Beyond Words: A Comprehensive Survey of Sentence Representations

Anonymous ACL submission

Abstract

Sentence representations are a critical component in several applications such as retrieval, question answering, and text classification. They capture the meaning of a sentence, enabling machines to understand and reason over human language. In recent years, significant progress has been made in developing methods for learning sentence representations, including unsupervised, supervised, and transfer learning approaches. In this paper, we provide an overview of the different methods for sentence representation learning, focusing mostly on deep learning models. We provide a systematic organization of the literature on sentence representation learning, highlighting the key contributions and challenges in this area. Overall, our review highlights the importance of this area in natural language processing, the progress made in sentence representation learning, and the challenges that remain. We conclude with directions for future research, suggesting potential avenues for improving the quality and efficiency of sentence representations.

1 Introduction

The *sentence*, together with the *word*, are the two fundamental grammatical units of human languages. Representing sentences for machine learning, which involves transforming a sentence into a vector or a fixed-length representation is a fundamental component of NLP. The quality of these representations affects the performance of downstream NLP tasks like text classification and text similarity (Conneau and Kiela, 2018).

Deep learning models have played a major role in obtaining sentence representations. While there have been significant advancements in the development of large language models (LLMs) such as GPT-3 (Brown et al., 2020), BLOOM (Workshop, 2023), they learn through effective word representations and modelling of the language at the (next) word level. Endowing the models the ability to learn effective representations of higher linguistic units beyond words – such as sentences – is useful.

For instance, sentence representations are useful in retrieving semantically similar documents prior to generation. LangChain[1] and various other frameworks, (Khattab et al., 2023), have underscored the critical demand for proficient sentence representations. The documents retrieved serve as valuable resources for generating fact-based responses, accommodating custom documents to address user queries, and fulfilling other essential functions.

However, current language models exhibit drawbacks in obtaining sentence representations out-of-the-box. For instance, Ethayarajh (2019) showed that out-of-the-box representations from BERT (Devlin et al., 2019) are fraught with problems such as anisotropy—representations occupying a narrow cone, making every representation closer to all others. Also, they are impractical for application scenarios: finding the best match for a query takes hours (Reimers and Gurevych, 2019).

To overcome the inadequacy of directly using sentence representations from language models, numerous methods have been developed. Several works have proposed to post-process the representations from BERT to alleviate the anisotropy (Li et al., 2020; Huang et al., 2021b) or repurpose representations from different layers of the model (Kim et al., 2021). But there has been a steadily growing body of works that move away from such post-processing and introduce new methods.

Perhaps due to the rapid advancements in the field, there are no literature reviews discussing the diverse range of techniques for learning sentence representations. The present paper offers a review of these techniques, with a specific emphasis on deep learning methods. Our review caters to two audiences: (a) Researchers from various fields seeking to get insights into recent breakthroughs in sentence representations, and (b) researchers aiming

[1] https://github.com/hwchase17/langchain

to advance the field of sentence representations.

1.1 Overview

We structure our literature review as follows:

- § 2 provides a brief history of methods to learn sentence representations and the different components of a modern framework.
- § 3 provides a review of supervised sentence representations that use labeled data to learn sentence representations.
- § 4 reviews methods that use unlabeled data to learn sentence representations (also called unsupervised sentence representation learning), a major focus of recent methods.
- § 5 describes methods that draw inspiration from other fields such as computer vision and
- § 6 provides a discussion of trends and analysis.
- § 7 discusses the challenges and suggests some future directions for research.

2 Background

2.1 Sentence Representations

Before the advent of neural networks, bag-of-words models were commonly used to represent sentences, but they suffered from limitations such as being unable to capture the relationships between words or the overall structure of the sentence.

Numerous efforts have aimed to improve sentence representations through neural networks. Inspired by Word2Vec (Pennington et al., 2014), Skip-Thought Vectors (Kiros et al., 2015) were trained to predict the surrounding sentences of a given target sentence. Subsequently, Conneau and Kiela (2018) employed various RNN networks to produce sentence embeddings, exploring their linguistic attributes, including part-of-speech tags, verb tense and named entity recognition. Notably, this study utilized NLI data for neural network training, predating the emergence of extensive pretrained models such as BERT (Devlin et al., 2019). BERT and similar models have since become a foundational framework for enhancing sentence representations.

2.2 Components of Sentence Representations

Neural networks have become the de-facto standard for learning sentence representations. The network takes two sentences as input and creates a vector for each sentence. These vectors are then trained to be similar for sentences that mean the same thing and different for sentences with different meanings. Learning sentence representations using neural networks has the following generic components (Figure 1):

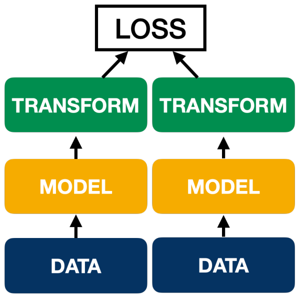

Figure 1: The components of an architecture to learn sentence representations. There are four main components: 1) ***Data*** - Obtaining positive and negative examples either using supervised data or some transformation 2) ***Model*** - Generally a pretrained model that has been trained on large quantities of gneeral text. 3) ***Transform*** - Some transformation applied to the representations from the model to obtain sentence representations and 4) ***Loss*** - Losses that bring semantically similar sentences closer together and others apart.

1. **Data**: Data used for learning sentence representations consists of pairs of semantically similar sentences, which can be either annotated by humans or generated through transformations to create positive and negative sentence pairs. (c.f. §§ 4.1 and 4.3).
2. **Model**: A sentence representation extraction model is a neural network backbone model unless specified otherwise. The backbone model can take the form of a RNN or pretrained transformer models like BERT (Devlin et al., 2019) or T5 (Raffel et al., 2020).
3. **Transform**: Neural network representations are often not well suited for use as sentence representations directly. While the `[CLS]` representations from BERT can serve as such, Reimers and Gurevych (2019) propose a pooling mechanism to obtain sentence representations by aggregating the representations of tokens. The type of transformation required depends on the type of model.
4. **Loss**: Contrastive learning is often used for sentence representations. The objective is to bring semantically similar examples closer together while pushing dissimilar examples further apart. Specifically, given a set of example pairs $\mathcal{D} = \{x_i, x_i^p\}$, a model is used to obtain representations for each pair, denoted h_i and h_i^p.

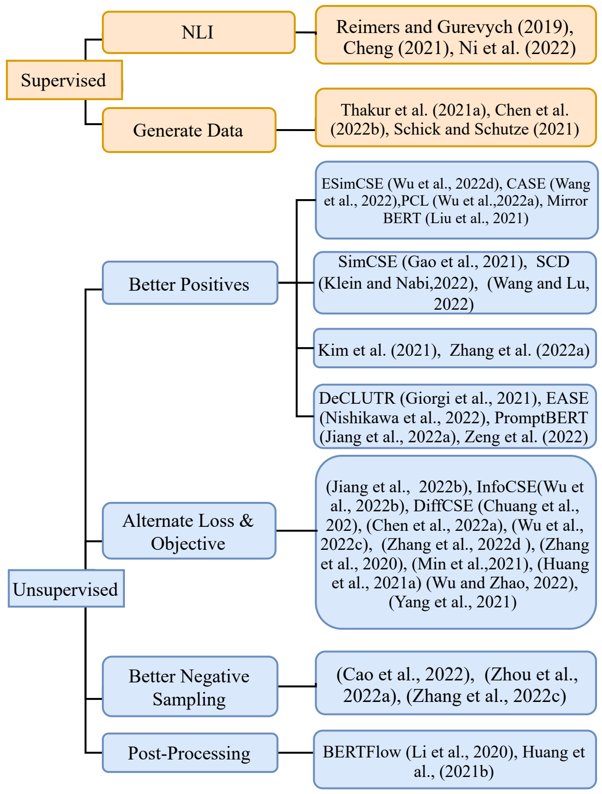

Figure 2: Overview of sentence representation methods.

The contrastive loss for an example is:

$$l_i = -\log \frac{e^{sim(h_i, h_i^p)}}{\sum_{j=1}^{N} e^{sim(h_i, h_j)}}$$

where N is the size of a mini-batch, $sim(\cdot,\cdot)$ is the similarity function which plays a crucial role. However, when selecting an appropriate loss function, several factors need to be considered. These factors include the choice of similarity measures and the characteristics of the negative examples.

In their influential paper, Reimers and Gurevych (2019) utilized this versatile framework to generate highly effective sentence embeddings, which has subsequently served as a cornerstone for further research. This framework, commonly referred to as the bi-encoder approach, involves encoding the *query* and *candidate* separately. However, an alternative approach exists where the *query* and *candidate* can be concatenated and encoded by a single model, facilitating interactions between words. This variant is known as the cross encoder.

Figure 2 illustrates the progression of work aimed at improving sentence representations. Two primary approaches stand out: supervised and unsupervised methods. For a clearer understanding of innovations, we categorize these methods based on variations of common techniques. Each category identifies contributions that target specific components (Figure 1): The "better positives" category focuses on refining augmentation techniques, primarily addressing the "data" component. Conversely, the "alternate loss and objectives" category explores improvements in the contrastive "loss" function. These dynamic interactions between categories are further depicted in Table 1.

3 Supervised Sentence Representations

Natural language understanding involves intricate reasoning. One way to learn better sentence representations is by excelling at tasks that demand reasoning. Large-scale supervised datasets for natural language understanding have emerged over the years: SNLI (Bowman et al., 2015), MNLI (Williams et al., 2018), ANLI (Nie et al., 2020). To that end, neural network methods utilize supervised datasets to learn sentence representations.

3.1 Natural Language Inference

Natural Language Inference (NLI) is the process of determining the logical relationship between a premise (an assumed true sentence) and a hypothesis (a possibly true sentence). The objective of NLI is to determine whether the hypothesis can be logically inferred from the premise (entailment), contradicts the premise (contradiction), or is neutral with respect to it (Dagan et al., 2013). NLI serves as a proxy for evaluating natural language understanding. According to Conneau et al. (2017), learning sentence representations using NLI data can be effectively transferred to other NLP tasks, demonstrating the generality of this approach.

In § 2.2, we discussed Siamese-BERT networks as presented by Reimers and Gurevych (2019). There are two noteworthy components to this model. First, processing inputs individually without promoting interaction between words; second, using an encoder like BERT that is not generative as its backbone model. The first component is computationally efficient but has been found to result in poorer performance compared to methods that promote interaction between words (Reimers and Gurevych, 2019). This lack of interaction can limit the network's ability to capture the nuances of language, and may result in less accurate sentence embeddings. In order to solve this, Cheng (2021) incorporated word-level interaction features into the sentence embedding while maintaining the efficiency of Siamese-BERT networks. Their approach makes use of ideas from knowledge distillation (Hinton et al., 2015): using the rich knowledge in pretrained cross-encoders and significantly improving the performance of Siamese-BERT.

Meanwhile, generative models – that generate text left to right, have been pretrained on huge amounts of data, and can perform a myriad of tasks. Ni et al. (2022a) examined the use of generative models as backbone for extracting sentence embeddings. They consider three methods to obtain sentence representations from a pretrained T5 model: the representation of the first token of the encoder, the representation of the first generated token of the decoder, or the mean of the representations from the encoder. They found them to be performant showing the utility of generative models for obtaining sentence representations.

3.2 Generating Data

Acquiring supervised data to train sentence representations is difficult task. However, in recent years, pre-trained models have emerged as a potential solution for generating training data. Furthermore, pre-trained models can serve as weak labelers to create silver data.

Cross-encoders that are pretrained on NLI data can be used to obtain silver data. In order to do this, Thakur et al. (2021a) suggest Augmented-SBERT. Their approach involves using different strategies to mine sentence pairs, followed by labeling them using a cross-encoder to create silver data. The silver data is then combined with the human-labelled training dataset, and a Siamese-BERT network is trained. However, this method requires mining appropriate sentence pairs first.

Rather than relying solely on obtaining supervised data, researchers are exploring the use of generative language models to create large amounts of synthetic training data for sentence encoders. This approach has the potential to produce high-quality training data at scale, addressing some of the challenges associated with supervised data acquisition. For instance, Chen et al. (2022b) demonstrate the use of a T5 model trained to generate entailment or contradiction pairs for a given sentence. However, this method still needs to provision a sentence to generate the entailment/contradiction pairs.

DINO, introduced by Schick and Schütze (2021), automates the generation of NLI data instructions using GPT2-XL. This approach eliminates the need for providing a sentence to generate entailment or contradiction pairs. Models trained on the resulting STS-Dino dataset outperform strong baselines on multiple semantic textual similarity datasets.

4 Unsupervised Sentence Representations

Unsupervised sentence representation learning does not require labeled data to learn sentence representations. Thus this approach has garnered significant attention in recent years. Unlike supervised methods, unsupervised learning techniques do not rely on explicit positive and negative examples but instead employ alternative techniques to mine them. Additionally, they may also modfiy the learning objectives.

4.1 Better Positives

Contrastive learning techniques optimize sentence representations by contrasting semantically similar examples against dissimilar ones (c.f § 2.2). A simple way to obtain a semantically similar example is to make minimal changes to it. In contrast to images, where simple transformations such as rotation, clipping, and color distortion can generate semantically similar examples, deleting or replacing a random word in a sentence can drastically change its meaning (Schlegel et al., 2021). Therefore, it is crucial to carefully select positive and negative examples for contrastive learning in NLP.

4.1.1 Surface Level

To create a sentence that carries the same meaning as another, one can modify the words or characters in the text. Recent research (Wang et al., 2022; Liu et al., 2021; Wu et al., 2022d) suggests certain transformations that preserve the semantic meaning. Wang et al. (2022) propose randomly flipping the case of some tokens, while Liu et al. (2021) mask spans of tokens to get positive instances, and Wu et al. (2022d) suggest to repeat certain words or subwords. Besides generating positive instances, these transformations help in fixing certain biases in representations generated by transformers. For example, Jiang et al. (2022a) found that avoiding high-frequency tokens can result in better sentence representations, and transformations that mask them out while learning sentence representations can improve its quality.

However, altering the surface characteristics of sentences can lead to models relying on shortcuts rather than learning semantics (Du et al., 2021). To address this issue, Wu et al. (2022a) propose the use of multiple augmentation strategies rather than a single transformation. They use shuffling, repeating, and dropping words as transformation strategies to improve model robustness. Additionally, they implement mechanisms to enhance learning from multiple positive examples.

4.1.2 Model Level

Another approach to generating positive examples is by leveraging the distinctive characteristics of the backbone model utilized in contrastive learning. These characteristics might be architectural choices, or using representation from certain components of the model.

Dropout is a regularization technique used in deep learning to prevent overfitting of a model. During training, some neurons in the layer are randomly deactivated, resulting in slightly different representations when the same training instance is passed through the model multiple times. These different representations can be used as positive examples for sentence representations. Recent studies such as Gao et al. (2021) have demonstrated the effectiveness of dropout as an augmentation strategy. Several other works have also incorporated this technique and improved upon it: promoting decorrelation between different dimensions (Klein and Nabi, 2022) and adding dropout in the transformation arsenal (Wu et al., 2022a,d).

Specific components of language models can be trained to generate semantically similar representations. One example is the use of prefix modules (Li and Liang, 2021), which are small, trainable modules added to a pretrained language model. Wang and Lu (2022) attach two prefix modules to the siamese bert network (c.f § 2) – one each for the two branches – and train them on NLI data. This enables the prefix modules to understand the nuances of the difference between representations. The authors show that representations from the two modules for the same sentence can then be used as positives.

4.1.3 Representation Level

Examining the latent representation of sentences generated by a model yields a valuable benefit. In this scenario, one can discover positive examples by exploring the representation space. These approaches offer the distinct advantage of obviating the need for any data augmentation.

Although BERT's `[CLS]` representation is commonly used as a sentence representation, it has been shown to be ineffective (Reimers and Gurevych, 2019). In fact, Kim et al. (2021) demonstrated that the various layers of BERT have differing levels of performance on the STS dataset. To address this issue, they propose reusing the intermediate BERT representations as positive examples. In contrast, Zhang et al. (2022a) identify the k-nearest neighbors of a sentence representation as positives.

4.1.4 Alternative Methods

Researchers have explored various other methods for obtaining positive samples for unsupervised sentence representations. One option is weak supervision: using spans from the same document (Giorgi et al., 2021), employing related entities (Nishikawa et al., 2022), and utilizing tweets and retweets-with-quotes (Di Giovanni and Brambilla, 2021). On the other hand, dialogue turns can be used as semantically related pairs of text for learning sentence representations (Zhou et al., 2022b).

Other approaches use the capability of large language models to perform tasks based on instructions—a technique called "prompting". Researchers have used prompts to obtain better sentence representations, as demonstrated in studies such as Jiang et al. (2022a), which employs the *"[X] means [MASK]"* prompt to extract sentence representations from the representation of the *"[MASK]"* token in a sentence. Another study by (Zeng et al., 2022) combines prompt-derived sentence representations with contrastive learning to improve the quality of the representations.

4.2 Alternative Loss and Objectives

In § 2 we discuss Contrastive loss, which is widely used in machine learning. However, this loss suffers from several limitations: for instance it only considers binary relationships between instances and lacks a mechanism to incorporate "hard negatives" (negatives that are difficult to distinguish from positive examples). To overcome these drawbacks, researchers have explored various strategies:

Supplementary Losses: used in addition to contrastive losses. These include: *(1)* hinge loss (Jiang et al., 2022b), which enhances discrimination between positive and negative pairs; *(2)* losses for reconstructing the original sentence from its representation to better capture sentence semantics (Wu et al., 2022b) ; *(3)* a loss to identify masked words and improve sensitivity to meaningless semantic transformations (Chuang et al., 2022); and *(4)* a loss to minimize redundant information in transformations by minimizing entropy (Chen et al., 2022a).

Modified Contrastive Loss: modifies the original contrastive loss to overcome drawbacks. Wu et al. (2022c) proposed an additional term that incorporates random noise from a Gaussian distribution as negative instances. Also, Zhang et al. (2022d) introduced two losses, angular loss and margin-based triplet loss, to address the intricacies of similarity between pairs of examples.

Different Loss: move away from contrastive loss to use a different loss function. For instance, Zhang et al. (2020) maximize the mutual information between a local and a global representation of a sentence. Min et al. (2021) identify an alternative sub-manifold within the sentence representation space that considers the geometric structure of sentences. Other objectives to learn sentence representations include disentangling the syntax and semantics from the representation (Huang et al., 2021a), generating important phrases from sentences instead of using contrastive learning (Wu and Zhao, 2022), or using sentence representation as a strong inductive bias to perform Masked Language Modeling (Yang et al., 2021).

4.3 Better Negative Sampling

The efficacy of contrastive learning hinges on the quality of negative samples used during training. While most methods prioritize selecting positive samples that bear similarity to the query text, it's equally crucial to include hard negatives that are dissimilar to the query text and pose a challenge for the model to classify. Failure to do so leads to a gradual diminution of the loss gradients, impeding the learning of useful representations (Zhang et al., 2022c). Additionally, using an adequate number of negative samples is also imperative for effective learning (Cao et al., 2022).

Given the importance of incorporating hard negatives, several innovative strategies have emerged. Researchers have found that mixed-negatives—a combination of representations of a positive and a randomly chosen negative—serve as an excellent hard negative representation (Zhang et al., 2022c). Similarly, Zhou et al. (2022a) leveraged noise from a uniform Gaussian distribution to foster uniformity in the learned representation space—a metric to assess learned sentence representation. To further refine their approach, they also implemented techniques to identify and penalize false negative instances, where similarity scores with the positives exceed a threshold.

4.4 Post-Processing

Ethayarajh (2019) suggest that the out-of-the-box representations from LLMs are not effective sentence representations. Consequently, several efforts have addressed this issue.

Almarwani et al. (2019) utilize the Discrete Cosine Transform, a widely used technique in signal processing, to condense word vectors into fixed-length vectors. This approach has demonstrated its effectiveness in capturing both syntax and semantics. Similarly, Li et al. (2020) employ normalizing flows to convert BERT's token representations into a Gaussian distribution, while Huang et al. (2021b) propose a simpler 'whitening' technique that enhances out-of-the-box sentence representations from LLMs by transforming the mean and covariance matrix of the sentence vectors.

5 Other Approaches

Multimodal: Human experiences are complex and involve multiple sensory modalities. Thus, it is beneficial to incorporate multiple modalities in learning sentence representations. Researchers have explored different approaches to use images to learn sentence representations: using contrastive loss that utilizes both images and text (Zhang et al., 2022b); optimizing a loss each for visual and textual representation (Jian et al., 2022); grounding text into image (Bordes et al., 2019). Other modalities like audio and video are yet to be incorporated in learning sentence representation.

Computer Vision Inspired: Momentum encoder, introduced by He et al. (2020), improves training stability in contrastive learning. It utilizes a queue of representations from previous batches as negatives for the current batch, decoupling batch size from the learning process. Several studies have integrated momentum encoder into sentence representation learning, leading to enhanced performance (Cao et al., 2022; Wu et al., 2022a,d; Tan et al., 2022).

Another popular technique, Bootstrap Your Own Latent (BYOL) (Grill et al., 2020), is a self-supervised learning method that dispenses with negative samples. It trains a neural network to predict a set of 'target' representations from an input data point, given an 'online' representation of the same data point. BYOL employs a contrastive loss

Name	Supervision	SentEval?	Base Model	Component	Average
Chen et al. (2022b)	Supervised	No	t5	Model	85.19
Gao et al. (2021)	Unsupervised	Yes	roberta-large	Data	83.76
Ni et al. (2022a)	Supervised	Yes	t5	Model	83.34
Wang et al. (2022)	Unsupervised	No	roberta-large	Data	80.84
Zhang et al. (2022d)	Unsupervised	Yes	sbert-large	Loss	80.69
Wang and Lu (2022)	Unsupervised	No	bert-base	Data	80.61
Wu et al. (2022b)	Unsupervised	Yes	bert-large	Loss	80.18
Wu et al. (2022a)	Unsupervised	Yes	bert-large	Data	79.94
Kim et al. (2021)	Unsupervised	Yes	roberta-large	Data	79.76
Wu et al. (2022d)	Unsupervised	Yes	roberta-large	Data	79.45
Zhou et al. (2022a)	Unsupervised	Yes	roberta-large	Data	79.30
Wu et al. (2022c)	Unsupervised	No	roberta-large	Loss	79.21
Jiang et al. (2022a)	Unsupervised	No	roberta-base	Loss	79.15
Cao et al. (2022)	Unsupervised	Yes	bert-large	Data	79.13
Zhang et al. (2022a)	Unsupervised	No	roberta-large	Data	79.04
Zhang et al. (2022c)	Unsupervised	Yes	bert-large	Data	78.8
Min et al. (2021)	Unsupervised	Yes	bert-large	-	78.79
Chuang et al. (2022)	Unsupervised	Yes	bert-base	Loss	78.49
Jiang et al. (2022b)	Unsupervised	Yes	bert-base	Loss	78.49
Chen et al. (2022a)	Unsupervised	Yes	roberta-large	Loss	78.08
Wu et al. (2022a)	Unsupervised	Yes	roberta-base	Data	77.91
Cheng (2021)	Supervised	No	roberta-large	-	77.47
Nishikawa et al. (2022)	Unsupervised	No	bert-base	Data	77.00
Reimers and Gurevych (2019)	Supervised	Yes	roberta-large	Transform Loss	76.68
Liu et al. (2021)	Unsupervised	No	roberta-base	Data	76.40
Wu and Zhao (2022)	Unsupervised	No	bert-base	Loss	76.16
Schick and Schütze (2021)	Unsupervised	No	roberta-base	Data	75.20
Klein and Nabi (2022)	Unsupervised	Yes	bert-base	Data	74.19
Huang et al. (2021b)	Unsupervised	No	LaBSE	Transform	71.71
Giorgi et al. (2021)	Unsupervised	Yes	roberta-base	Data	69.99
Yang et al. (2021)	Unsupervised	No	bert-base	Loss	67.22
Zhang et al. (2020)	Unsupervised	Yes	bert-base	Loss	66.58
Li et al. (2020)	Unsupervised	No	bert-base	Data	66.55

Table 1: Comparison of methods. SentEval indicates whether the work benchmarks against SentEval (Conneau and Kiela, 2018), component indicates the component from Figure 1 that the work targets, and average shows the average score on the STS benchmark.

function to encourage similarity between the online and target representations. An advantage of BYOL is the elimination of the need for negative samples; instead, it uses augmented versions of the same data point as positive samples. This method has been effectively applied to natural language processing by Zhang et al. (2021).

6 Trends & Analysis

Limited advantages of supervision: Table 1 summarizes all the results. Surprisingly, a simple dropout-based data augmentation technique (Gao et al., 2021) demonstrates superior performance compared to most other methods, including those which use T5, which is trained on billions of tokens (Ni et al., 2022a). Leveraging unsupervised data first to learn sentence representations, followed by supervised training, may be more practical.

Downplaying downstream task evaluation: The neglect of evaluating sentence representations in downstream tasks, as exemplified in Table 1, is noticeable. With LLMs demonstrating remarkable zero-shot performance across various tasks, the utility of sentence representations for tasks beyond semantic similarity and retrieval seems to dwindle. Nevertheless, recent research underscores how sentence representations can enhance few-shot text classification performance (Tunstall et al., 2022). The ongoing debate regarding their practicality remains unsettled, and further exploration of diverse applications is essential.

Data-centric innovations: Most innovations in this field focus on improving the DATA aspect, including obtaining better positives or negatives and generating data using large language models

(Schick and Schütze, 2021; Chen et al., 2022b). While generative models like T5 can boost performance, other LLMs like ChatGPT can bring additional benefits because of their scale.

Keeping up with LLMs: We have identified several noteworthy endeavors using massive language models with billions of parameters for sentence representations. SGPT (Muennighoff, 2022) has successfully trained an open-source GPT decoder-only model on the SNLI and MNLI datasets, surpassing OpenAI's 175B parameter model. Additionally, GTR (Ni et al., 2022b) examined scaling laws, revealing larger T5 models have better performance. Nonetheless, recent developments such as GTE (Li et al., 2023) and BGE (Xiao et al., 2023) highlight that a collection of high-quality datasets for contrastive training can yield significantly enhanced results compared to just using bigger models.

7 Challenges

Practical Applications and the rise of Tools: Sentence representations are commonly employed for sentence retrieval in practical applications, as evidenced by the increasing number of benchmarks (Thakur et al., 2021b). However, their utility extends beyond retrieval, as demonstrated by recent work (Schuster et al., 2022), which leverages sentence representations for identifying documents that share a similar stance on a topic and for isolating documents that diverge from the consensus.

The increasing use of sentence representations in practical applications such as retrieval requires efficient storage and indexing solutions that enable fast retrieval. These solutions are commonly referred to as vector databases and include popular options such as Pinecone[2] and Milvus.[3] These vector databases can be integrated with other frameworks such as LangChain that facilitate the development of applications using LLMs.

Adapting to different Domains: Research has shown that sentence representations learned in one domain may not accurately capture the semantic meaning of sentences in another domain (Jiang et al., 2022b; Thakur et al., 2021a). Some solutions have been proposed in the literature, such as generating queries using a pretrained T5 model on a paragraph from the target domain, or using a pretrained cross-encoder to label the query and paragraph, or using a denoising objective (Wang et al., 2021). Nonetheless, training models that work well across domains remains challenging.

Cross-lingual Sentence Representations: Creating sentence representations that can be used across languages, especially those with limited annotated data, poses a significant challenge. New solutions for cross-lingual retrieval are being developed and deployed for real-world use cases.[4] Many scholarly works (Nishikawa et al., 2022; Feng et al., 2022; Wieting et al., 2020) have addressed cross-lingual sentence representation learning in recent times, but they require aligned data between languages, which is hard to obtain.

How Universal are Sentence Representations? The original purpose of sentence representations was to serve as a versatile tool for various NLP tasks. One prominent effort to evaluate the universality of sentence representations was the SentEval task (Conneau and Kiela, 2018), which tested the representations' performance on text classification, natural language inference, and semantic text similarity tasks. However, many recent works on sentence representation tend to emphasize their effectiveness on semantic text similarity datasets (Table 1). This shift raises questions about the universal nature of these representations—are sentence representations useful only for retrieval, or do they indeed have other applications? Such questions are put back into spotlight by recent benchmarks such as MTEB (Muennighoff et al., 2022).

8 Conclusions

This survey offers an overview of sentence representations, presenting a taxonomy of methods. While major innovations focused on obtaining better quality data for contrastive learning, modern advances in generative technologies can accelerate the automatic generation of supervised data at low cost. Although LLMs play a crucial role in informing the advancement of sentence representations, further enhancements in sentence representation learning are necessary to personalize current LLMs to achieve tailored results. We highlighted that better multilingual and multidomain sentence representations are needed, now that LLMs are being deployed in different domains at a rapid pace. We hope that this survey can accelerate advances in sentence representation learning.

[2] https://www.pinecone.io/

[3] https://milvus.io/

[4] https://txt.cohere.com/multilingual/

9 Limitations

While we have made an effort to encompass a comprehensive range of literature on sentence representations, it is possible that certain papers may have been inadvertently excluded from our literature review. Additionally, we acknowledge that our approach assumes the majority of methods primarily focus on sentences or a limited number of tokens, typically within a few hundred. However, it is important to note that representation learning for documents or longer contexts—an active area of research—utilizes similar techniques. This survey does not cover those specific areas, which may warrant further attention.

References

Nada Almarwani, Hanan Aldarmaki, and Mona Diab. 2019. Efficient sentence embedding using discrete cosine transform. 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 3672–3678, Hong Kong, China. Association for Computational Linguistics.

Patrick Bordes, Eloi Zablocki, Laure Soulier, Benjamin Piwowarski, and Patrick Gallinari. 2019. Incorporating visual semantics into sentence representations within a grounded space. 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 696–707, Hong Kong, China. Association for Computational Linguistics.

Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. 2015. A large annotated corpus for learning natural language inference. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 632–642, Lisbon, Portugal. Association for Computational Linguistics.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.

Rui Cao, Yihao Wang, Yuxin Liang, Ling Gao, Jie Zheng, Jie Ren, and Zheng Wang. 2022. Exploring the impact of negative samples of contrastive learning: A case study of sentence embedding. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 3138–3152, Dublin, Ireland. Association for Computational Linguistics.

Shaobin Chen, Jie Zhou, Yuling Sun, and Liang He. 2022a. An information minimization based contrastive learning model for unsupervised sentence embeddings learning. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 4821–4831, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Yiming Chen, Yan Zhang, Bin Wang, Zuozhu Liu, and Haizhou Li. 2022b. Generate, discriminate and contrast: A semi-supervised sentence representation learning framework. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 8150–8161, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Xingyi Cheng. 2021. Dual-view distilled bert for sentence embedding. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '21, page 2151–2155, New York, NY, USA. Association for Computing Machinery.

Yung-Sung Chuang, Rumen Dangovski, Hongyin Luo, Yang Zhang, Shiyu Chang, Marin Soljacic, Shang-Wen Li, Scott Yih, Yoon Kim, and James Glass. 2022. DiffCSE: Difference-based contrastive learning for sentence embeddings. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4207–4218, Seattle, United States. Association for Computational Linguistics.

Alexis Conneau and Douwe Kiela. 2018. SentEval: An evaluation toolkit for universal sentence representations. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Alexis Conneau, Douwe Kiela, Holger Schwenk, Loïc Barrault, and Antoine Bordes. 2017. Supervised learning of universal sentence representations from natural language inference data. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 670–680, Copenhagen, Denmark. Association for Computational Linguistics.

Ido Dagan, Dan Roth, Mark Sammons, and Fabio Massimo Zanzotto. 2013. Recognizing textual entailment: Models and applications. *Synthesis Lectures on Human Language Technologies*, 6(4):1–220.

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.

Marco Di Giovanni and Marco Brambilla. 2021. Exploiting Twitter as source of large corpora of weakly similar pairs for semantic sentence embeddings. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9902–9910, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Mengnan Du, Varun Manjunatha, Rajiv Jain, Ruchi Deshpande, Franck Dernoncourt, Jiuxiang Gu, Tong Sun, and Xia Hu. 2021. Towards interpreting and mitigating shortcut learning behavior of NLU models. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 915–929, Online. Association for Computational Linguistics.

Kawin Ethayarajh. 2019. How contextual are contextualized word representations? Comparing the geometry of BERT, ELMo, and GPT-2 embeddings. 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 55–65, Hong Kong, China. Association for Computational Linguistics.

Fangxiaoyu Feng, Yinfei Yang, Daniel Cer, Naveen Arivazhagan, and Wei Wang. 2022. Language-agnostic BERT sentence embedding. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 878–891, Dublin, Ireland. Association for Computational Linguistics.

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.

John Giorgi, Osvald Nitski, Bo Wang, and Gary Bader. 2021. DeCLUTR: Deep contrastive learning for unsupervised textual representations. 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 879–895, Online. Association for Computational Linguistics.

Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H. Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Daniel Guo, Mohammad Gheshlaghi Azar, Bilal Piot, Koray Kavukcuoglu, Rémi Munos, and Michal Valko. 2020. Bootstrap your own latent a new approach to self-supervised learning. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, Red Hook, NY, USA. Curran Associates Inc.

Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. 2020. Momentum contrast for unsupervised visual representation learning. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9726–9735.

Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. 2015. Distilling the knowledge in a neural network. *arXiv*, (1503.02531).

James Y. Huang, Kuan-Hao Huang, and Kai-Wei Chang. 2021a. Disentangling semantics and syntax in sentence embeddings with pre-trained language models. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1372–1379, Online. Association for Computational Linguistics.

Junjie Huang, Duyu Tang, Wanjun Zhong, Shuai Lu, Linjun Shou, Ming Gong, Daxin Jiang, and Nan Duan. 2021b. WhiteningBERT: An easy unsupervised sentence embedding approach. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 238–244, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Yiren Jian, Chongyang Gao, and Soroush Vosoughi. 2022. Non-linguistic supervision for contrastive learning of sentence embeddings. In *Advances in Neural Information Processing Systems*.

Ting Jiang, Jian Jiao, Shaohan Huang, Zihan Zhang, Deqing Wang, Fuzhen Zhuang, Furu Wei, Haizhen Huang, Denvy Deng, and Qi Zhang. 2022a. PromptBERT: Improving BERT sentence embeddings with prompts. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 8826–8837, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Yuxin Jiang, Linhan Zhang, and Wei Wang. 2022b. Improved universal sentence embeddings with prompt-based contrastive learning and energy-based learning. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 3021–3035, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Omar Khattab, Arnav Singhvi, Paridhi Maheshwari, Zhiyuan Zhang, Keshav Santhanam, Sri Vardhamanan, Saiful Haq, Ashutosh Sharma, Thomas T. Joshi, Hanna Moazam, Heather Miller, Matei Zaharia, and Christopher Potts. 2023. Dspy: Compiling declarative language model calls into self-improving pipelines. *arXiv preprint arXiv:2310.03714*.

Taeuk Kim, Kang Min Yoo, and Sang-goo Lee. 2021. Self-guided contrastive learning for BERT sentence representations. 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 2528–2540, Online. Association for Computational Linguistics.

Ryan Kiros, Yukun Zhu, Ruslan Salakhutdinov, Richard S. Zemel, Antonio Torralba, Raquel Urtasun, and Sanja Fidler. 2015. Skip-thought vectors. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*, page 3294–3302, Cambridge, MA, USA. MIT Press.

Tassilo Klein and Moin Nabi. 2022. SCD: Self-contrastive decorrelation of sentence embeddings. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 394–400, Dublin, Ireland. Association for Computational Linguistics.

Bohan Li, Hao Zhou, Junxian He, Mingxuan Wang, Yiming Yang, and Lei Li. 2020. On the sentence embeddings from pre-trained language models. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 9119–9130, Online. Association for Computational Linguistics.

Xiang Lisa Li and Percy Liang. 2021. Prefix-tuning: Optimizing continuous prompts for generation. 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 4582–4597, Online. Association for Computational Linguistics.

Zehan Li, Xin Zhang, Yanzhao Zhang, Dingkun Long, Pengjun Xie, and Meishan Zhang. 2023. Towards general text embeddings with multi-stage contrastive learning.

Fangyu Liu, Ivan Vulić, Anna Korhonen, and Nigel Collier. 2021. Fast, effective, and self-supervised: Transforming masked language models into universal lexical and sentence encoders. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 1442–1459, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Changrong Min, Yonghe Chu, Liang Yang, Bo Xu, and Hongfei Lin. 2021. Locality preserving sentence encoding. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 3050–3060, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Niklas Muennighoff. 2022. Sgpt: Gpt sentence embeddings for semantic search. *arXiv preprint arXiv:2202.08904*.

Niklas Muennighoff, Nouamane Tazi, Loïc Magne, and Nils Reimers. 2022. Mteb: Massive text embedding benchmark. *arXiv preprint arXiv:2210.07316*.

Jianmo Ni, Gustavo Hernandez Abrego, Noah Constant, Ji Ma, Keith Hall, Daniel Cer, and Yinfei Yang. 2022a. Sentence-T5: Scalable sentence encoders from pre-trained text-to-text models. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 1864–1874, Dublin, Ireland. Association for Computational Linguistics.

Jianmo Ni, Chen Qu, Jing Lu, Zhuyun Dai, Gustavo Hernandez Abrego, Ji Ma, Vincent Zhao, Yi Luan, Keith Hall, Ming-Wei Chang, and Yinfei Yang. 2022b. Large dual encoders are generalizable retrievers. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 9844–9855, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Yixin Nie, Adina Williams, Emily Dinan, Mohit Bansal, Jason Weston, and Douwe Kiela. 2020. Adversarial NLI: A new benchmark for natural language understanding. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4885–4901, Online. Association for Computational Linguistics.

Sosuke Nishikawa, Ryokan Ri, Ikuya Yamada, Yoshimasa Tsuruoka, and Isao Echizen. 2022. EASE: Entity-aware contrastive learning of sentence embedding. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3870–3885, Seattle, United States. Association for Computational Linguistics.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar. Association for Computational Linguistics.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67.

Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. 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 3982–3992, Hong Kong, China. Association for Computational Linguistics.

Timo Schick and Hinrich Schütze. 2021. Generating datasets with pretrained language models. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6943–

6951, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Viktor Schlegel, Goran Nenadic, and Riza Batista-Navarro. 2021. Semantics altering modifications for evaluating comprehension in machine reading. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 13762–13770.

Tal Schuster, Sihao Chen, Senaka Buthpitiya, Alex Fabrikant, and Donald Metzler. 2022. Stretching sentence-pair NLI models to reason over long documents and clusters. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 394–412, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Haochen Tan, Wei Shao, Han Wu, Ke Yang, and Linqi Song. 2022. A sentence is worth 128 pseudo tokens: A semantic-aware contrastive learning framework for sentence embeddings. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 246–256, Dublin, Ireland. Association for Computational Linguistics.

Nandan Thakur, Nils Reimers, Johannes Daxenberger, and Iryna Gurevych. 2021a. Augmented SBERT: Data augmentation method for improving bi-encoders for pairwise sentence scoring tasks. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 296–310, Online. Association for Computational Linguistics.

Nandan Thakur, Nils Reimers, Andreas Rücklé, Abhishek Srivastava, and Iryna Gurevych. 2021b. BEIR: A heterogeneous benchmark for zero-shot evaluation of information retrieval models. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*.

Lewis Tunstall, Nils Reimers, Unso Eun Seo Jo, Luke Bates, Daniel Korat, Moshe Wasserblat, and Oren Pereg. 2022. Efficient few-shot learning without prompts. *arXiv*, (2209.11055).

Kexin Wang, Nils Reimers, and Iryna Gurevych. 2021. TSDAE: Using transformer-based sequential denoising auto-encoderfor unsupervised sentence embedding learning. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 671–688, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Tianduo Wang and Wei Lu. 2022. Differentiable data augmentation for contrastive sentence representation learning. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 7640–7653, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Wei Wang, Liangzhu Ge, Jingqiao Zhang, and Cheng Yang. 2022. Improving contrastive learning of sentence embeddings with case-augmented positives and retrieved negatives. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '22, page 2159–2165, New York, NY, USA. Association for Computing Machinery.

John Wieting, Graham Neubig, and Taylor Berg-Kirkpatrick. 2020. A bilingual generative transformer for semantic sentence embedding. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1581–1594, Online. Association for Computational Linguistics.

Adina Williams, Nikita Nangia, and Samuel Bowman. 2018. A broad-coverage challenge corpus for sentence understanding through inference. 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 1112–1122, New Orleans, Louisiana. Association for Computational Linguistics.

BigScience Workshop. 2023. Bloom: A 176b-parameter open-access multilingual language model.

Bohong Wu and Hai Zhao. 2022. Sentence representation learning with generative objective rather than contrastive objective. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 3356–3368, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Qiyu Wu, Chongyang Tao, Tao Shen, Can Xu, Xiubo Geng, and Daxin Jiang. 2022a. PCL: Peer-contrastive learning with diverse augmentations for unsupervised sentence embeddings. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 12052–12066, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Xing Wu, Chaochen Gao, Zijia Lin, Jizhong Han, Zhongyuan Wang, and Songlin Hu. 2022b. InfoCSE: Information-aggregated contrastive learning of sentence embeddings. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 3060–3070, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Xing Wu, Chaochen Gao, Yipeng Su, Jizhong Han, Zhongyuan Wang, and Songlin Hu. 2022c. Smoothed contrastive learning for unsupervised sentence embedding. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 4902–4906, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Xing Wu, Chaochen Gao, Liangjun Zang, Jizhong Han, Zhongyuan Wang, and Songlin Hu. 2022d. ESimCSE: Enhanced sample building method for contrastive learning of unsupervised sentence embedding. In *Proceedings of the 29th International Con-*

ference on Computational Linguistics, pages 3898–3907, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Shitao Xiao, Zheng Liu, Peitian Zhang, and Niklas Muennighoff. 2023. C-pack: Packaged resources to advance general chinese embedding.

Ziyi Yang, Yinfei Yang, Daniel Cer, Jax Law, and Eric Darve. 2021. Universal sentence representation learning with conditional masked language model. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6216–6228, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Jiali Zeng, Yongjing Yin, Yufan Jiang, Shuangzhi Wu, and Yunbo Cao. 2022. Contrastive learning with prompt-derived virtual semantic prototypes for unsupervised sentence embedding. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 7042–7053, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Dejiao Zhang, Wei Xiao, Henghui Zhu, Xiaofei Ma, and Andrew Arnold. 2022a. Virtual augmentation supported contrastive learning of sentence representations. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 864–876, Dublin, Ireland. Association for Computational Linguistics.

Miaoran Zhang, Marius Mosbach, David Adelani, Michael Hedderich, and Dietrich Klakow. 2022b. MCSE: Multimodal contrastive learning of sentence embeddings. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5959–5969, Seattle, United States. Association for Computational Linguistics.

Yan Zhang, Ruidan He, Zuozhu Liu, Lidong Bing, and Haizhou Li. 2021. Bootstrapped unsupervised sentence representation 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 5168–5180, Online. Association for Computational Linguistics.

Yan Zhang, Ruidan He, Zuozhu Liu, Kwan Hui Lim, and Lidong Bing. 2020. An unsupervised sentence embedding method by mutual information maximization. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1601–1610, Online. Association for Computational Linguistics.

Yanzhao Zhang, Richong Zhang, Samuel Mensah, Xudong Liu, and Yongyi Mao. 2022c. Unsupervised sentence representation via contrastive learning with mixing negatives. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 11730–11738.

Yuhao Zhang, Hongji Zhu, Yongliang Wang, Nan Xu, Xiaobo Li, and Binqiang Zhao. 2022d. A contrastive framework for learning sentence representations from pairwise and triple-wise perspective in angular space. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 4892–4903, Dublin, Ireland. Association for Computational Linguistics.

Kun Zhou, Beichen Zhang, Xin Zhao, and Ji-Rong Wen. 2022a. Debiased contrastive learning of unsupervised sentence representations. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6120–6130, Dublin, Ireland. Association for Computational Linguistics.

Zhihan Zhou, Dejiao Zhang, Wei Xiao, Nicholas Dingwall, Xiaofei Ma, Andrew Arnold, and Bing Xiang. 2022b. Learning dialogue representations from consecutive utterances. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 754–768, Seattle, United States. Association for Computational Linguistics.