1.7%)	.364 (+2.2%)
-CL	.373	.469	.342

4.5.2 Method Generalizability. In this work, we propose a contrastive learning method for search result diversification. To contrast documents at a finer-grained, we initialize the document encoder with BERT, train it with contrastive learning tasks, and obtain the contextualized representation. In contrast, most existing methods (*e.g.*, DALETOR [38], DESA [25], and Graph4DIV [30]) apply doc2vec embeddings to represent documents. To validate the generalizability of our method, we equip several strong baseline methods with the BERT representation (denoted as "+BERT") and our contrastive learning pretraining strategy (denoted as "+BERT+CL") and test their performance. We also validate the performance of our single diversified ranking model which leverages BERT as initial document representation by removing all contrastive learning tasks (denoted as "-CL"). Results are shown in Table 3 and we can observe:



(a) Threshold (b) Subtopic Number

Figure 3: Results of CL4DIV with different hyperparameters.

First, our contrastive learning tasks can bring improvement for all diversification methods. For example, DALETOR's performance can be enhanced by more than about 3% in terms of all metrics. This demonstrates that our method is both effective and highly compatible. Second, although the performance can be enhanced by using the BERT representation, the improvement is less than if our contrastive learning tasks are used. This indicates that the superiority of our method stems from our proposed contrastive learning tasks rather than merely using BERT's parameters.

4.5.3 Impact of Different Hyperparameter Settings. In our method, the threshold α and the number of generated subtopics m are two important hyperparameters. In this section, we conduct experiments to explore their impact on the final document ranking

The threshold α is an essential hyperparameter for determining the positive document pairs in SUB tasks. In the SEQ task, a larger α will restrict the document replacement operation replacing with documents that are more similar. Results in Figure 3 (a) show that the final performance can be gradually improved as α increases from 0.8 to 0.99. This implies that identifying more similar sequences benefits our contrastive learning method. However, a larger α will also result in fewer documents that can be used for training, *i.e.*, fewer similar sequences can be generated. This will also hurt the performance of contrastive learning (decreases when $\alpha > 0.99$).

As for the subtopic number m, according to the results in Figure 3 (b), generating more subtopics can represent document content more comprehensively and assist the model in measuring the similarity between documents more accurately. The performance starts to degrade when more than 20 subtopics are generated. This is because more subtopics will lead to redundancy, which will disrupt the document similarity computation.

4.5.4 Analysis of Generated Subtopics. We employ a T5 model to generate subtopics for each document instead of relying on the annotated users' intents. To verify the quality of the generated subtopics, we conduct both quantitative and qualitative analyses.

First, we directly compare the generated subtopics of each document with their corresponding annotated users' intents. The word-level F1 / Rouge-L score is used to evaluate the results. After calculation, we discover that all documents have an average F1 score of about 0.25. This result indicates that there is still a big gap between our generated subtopics and the human labels. Yet, our method can still provide promising results. Therefore, we speculate that if more accurate subtopics can be generated (*e.g.*, using recently proposed large language models), our method may obtain further improvement. Second, we randomly select two queries and their candidate documents. The results are shown in Table 4, where the key information is highlighted in blue. Considering the first query "Hip fractures", the generated subtopics cover the key document content (marked in blue). However, we also find that T5 generates some duplicated subtopics. For example, for the query "South Africa", the second and the third subtopic convey similar meanings. This result is consistent with our quantitative analysis. Although the performance of using T5 to generate subtopics is far from ideal, we can still get good results. We plan to investigate more powerful generative models as the subtopic generator in the future.

Table 4: Examples of subtopics generated by T5.

Query #173: *Hip fractures*	Doc: **clueweb09-en0060-60-20130**
Practical geriatrics: ... Americans spend more than $10 billion in direct costs of care[1] for the 250, 000 hip fractures that occur each year[2] ... Surgeons perform 125, 000 hip replacements[3] annually ... **Generated subtopics**: [1] how much money do people spend on hip replacements each year [2] how many hip fractures each year [3] what is the primary treatment for hip fracture for older adult	
Query #97: *South Africa*	Doc: **clueweb09-en0004-80-05666**
South Africa bed and breakfast accommodation ... Popular major towns[1]: Cape Town, Somerset West, ... Featured Listing[2,3]: The stanville inn is a selected service budget hotel ... Featured Listing[2,3]: ... **Generated subtopics**: [1] what is South Africa's major cities [2] where are the famous hotels in South Africa [3] hotels to stay in South Africa	

5 CONCLUSION AND FUTURE WORK

In this work, we proposed a two-stage self-supervised pretraining framework with contrastive learning to facilitate diversification-oriented data representation. First, we designed three contrastive learning tasks from the perspectives of subtopics, documents, and document sequences. The first two tasks helped the document encoder learn the fine-grained subtopic coverage differences of documents and focused on the main content of documents. The last task utilized the relationship between documents to enhance the document sequence encoder. Second, we employed the pretrained parameters to initialize the two encoders and further train them with diversified ranking objectives. Experimental results demonstrate the benefit of introducing contrastive learning to relieve rough data representation problems in search result diversification.

In the future, we plan to model longer documents and apply large language models for better document representation.

ACKNOWLEDGMENTS

Zhicheng Dou is the corresponding author. This work was supported by the National Natural Science Foundation of China No. 62272467, 61832017, and U2001212, and the fund for building world-class universities (disciplines) of Renmin University of China. The work was partially done at the Engineering Research Center of Next Generation Intelligent Search and Recommendation, Ministry of Education.

REFERENCES

[1] Samuel R. Bowman, Luke Vilnis, Oriol Vinyals, Andrew M. Dai, Rafal Józefowicz, and Samy Bengio. 2016. Generating Sentences from a Continuous Space. In *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning, CoNLL 2016, Berlin, Germany, August 11-12, 2016*, Yoav Goldberg and Stefan Riezler (Eds.). ACL, 10–21. https://doi.org/10.18653/v1/k16-1002

[2] Jamie Callan, Mark Hoy, Changkuk Yoo, and Le Zhao. 2009. The clueweb09 dataset, 2009. *URL http://boston. lti. cs. cmu. edu/Data/clueweb09* (2009).

[3] Jaime G. Carbonell and Jade Goldstein. 1998. The Use of MMR, Diversity-Based Reranking for Reordering Documents and Producing Summaries. In *SIGIR '98: Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, August 24-28 1998, Melbourne, Australia*, W. Bruce Croft, Alistair Moffat, C. J. van Rijsbergen, Ross Wilkinson, and Justin Zobel (Eds.). ACM, 335–336. https://doi.org/10.1145/290941.291025

[4] Olivier Chapelle, Donald Metlzer, Ya Zhang, and Pierre Grinspan. 2009. Expected reciprocal rank for graded relevance. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management, CIKM 2009, Hong Kong, China, November 2-6, 2009*, David Wai-Lok Cheung, Il-Yeol Song, Wesley W. Chu, Xiaohua Hu, and Jimmy Lin (Eds.). ACM, 621–630. https://doi.org/10.1145/1645953.1646033

[5] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey E. Hinton. 2020. A Simple Framework for Contrastive Learning of Visual Representations. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event (Proceedings of Machine Learning Research, Vol. 119)*. PMLR, 1597–1607. http://proceedings.mlr.press/v119/chen20j.html

[6] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey E. Hinton. 2020. A Simple Framework for Contrastive Learning of Visual Representations. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event (Proceedings of Machine Learning Research, Vol. 119)*. PMLR, 1597–1607. http://proceedings.mlr.press/v119/chen20j.html

[7] Charles L. A. Clarke, Maheedhar Kolla, Gordon V. Cormack, Olga Vechtomova, Azin Ashkan, Stefan Büttcher, and Ian MacKinnon. 2008. Novelty and diversity in information retrieval evaluation. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2008, Singapore, July 20-24, 2008*, Sung-Hyon Myaeng, Douglas W. Oard, Fabrizio Sebastiani, Tat-Seng Chua, and Mun-Kew Leong (Eds.). ACM, 659–666. https://doi.org/10.1145/1390334.1390446

[8] Charles L. A. Clarke, Maheedhar Kolla, and Olga Vechtomova. 2009. An Effectiveness Measure for Ambiguous and Underspecified Queries. In *Advances in Information Retrieval Theory, Second International Conference on the Theory of Information Retrieval, ICTIR 2009, Cambridge, UK, September 10-12, 2009, Proceedings (Lecture Notes in Computer Science, Vol. 5766)*, Leif Azzopardi, Gabriella Kazai, Stephen E. Robertson, Stefan M. Rüger, Milad Shokouhi, Dawei Song, and Emine Yilmaz (Eds.). Springer, 188–199. https://doi.org/10.1007/978-3-642-04417-5_17

[9] Andrew M. Dai and Quoc V. Le. 2015. Semi-supervised Sequence Learning. In *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, Corinna Cortes, Neil D. Lawrence, Daniel D. Lee, Masashi Sugiyama, and Roman Garnett (Eds.). 3079–3087. https://proceedings.neurips.cc/paper/2015/hash/7137debd45ae4d0ab9aa953017286b20-Abstract.html

[10] Van Dang and W. Bruce Croft. 2012. Diversity by proportionality: an election-based approach to search result diversification. In *The 35th International ACM SIGIR conference on research and development in Information Retrieval, SIGIR '12, Portland, OR, USA, August 12-16, 2012*, William R. Hersh, Jamie Callan, Yoelle Maarek, and Mark Sanderson (Eds.). ACM, 65–74. https://doi.org/10.1145/2348283.2348296

[11] Van Dang and W. Bruce Croft. 2013. Term level search result diversification. In *The 36th International ACM SIGIR conference on research and development in Information Retrieval, SIGIR '13, Dublin, Ireland - July 28 - August 01, 2013*, Gareth J. F. Jones, Paraic Sheridan, Diane Kelly, Maarten de Rijke, and Tetsuya Sakai (Eds.). ACM, 603–612. https://doi.org/10.1145/2484028.2484095

[12] 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, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, Jill Burstein, Christy Doran, and Thamar Solorio (Eds.). Association for Computational Linguistics, 4171–4186. https://doi.org/10.18653/v1/n19-1423

[13] 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, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021*, Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih (Eds.). Association for Computational Linguistics, 6894–6910. https://doi.org/10.18653/v1/2021.emnlp-main.552

[14] Sha Hu, Zhicheng Dou, Xiaojie Wang, Tetsuya Sakai, and Ji-Rong Wen. 2015. Search Result Diversification Based on Hierarchical Intents. In *Proceedings of the 24th ACM International Conference on Information and Knowledge Management, CIKM 2015, Melbourne, VIC, Australia, October 19 - 23, 2015*, James Bailey, Alistair Moffat, Charu C. Aggarwal, Maarten de Rijke, Ravi Kumar, Vanessa Murdock, Timos K. Sellis, and Jeffrey Xu Yu (Eds.). ACM, 63–72. https://doi.org/10.1145/2806416.2806455

[15] Zhengbao Jiang, Ji-Rong Wen, Zhicheng Dou, Wayne Xin Zhao, Jian-Yun Nie, and Ming Yue. 2017. Learning to Diversify Search Results via Subtopic Attention. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval, Shinjuku, Tokyo, Japan, August 7-11, 2017*, Noriko Kando, Tetsuya Sakai, Hideo Joho, Hang Li, Arjen P. de Vries, and Ryen W. White (Eds.). ACM, 545–554. https://doi.org/10.1145/3077136.3080805

[16] Quoc V. Le and Tomás Mikolov. 2014. Distributed Representations of Sentences and Documents. In *Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014 (JMLR Workshop and Conference Proceedings, Vol. 32)*. JMLR.org, 1188–1196. http://proceedings.mlr.press/v32/le14.html

[17] Yizhi Li, Zhenghao Liu, Chenyan Xiong, and Zhiyuan Liu. 2021. More Robust Dense Retrieval with Contrastive Dual Learning. In *ICTIR '21: The 2021 ACM SIGIR International Conference on the Theory of Information Retrieval, Virtual Event, Canada, July 11, 2021*, Faegheh Hasibi, Yi Fang, and Akiko Aizawa (Eds.). ACM, 287–296. https://doi.org/10.1145/3471158.3472245

[18] Jiongnan Liu, Zhicheng Dou, Xiaojie Wang, Shuqi Lu, and Ji-Rong Wen. 2020. DVGAN: A Minimax Game for Search Result Diversification Combining Explicit and Implicit Features. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020, Virtual Event, China, July 25-30, 2020*, Jimmy Huang, Yi Chang, Xueqi Cheng, Jaap Kamps, Vanessa Murdock, Ji-Rong Wen, and Yiqun Liu (Eds.). ACM, 479–488. https://doi.org/10.1145/3397271.3401084

[19] Ilya Loshchilov and Frank Hutter. 2019. Decoupled Weight Decay Regularization. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net. https://openreview.net/forum?id=Bkg6RiCqY7

[20] Xinyu Ma, Jiafeng Guo, Ruqing Zhang, Yixing Fan, Xiang Ji, and Xueqi Cheng. 2021. Prop: Pre-training with representative words prediction for ad-hoc retrieval. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*. 283–291.

[21] Xinyu Ma, Jiafeng Guo, Ruqing Zhang, Yixing Fan, Yingyan Li, and Xueqi Cheng. 2021. B-PROP: Bootstrapped Pre-training with Representative Words Prediction for Ad-hoc Retrieval. In *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021*, Fernando Diaz, Chirag Shah, Torsten Suel, Pablo Castells, Rosie Jones, and Tetsuya Sakai (Eds.). ACM, 1318–1327. https://doi.org/10.1145/3404835.3462869

[22] Rodrigo Nogueira, Jimmy Lin, and AI Epistemic. 2019. From doc2query to docTTTTTquery. *Online preprint* 6 (2019).

[23] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*, Alessandro Moschitti, Bo Pang, and Walter Daelemans (Eds.). ACL, 1532–1543. https://doi.org/10.3115/v1/d14-1162

[24] Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep Contextualized Word Representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 1 (Long Papers)*, Marilyn A. Walker, Heng Ji, and Amanda Stent (Eds.). Association for Computational Linguistics, 2227–2237. https://doi.org/10.18653/v1/n18-1202

[25] Xubo Qin, Zhicheng Dou, and Ji-Rong Wen. 2020. Diversifying Search Results using Self-Attention Network. In *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020*, Mathieu d'Aquin, Stefan Dietze, Claudia Hauff, Edward Curry, and Philippe Cudré-Mauroux (Eds.). ACM, 1265–1274. https://doi.org/10.1145/3340531.3411914

[26] Xubo Qin, Zhicheng Dou, Yutao Zhu, and Ji-Rong Wen. 2022. GDESA: Greedy Diversity Encoder with Self-Attention for Search Results Diversification. *ACM Transactions on Information Systems (TOIS)* (2022).

[27] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J Liu, et al. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.* 21, 140 (2020), 1–67.

[28] Rodrygo L. T. Santos, Craig Macdonald, and Iadh Ounis. 2010. Exploiting query reformulations for web search result diversification. In *Proceedings of the 19th International Conference on World Wide Web, WWW 2010, Raleigh, North Carolina, USA, April 26-30, 2010*, Michael Rappa, Paul Jones, Juliana Freire, and Soumen Chakrabarti (Eds.). ACM, 881–890. https://doi.org/10.1145/1772690.1772780

[29] Nitish Srivastava, Geoffrey E. Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15, 1 (2014), 1929–1958. https://doi.org/10.5555/2627435.2670313

[30] Zhan Su, Zhicheng Dou, Yutao Zhu, Xubo Qin, and Ji-Rong Wen. 2021. Modeling Intent Graph for Search Result Diversification. In *SIGIR '21: The 44th International*

ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021, Fernando Diaz, Chirag Shah, Torsten Suel, Pablo Castells, Rosie Jones, and Tetsuya Sakai (Eds.). ACM, 736–746. https://doi.org/10.1145/3404835.3462872

[31] Zhan Su, Zhicheng Dou, Yutao Zhu, and Ji-Rong Wen. 2022. Knowledge Enhanced Search Result Diversification. In *KDD '22: The 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, August 14 - 18, 2022*, Aidong Zhang and Huzefa Rangwala (Eds.). ACM, 1687–1695. https://doi.org/10.1145/3534678.3539459

[32] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).

[33] Tongzhou Wang and Phillip Isola. 2020. Understanding Contrastive Representation Learning through Alignment and Uniformity on the Hypersphere. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event (Proceedings of Machine Learning Research, Vol. 119)*. PMLR, 9929–9939. http://proceedings.mlr.press/v119/wang20k.html

[34] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2020. Transformers: State-of-the-Art Natural Language Processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, EMNLP 2020 - Demos, Online, November 16-20, 2020*, Qun Liu and David Schlangen (Eds.). Association for Computational Linguistics, 38–45. https://doi.org/10.18653/v1/2020.emnlp-demos.6

[35] Fen Xia, Tie-Yan Liu, Jue Wang, Wensheng Zhang, and Hang Li. 2008. Listwise approach to learning to rank: theory and algorithm. In *Machine Learning, Proceedings of the Twenty-Fifth International Conference (ICML 2008), Helsinki, Finland, June 5-9, 2008 (ACM International Conference Proceeding Series, Vol. 307)*, William W. Cohen, Andrew McCallum, and Sam T. Roweis (Eds.). ACM, 1192–1199. https://doi.org/10.1145/1390156.1390306

[36] Long Xia, Jun Xu, Yanyan Lan, Jiafeng Guo, and Xueqi Cheng. 2015. Learning Maximal Marginal Relevance Model via Directly Optimizing Diversity Evaluation Measures. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, Santiago, Chile, August 9-13, 2015*, Ricardo Baeza-Yates, Mounia Lalmas, Alistair Moffat, and Berthier A. Ribeiro-Neto (Eds.). ACM, 113–122. https://doi.org/10.1145/2766462.2767710

[37] Long Xia, Jun Xu, Yanyan Lan, Jiafeng Guo, and Xueqi Cheng. 2016. Modeling Document Novelty with Neural Tensor Network for Search Result Diversification. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval, SIGIR 2016, Pisa, Italy, July 17-21, 2016*, Raffaele Perego, Fabrizio Sebastiani, Javed A. Aslam, Ian Ruthven, and Justin Zobel (Eds.). ACM, 395–404. https://doi.org/10.1145/2911451.2911498

[38] Le Yan, Zhen Qin, Rama Kumar Pasumarthi, Xuanhui Wang, and Michael Bendersky. 2021. Diversification-Aware Learning to Rank using Distributed Representation. In *WWW '21: The Web Conference 2021, Virtual Event / Ljubljana, Slovenia, April 19-23, 2021*, Jure Leskovec, Marko Grobelnik, Marc Najork, Jie Tang, and Leila Zia (Eds.). ACM / IW3C2, 127–136. https://doi.org/10.1145/3442381.3449831

[39] Hai-Tao Yu. 2022. Optimize What You Evaluate With: Search Result Diversification Based on Metric Optimization. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022.* AAAI Press, 10399–10407. https://ojs.aaai.org/index.php/AAAI/article/view/21282

[40] Yisong Yue and Thorsten Joachims. 2008. Predicting diverse subsets using structural SVMs. In *Machine Learning, Proceedings of the Twenty-Fifth International Conference (ICML 2008), Helsinki, Finland, June 5-9, 2008 (ACM International Conference Proceeding Series, Vol. 307)*, William W. Cohen, Andrew McCallum, and Sam T. Roweis (Eds.). ACM, 1224–1231. https://doi.org/10.1145/1390156.1390310

[41] Hao Zhang, Aixin Sun, Wei Jing, Guoshun Nan, Liangli Zhen, Joey Tianyi Zhou, and Rick Siow Mong Goh. 2021. Video Corpus Moment Retrieval with Contrastive Learning. In *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11-15, 2021*, Fernando Diaz, Chirag Shah, Torsten Suel, Pablo Castells, Rosie Jones, and Tetsuya Sakai (Eds.). ACM, 685–695. https://doi.org/10.1145/3404835.3462874

[42] Yadong Zhu, Yanyan Lan, Jiafeng Guo, Xueqi Cheng, and Shuzi Niu. 2014. Learning for search result diversification. In *The 37th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '14, Gold Coast , QLD, Australia - July 06 - 11, 2014*, Shlomo Geva, Andrew Trotman, Peter Bruza, Charles L. A. Clarke, and Kalervo Järvelin (Eds.). ACM, 293–302. https://doi.org/10.1145/2600428.2609634

[43] Yutao Zhu, Jian-Yun Nie, Zhicheng Dou, Zhengyi Ma, Xinyu Zhang, Pan Du, Xiaochen Zuo, and Hao Jiang. 2021. Contrastive Learning of User Behavior Sequence for Context-Aware Document Ranking. In *CIKM '21: The 30th ACM International Conference on Information and Knowledge Management, Virtual Event, Queensland, Australia, November 1 - 5, 2021*, Gianluca Demartini, Guido Zuccon, J. Shane Culpepper, Zi Huang, and Hanghang Tong (Eds.). ACM, 2780–2791. https://doi.org/10.1145/3459637.3482243