Contrastive Data and Learning for Natural Language Processing

Rui Zhang
Penn State University
rmz5227@psu.edu

Yangfeng Ji
University of Virginia
yangfeng@virginia.edu

Yue Zhang
Westlake University
yue.zhang@wias.org.cn

Rebecca J. Passonneau
Penn State University
rjp49@psu.edu

1 Brief Description

Current NLP models heavily rely on effective representation learning algorithms. Contrastive learning is one such technique to learn an embedding space such that similar data sample pairs have close representations while dissimilar samples stay far apart from each other. It can be used in supervised or unsupervised settings using different loss functions to produce task-specific or general-purpose representations. While it has originally enabled the success for vision tasks, recent years have seen a growing number of publications in contrastive NLP as shown in Figure 1. This first line of works not only delivers promising performance improvements in various NLP tasks, but also provides desired characteristics such as task-agnostic sentence representation, faithful text generation, data-efficient learning in zero-shot and few-shot settings, interpretability and explainability.

In this tutorial, we aim to provide a gentle introduction to the fundamentals of contrastive learning approaches and the theory behind them. We then survey the benefits and the best practices of contrastive learning for various downstream NLP applications including Text Classification, Question Answering, Summarization, Text Generation, Interpretability and Explainability, Commonsense Knowledge and Reasoning, Vision-and-Language. This tutorial intends to help researchers in the NLP and computational linguistics community to understand this emerging topic and promote future research directions of using contrastive learning for NLP applications.[1]

Type of Tutorial: Cutting-edge As an emerging approach, recent years have seen a growing number of NLP papers using contrastive learning (Figure 1). Contrastive learning still has a huge potential in other applications and challenges, and

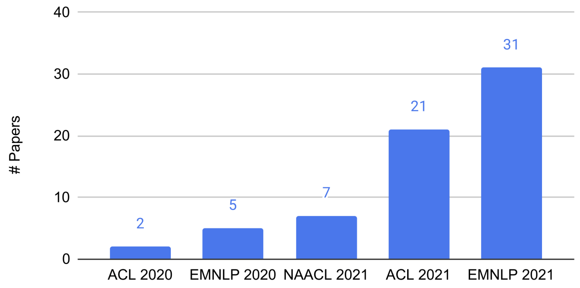

Figure 1: The number of papers in recent *ACL conferences with "contrastive learning" in the title. We anticipate there will be even more papers in 2022.

we anticipate there will be even more papers in the next year before this tutorial. However, there is no tutorial yet that systematically introduces contrastive learning and its application to NLP.

Target Audience and Expected Background This tutorial is targeted at a broad and general audience who is interested using contrastive learning for NLP tasks. The tutorial will be self-contained. The expected prerequisite only includes basic a understanding of machine learning concepts such as classification, loss functions, and gradient-based optimization. We also expect the audience to be familiar with the definition of different NLP tasks.

2 Tutorial Structure and Content

This tutorial first gives an introduction to the foundation of contrastive learning and then reviews the NLP application of contrastive learning. Our tutorial covers both **contrastive data augmentation for NLP** and **contrastive representation learning for NLP**. The former focuses on the data side: how we can create contrastive data examples. This is useful not only for contrastive learning signals, but also for many other reasons such as evaluating model behaviors, augmenting data for low-resource training, producing contrastive explanation, promoting faithful text generation. The latter focuses

[1] Tutorial materials are available at https://contrastive-nlp-tutorial.github.io/

on the learning algorithm side: how we can use contrastive learning broadly in different NLP tasks. Here is the outline with an estimated schedule.

Part 1: Foundations of Contrastive Learning (60 min)

- Contrastive Learning Objectives (15 min)
- Contrastive Data Sampling and Augmentation Strategies (15 min)
- Successful Applications (15 min)
- Analysis of Contrastive Learning (15 min)

Part 2: Contrastive Learning for NLP (90 min)

- Contrastive Learning in NLP Tasks (30 min)
- Task-agnostics Representation (15 min)
- Faithful Text Generation (15 min)
- Data-efficient Learning (15 min)
- Interpretability and Explainability (15 min)

Part 3: Lessons Learned, Practical Advice, and Future Directions (30 min)

- Lessons Learned (10 min)
- Practical Advice (10 min)
- Future Directions (10 min)

The following subsections give more details with reference papers for each part.

2.1 Foundations of Contrastive Learning

In the first part, we will provide a brief overview of contrastive learning foundations and introduce the most well-known contrastive learning approaches. We start with different contrastive learning objectives including Contrastive Loss (Chopra et al., 2005), Triplet Loss (Schroff et al., 2015), Lifted Structured Loss (Oh Song et al., 2016), N-pair Loss (Sohn, 2016), Noise Contrastive Estimation (NCE) (Gutmann and Hyvärinen, 2010), InfoNCE (van den Oord et al., 2018), and Soft-Nearest Neighbors Loss (Salakhutdinov and Hinton, 2007; Frosst et al., 2019). We then overview different sampling strategies to create contrastive pairs including debiased constrastive learning (Chuang et al., 2020), hard negative samples (Robinson et al., 2020), supervised contrastive learning (Chen et al., 2020), and adversarial contrastive learning (Kim et al., 2020). We will also talk about contrastive learning with deep neural networks that have shown great successes in vision and language applications such as word2vec (Mikolov et al., 2013), SimCLR (Chen et al., 2020), SimCSE (Gao et al., 2021b), and CLIP (Radford et al., 2021). We will also discuss work on intriguing analyses of contrastive learning (Tian et al., 2020; Purushwalkam and Gupta, 2020; Xiao et al., 2021).

2.2 Contrastive Learning for NLP

In this part, we will first survey the usage of contrastive learning in different NLP tasks. Later, we will also highlight four characteristics that contrastive learning has demonstrated in addition to the promising performance improvement.

Contrastive learning has shown success in many NLP tasks. We plan cover the following: **Contrastive Data Augmentation for NLP** (Shen et al., 2020; Ye et al., 2021; Qu et al., 2021); **Text Classification** (Pappas and Henderson, 2018; Fang et al., 2020; Kachuee et al., 2020; Suresh and Ong, 2021; Du et al., 2021; Carlsson et al., 2021; Xiong et al., 2021; Qiu et al., 2021; Xu et al., 2021b; Klein and Nabi, 2021); **Sentence Embeddings** (Kim et al., 2021; Zhang et al., 2021a; Sedghamiz et al., 2021) including Quick-Thought (Logeswaran and Lee, 2018),Sentence-BERT (Reimers and Gurevych, 2019), Info-Sentence BERT (Zhang et al., 2020a), SimCSE (Gao et al., 2021b), DeCLUTR (Giorgi et al., 2020), ConSERT (Yan et al., 2021b), DialogueCSE (Liu et al., 2021a). We will also cover discourse analysis (Iter et al., 2020; Kiyomaru and Kurohashi, 2021); **Information Extraction** (Qin et al., 2020; Chen et al., 2021b; Wang et al., 2021d) **Machine Translation** (Pan et al., 2021; Vamvas and Sennrich, 2021); **Question Answering** (Karpukhin et al., 2020; You et al., 2021; Yang et al., 2021b; Yue et al., 2021); **Summarization** (Duan et al., 2019; Liu and Liu, 2021) including faithfulness (Cao and Wang, 2021), summary evaluation (Wu et al., 2020a), multilingual summarization (Wang et al., 2021a), and dialogue summarization (Liu et al., 2021d); **Text Generation** (Chai et al., 2021; Lee et al., 2021b) including logic-consistent text generation (Shu et al., 2021), paraphrase generation (Yang et al., 2021a), grammatical error correction (Cao et al., 2021), dialogue generation (Cai et al., 2020), x-ray report generation (Liu et al., 2021b; Yan et al., 2021a), data-to-text generation (Uehara et al., 2020); **Few-shot Learning** (Liu et al., 2021c; Zhang et al., 2021c; Wang et al., 2021c; Luo et al., 2021; Das et al., 2021); **Language Model Contrastive Pretraining** (Wu et al., 2020b; Gunel et al., 2020; Clark et al., 2020; Yu et al., 2020; Rethmeier and Augenstein, 2020, 2021; Meng et al., 2021; Li et al., 2021b); **Interpretability and Explainability** (Gardner et al., 2020; Liang et al., 2020; Ross et al., 2020; Chen et al., 2021a; Jacovi et al., 2021); **Commonsense Knowledge and Reasoning** (Klein and Nabi, 2020; Paranjape et al., 2021; Li et al., 2021a); **Vision-and-Language** (Zhang et al., 2020b; Li et al., 2020; Dharur et al., 2020; Cui et al., 2020; Radford et al., 2021; Xu et al., 2021a; Jia et al., 2021; Lee et al., 2021a). We will also briefly talk about other applications such as distillation and model compression (Sun et al., 2020), debiasing (Cheng et al., 2021), fact verification (Schuster et al., 2021), short text clustering (Zhang et al., 2021b), out-of-domain detection (Zeng et al., 2021; Zhou and Chen, 2021), robustness (Ma et al., 2021), code representation learning (Jain et al., 2020), active learning (Margatina et al., 2021), knowledge representation learning (Ouyang et al., 2021), adversarial learning (Rim et al., 2021).

In addition to the performance benefit, we highlight that contrastive learning is particularly interesting for NLP because it offers four advantages:

Task-agnostic Sentence Representation As a representation learning approach, contrastive learning has demonstrated its effectiveness to learn task-agnostic sentence embeddings that can be applied across different tasks. Such progress enables efficient encoding of sentences to support large-scale semantic similarity comparison, clustering, and information retrieval via semantic search. The most successful framework is Sentence-BERT (Reimers and Gurevych, 2019) that uses siamese networks with triplet loss to learn sentence embeddings based on cosine similarity. Another example is CERT (Fang et al., 2020) that employs contrastive self-supervised learning at the sentence level with back-translation data augmentation. It outperforms BERT on 7 out of 11 natural language understanding tasks on the GLUE benchmark. Later, SimCSE (Gao et al., 2021b) uses both unsupervised denoising objective and supervised natural language inference signals to learn sentence embeddings. It achieves substantial improvements on several standard semantic textual similarity benchmarks.

Faithful and Factual Consistent Text Generation Contrastive learning is also used to improve faithfulness and factuality of data-to-text generation and abstractive summarization, which has been shown a very challenging issue with the pretrained language models that often hallucinate (Kryscinski et al., 2019; Parikh et al., 2020; Maynez et al., 2020). Shu et al. (2021) propose to improve logic-to-text generation models by designing rule-based data augmentation to create contrastive examples to cover variations of logic forms paired with diverse natural language expressions to improve the generalizability. CLIFF (Cao and Wang, 2021) propose to improve faithful and factual consistency for abstractive summarization by contrasting reference summaries as positive training data and automatically generated erroneous summaries as negative training data. Wu et al. (2020a) also propose to use contrastive learning for unsupervised reference-free summary quality evaluation.

Data-efficient Learning Another advantage of contrastive learning is to facilitate data-efficient learning when training data is not abundantly available such as in zero-shot and few-shot settings. CoDA (Qu et al., 2021) is a data augmentation framework that synthesizes contrast-enhanced and diverse examples by integrating multiple transformations over text. CLESS (Rethmeier and Augenstein, 2020) analyze data-efficient pretraining via contrastive self-supervision through pretraining data efficiency, zero to few-shot label efficiency, and long-tail generalization. CONTaiNER (Das et al., 2021) improves few-shot named entity recognition by performing contrastive learning over Gaussian distributions of token embeddings. VideoCLIP (Xu et al., 2021a) uses contrastive pretraining for zero-shot video-text understanding.

Interpretability and Explainability Contrastive learning provides a new way for promoting model interpretability and explainability. Contrast Sets (Gardner et al., 2020) evaluate local decision boundaries of models by manually perturbing the test instances in small but meaningful ways. Jacovi et al. (2021) propose to produce contrastive explanations for classification models by modifying model representation and model behavior based on contrastive reasoning. Paranjape et al. (2021) leverage prompt engineering over pretrained language models to create contrastive explanations for commonsense reasoning tasks.

2.3 Lessons Learned, Practical Advice, and Future Directions

In this part, we will summarize our discussions of existing work with lessons learned and practical advice. We will also envision the future directions of contrastive learning for NLP such as data augmentation quality and efficiency (Wang et al., 2021b), hard negative examples (Zhang and Stratos, 2021), under-explored NLP applications (Li et al., 2021b), large batch size (Gao et al., 2021a).

3 Reading List

We compile the a light reading list for the audience learning before coming to the tutorial:

- SimCLR (Chen et al., 2020)
- CLIP (Radford et al., 2021)
- SimCSE (Gao et al., 2021b)
- Contrast Sets (Gardner et al., 2020)

4 Diversity

Our presenters come from 3 institutions based in the U.S. and China including 3 male and 1 female researchers on different levels of academic seniority. As contrastive learning can be applied broadly, our tutorial spans many different NLP tasks and domains covering Text Classification and Sentence Embeddings, Information Extraction, Machine Translation, Question Answering, Summarization, Text Generation, Few-shot Learning, Interpretability and Explainability, Commonsense Knowledge and Reasoning, Vision-and-Language, Distillation and Model Compression. Therefore, the audience will come from diverse backgrounds.

5 Presenters

Rui Zhang is an Assistant Professor in the Computer Science and Engineering Department of Penn State University and a co-director of the PSU NLP Lab. He is one of the recipients of 2020 Amazon Research Awards. He serves as an Area Chair at NAACL 2021, EMNLP 2021, and NLPCC 2021. He co-organizes the Interactive and Executable Semantic Parsing workshop at EMNLP 2020 which attracted an international audience with 100+ researchers from diverse academic and demographic backgrounds. He has been working on contrastive learning for few-shot named entity recognition (Das et al., 2021) and text generation (Shu et al., 2021). https://ryanzhumich.github.io/

Yangfeng Ji is the William Wulf Assistant Professor in the Department of Computer Science at the University of Virginia, where he leads the Natural Language Processing group. His research interests include building machine learning models for text understanding and generation. His work on entity-driven story generation won an Outstanding Paper Award at NAACL 2018. He is a co-author of an EMNLP 2020 tutorial on The Amazing World of Neural Language Generation. https://yangfengji.net/

Yue Zhang is an Associate Professor at Westlake University. His research interests include NLP and its underlying machine learning algorithms and downstream applications. He was the area chairs of ACL (2017/18/19/20/21), COLING (2014/18), NAACL (2015/19/21), EMNLP (2015/17/19/20), EACL (2021) and IJCAI (2021). He won the best paper awards of IALP (2017), COLING (2018) and best paper honorable mention of SemEval (2020). He is the author of EMNLP 2018 tutorial on Joint models for NLP. https://frcchang.github.io/

Rebecca J. Passonneau is a Professor in the Computer Science and Engineering Department of Penn State University and a co-director of the PSU NLP Lab. Her area of research is natural language processing, with a focus on semantics and pragmatics. Her work is reported in over 130 journal and refereed conference publications. She won a Best Paper Runner Up at NAACL 2010. She is a tutorial co-chair for NAACL 2018. https://sites.psu.edu/becky/

6 Ethics Statement

As contrastive learning often involves data augmentation and manipulation, our ethical consideration mainly focuses on properly dealing with bias in the dataset. As bias and fairness created by contrastive learning algorithms are still under-explored, we will also discuss such relevant topics in the section on future directions.

References

Hengyi Cai, Hongshen Chen, Yonghao Song, Zhuoye Ding, Yongjun Bao, Weipeng Yan, and Xiaofang Zhao. 2020. Group-wise contrastive learning for neural dialogue generation. *arXiv preprint arXiv:2009.07543*.

Hannan Cao, Wenmian Yang, and Hwee Tou Ng. 2021. Grammatical error correction with contrastive learning in low error density domains. *arXiv preprint arXiv:2109.01484*.

Shuyang Cao and Lu Wang. 2021. Cliff: Contrastive learning for improving faithfulness and factuality in abstractive summarization. *arXiv preprint arXiv:2109.09209*.

Fredrik Carlsson, Amaru Cuba Gyllensten, Evangelia Gogoulou, Erik Ylipää Hellqvist, and Magnus Sahlgren. 2021. Semantic re-tuning with contrastive tension. In *International Conference on Learning Representations*.

Yekun Chai, Haidong Zhang, Qiyue Yin, and Junge Zhang. 2021. Counter-contrastive learning for language gans. *EMNLP-Findings 2021*.

Qianglong Chen, Feng Ji, Xiangji Zeng, Feng-Lin Li, Ji Zhang, Haiqing Chen, and Yin Zhang. 2021a. Kace: Generating knowledge aware contrastive explanations for natural language inference. 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 2516–2527.

Tao Chen, Haizhou Shi, Siliang Tang, Zhigang Chen, Fei Wu, and Yueting Zhuang. 2021b. Cil: Contrastive instance learning framework for distantly supervised relation extraction. *arXiv preprint arXiv:2106.10855*.

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

Pengyu Cheng, Weituo Hao, Siyang Yuan, Shijing Si, and Lawrence Carin. 2021. Fairfil: Contrastive neural debiasing method for pretrained text encoders. In *International Conference on Learning Representations*.

Sumit Chopra, Raia Hadsell, and Yann LeCun. 2005. Learning a similarity metric discriminatively, with application to face verification. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 539–546. IEEE.

Ching-Yao Chuang, Joshua Robinson, Lin Yen-Chen, Antonio Torralba, and Stefanie Jegelka. 2020. Debiased contrastive learning. *arXiv preprint arXiv:2007.00224*.

Kevin Clark, Minh-Thang Luong, Quoc V. Le, and Christopher D. Manning. 2020. Pre-training transformers as energy-based cloze models. *CoRR*.

Wanyun Cui, Guangyu Zheng, and Wei Wang. 2020. Unsupervised natural language inference via decoupled multimodal contrastive learning. *arXiv preprint arXiv:2010.08200*.

Sarkar Snigdha Sarathi Das, Arzoo Katiyar, Rebecca J Passonneau, and Rui Zhang. 2021. Container: Few-shot named entity recognition via contrastive learning. *arXiv preprint arXiv:2109.07589*.

Sameer Dharur, Purva Tendulkar, Dhruv Batra, Devi Parikh, and Ramprasaath R Selvaraju. 2020. Sort-ing vqa models: Contrastive gradient learning for improved consistency. *arXiv preprint arXiv:2010.10038*.

Yangkai Du, Tengfei Ma, Lingfei Wu, Fangli Xu, Xuhong Zhang, and Shouling Ji. 2021. Constructing contrastive samples via summarization for text classification with limited annotations. *arXiv preprint arXiv:2104.05094*.

Xiangyu Duan, Hongfei Yu, Mingming Yin, Min Zhang, Weihua Luo, and Yue Zhang. 2019. Contrastive attention mechanism for abstractive sentence summarization. *CoRR*, abs/1910.13114.

Hongchao Fang, Sicheng Wang, Meng Zhou, Jiayuan Ding, and Pengtao Xie. 2020. Cert: Contrastive self-supervised learning for language understanding. *arXiv preprint arXiv:2005.12766*.

Nicholas Frosst, Nicolas Papernot, and Geoffrey Hinton. 2019. Analyzing and improving representations with the soft nearest neighbor loss. In *Proceedings of the 36th International Conference on Machine Learning*.

Luyu Gao, Yunyi Zhang, Jiawei Han, and Jamie Callan. 2021a. Scaling deep contrastive learning batch size under memory limited setup. In *Proceedings of the 6th Workshop on Representation Learning for NLP (RepL4NLP-2021)*, Online. Association for Computational Linguistics.

Tianyu Gao, Xingcheng Yao, and Danqi Chen. 2021b. Simcse: Simple contrastive learning of sentence embeddings. *arXiv preprint arXiv:2104.08821*.

Matt Gardner, Yoav Artzi, Victoria Basmova, Jonathan Berant, Ben Bogin, Sihao Chen, Pradeep Dasigi, Dheeru Dua, Yanai Elazar, Ananth Gottumukkala, et al. 2020. Evaluating models' local decision boundaries via contrast sets. *arXiv preprint arXiv:2004.02709*.

John M Giorgi, Osvald Nitski, Gary D Bader, and Bo Wang. 2020. Declutr: Deep contrastive learning for unsupervised textual representations. *arXiv preprint arXiv:2006.03659*.

Beliz Gunel, Jingfei Du, Alexis Conneau, and Ves Stoyanov. 2020. Supervised contrastive learning for pretrained language model fine-tuning. *arXiv preprint arXiv:2011.01403*.

Michael Gutmann and Aapo Hyvärinen. 2010. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*.

Dan Iter, Kelvin Guu, Larry Lansing, and Dan Jurafsky. 2020. Pretraining with contrastive sentence objectives improves discourse performance of language models. *arXiv preprint arXiv:2005.10389*.

Alon Jacovi, Swabha Swayamdipta, Shauli Ravfogel, Yanai Elazar, Yejin Choi, and Yoav Goldberg. 2021. Contrastive explanations for model interpretability. *arXiv preprint arXiv:2103.01378*.

Paras Jain, Ajay Jain, Tianjun Zhang, Pieter Abbeel, Joseph E Gonzalez, and Ion Stoica. 2020. Contrastive code representation learning. *arXiv preprint arXiv:2007.04973*.

Chao Jia, Yinfei Yang, Ye Xia, Yi-Ting Chen, Zarana Parekh, Hieu Pham, Quoc V. Le, Yun-Hsuan Sung, Zhen Li, and Tom Duerig. 2021. Scaling up visual and vision-language representation learning with noisy text supervision. *CoRR*, abs/2102.05918.

Mohammad Kachuee, Hao Yuan, Young-Bum Kim, and Sungjin Lee. 2020. Self-supervised contrastive learning for efficient user satisfaction prediction in conversational agents. *arXiv preprint arXiv:2010.11230*.

Vladimir Karpukhin, Barlas Oğuz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen-tau Yih. 2020. Dense passage retrieval for open-domain question answering. *arXiv preprint arXiv:2004.04906*.

Minseon Kim, Jihoon Tack, and Sung Ju Hwang. 2020. Adversarial self-supervised contrastive learning. In *Advances in Neural Information Processing Systems*.

Taeuk Kim, Kang Min Yoo, and Sang-goo Lee. 2021. Self-guided contrastive learning for bert sentence representations. *arXiv preprint arXiv:2106.07345*.

Hirokazu Kiyomaru and Sadao Kurohashi. 2021. Contextualized and generalized sentence representations by contrastive self-supervised learning: A case study on discourse relation analysis. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5578–5584.

Tassilo Klein and Moin Nabi. 2020. Contrastive self-supervised learning for commonsense reasoning. *arXiv preprint arXiv:2005.00669*.

Tassilo Klein and Moin Nabi. 2021. Attention-based contrastive learning for winograd schemas. *arXiv preprint arXiv:2109.05108*.

Wojciech Kryscinski, Bryan McCann, Caiming Xiong, and Richard Socher. 2019. Evaluating the factual consistency of abstractive text summarization. *CoRR*, abs/1910.12840.

Hwanhee Lee, Seunghyun Yoon, Franck Dernoncourt, Trung Bui, and Kyomin Jung. 2021a. Umic: An unreferenced metric for image captioning via contrastive learning. *arXiv preprint arXiv:2106.14019*.

Seanie Lee, Dong Bok Lee, and Sung Ju Hwang. 2021b. Contrastive learning with adversarial perturbations for conditional text generation. In *International Conference on Learning Representations*.

Haonan Li, Yeyun Gong, Jian Jiao, Ruofei Zhang, Timothy Baldwin, and Nan Duan. 2021a. Kfcnet: Knowledge filtering and contrastive learning network for generative commonsense reasoning. *arXiv preprint arXiv:2109.06704*.

Shicheng Li, Pengcheng Yang, Fuli Luo, and Jun Xie. 2021b. Multi-granularity contrasting for cross-lingual pre-training. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1708–1717.

Wei Li, Can Gao, Guocheng Niu, Xinyan Xiao, Hao Liu, Jiachen Liu, Hua Wu, and Haifeng Wang. 2020. Unimo: Towards unified-modal understanding and generation via cross-modal contrastive learning. *arXiv preprint arXiv:2012.15409*.

Weixin Liang, James Zou, and Zhou Yu. 2020. Alice: Active learning with contrastive natural language explanations. *arXiv preprint arXiv:2009.10259*.

Che Liu, Rui Wang, Jinghua Liu, Jian Sun, Fei Huang, and Luo Si. 2021a. Dialoguecse: Dialogue-based contrastive learning of sentence embeddings. *arXiv preprint arXiv:2109.12599*.

Fenglin Liu, Changchang Yin, Xian Wu, Shen Ge, Ping Zhang, and Xu Sun. 2021b. Contrastive attention for automatic chest x-ray report generation. *arXiv preprint arXiv:2106.06965*.

Han Liu, Feng Zhang, Xiaotong Zhang, Siyang Zhao, and Xianchao Zhang. 2021c. An explicit-joint and supervised-contrastive learning framework for few-shot intent classification and slot filling. In *EMNLP-Findings 2021*.

Junpeng Liu, Yanyan Zou, Hainan Zhang, Hongshen Chen, Zhuoye Ding, Caixia Yuan, and Xiaojie Wang. 2021d. Topic-aware contrastive learning for abstractive dialogue summarization. *arXiv preprint arXiv:2109.04994*.

Yixin Liu and Pengfei Liu. 2021. Simcls: A simple framework for contrastive learning of abstractive summarization. *arXiv preprint arXiv:2106.01890*.

Lajanugen Logeswaran and Honglak Lee. 2018. An efficient framework for learning sentence representations. *CoRR*, abs/1803.02893.

Ruikun Luo, Guanhuan Huang, and Xiaojun Quan. 2021. Bi-granularity contrastive learning for post-training in few-shot scene. *arXiv preprint arXiv:2106.02327*.

Xiaofei Ma, Cicero Nogueira dos Santos, and Andrew O Arnold. 2021. Contrastive fine-tuning improves robustness for neural rankers. *arXiv preprint arXiv:2105.12932*.

Katerina Margatina, Giorgos Vernikos, Loïc Barrault, and Nikolaos Aletras. 2021. Active learning by acquiring contrastive examples. *arXiv preprint arXiv:2109.03764*.

Joshua Maynez, Shashi Narayan, Bernd Bohnet, and Ryan McDonald. 2020. On faithfulness and factuality in abstractive summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online. Association for Computational Linguistics.

Yu Meng, Chenyan Xiong, Payal Bajaj, Saurabh Tiwary, Paul Bennett, Jiawei Han, and Xia Song. 2021. COCO-LM: correcting and contrasting text sequences for language model pretraining. *CoRR*, abs/2102.08473.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Hyun Oh Song, Yu Xiang, Stefanie Jegelka, and Silvio Savarese. 2016. Deep metric learning via lifted structured feature embedding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4004–4012.

Bo Ouyang, Wenbing Huang, Runfa Chen, Zhixing Tan, Yang Liu, Maosong Sun, and Jihong Zhu. 2021. Knowledge representation learning with contrastive completion coding. *EMNLP-Findings 2021*.

Xiao Pan, Mingxuan Wang, Liwei Wu, and Lei Li. 2021. Contrastive learning for many-to-many multilingual neural machine translation. *arXiv preprint arXiv:2105.09501*.

Nikolaos Pappas and James Henderson. 2018. Joint input-label embedding for neural text classification. *CoRR*, abs/1806.06219.

Bhargavi Paranjape, Julian Michael, Marjan Ghazvininejad, Luke Zettlemoyer, and Hannaneh Hajishirzi. 2021. Prompting contrastive explanations for commonsense reasoning tasks. *arXiv preprint arXiv:2106.06823*.

Ankur P. Parikh, Xuezhi Wang, Sebastian Gehrmann, Manaal Faruqui, Bhuwan Dhingra, Diyi Yang, and Dipanjan Das. 2020. Totto: A controlled table-to-text generation dataset. *CoRR*, abs/2004.14373.

Senthil Purushwalkam and Abhinav Gupta. 2020. Demystifying contrastive self-supervised learning: Invariances, augmentations and dataset biases. *arXiv preprint arXiv:2007.13916*.

Yujia Qin, Yankai Lin, Ryuichi Takanobu, Zhiyuan Liu, Peng Li, Heng Ji, Minlie Huang, Maosong Sun, and Jie Zhou. 2020. Erica: improving entity and relation understanding for pre-trained language models via contrastive learning. *arXiv preprint arXiv:2012.15022*.

Yao Qiu, Jinchao Zhang, and Jie Zhou. 2021. Improving gradient-based adversarial training for text classification by contrastive learning and auto-encoder. *arXiv preprint arXiv:2109.06536*.

Yanru Qu, Dinghan Shen, Yelong Shen, Sandra Sajeev, Weizhu Chen, and Jiawei Han. 2021. Co{da}: Contrast-enhanced and diversity-promoting data augmentation for natural language understanding. In *International Conference on Learning Representations*.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. *arXiv preprint arXiv:2103.00020*.

Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

Nils Rethmeier and Isabelle Augenstein. 2020. Long-tail zero and few-shot learning via contrastive pretraining on and for small data. *CoRR*, abs/2010.01061.

Nils Rethmeier and Isabelle Augenstein. 2021. A primer on contrastive pretraining in language processing: Methods, lessons learned and perspectives. *arXiv preprint arXiv:2102.12982*.

Daniela N Rim, DongNyeong Heo, and Heeyoul Choi. 2021. Adversarial training with contrastive learning in nlp. *arXiv preprint arXiv:2109.09075*.

Joshua Robinson, Ching-Yao Chuang, Suvrit Sra, and Stefanie Jegelka. 2020. Contrastive learning with hard negative samples. *arXiv preprint arXiv:2010.04592*.

Alexis Ross, Ana Marasović, and Matthew E Peters. 2020. Explaining nlp models via minimal contrastive editing (mice). *arXiv preprint arXiv:2012.13985*.

Ruslan Salakhutdinov and Geoff Hinton. 2007. Learning a nonlinear embedding by preserving class neighbourhood structure. In *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics*.

Florian Schroff, Dmitry Kalenichenko, and James Philbin. 2015. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823.

Tal Schuster, Adam Fisch, and Regina Barzilay. 2021. Get your vitamin c! robust fact verification with contrastive evidence. *arXiv preprint arXiv:2103.08541*.

Hooman Sedghamiz, Shivam Raval, Enrico Santus, Tuka Alhanai, and Mohammad Ghassemi. 2021. Supcl-seq: Supervised contrastive learning for downstream optimized sequence representations. *arXiv preprint arXiv:2109.07424*.

Dinghan Shen, Mingzhi Zheng, Yelong Shen, Yanru Qu, and Weizhu Chen. 2020. A simple but tough-to-beat data augmentation approach for natural language understanding and generation. *arXiv preprint arXiv:2009.13818*.

Chang Shu, Yusen Zhang, Xiangyu Dong, Peng Shi, Tao Yu, and Rui Zhang. 2021. Logic-consistency text generation from semantic parses. *arXiv preprint arXiv:2108.00577*.

Kihyuk Sohn. 2016. Improved deep metric learning with multi-class n-pair loss objective. In *Advances in neural information processing systems*, pages 1857–1865.

Siqi Sun, Zhe Gan, Yu Cheng, Yuwei Fang, Shuohang Wang, and Jingjing Liu. 2020. Contrastive distillation on intermediate representations for language model compression. *arXiv preprint arXiv:2009.14167*.

Varsha Suresh and Desmond C Ong. 2021. Not all negatives are equal: Label-aware contrastive loss for fine-grained text classification. *arXiv preprint arXiv:2109.05427*.

Yonglong Tian, Chen Sun, Ben Poole, Dilip Krishnan, Cordelia Schmid, and Phillip Isola. 2020. What makes for good views for contrastive learning? *arXiv preprint arXiv:2005.10243*.

Yui Uehara, Tatsuya Ishigaki, Kasumi Aoki, Hiroshi Noji, Keiichi Goshima, Ichiro Kobayashi, Hiroya Takamura, and Yusuke Miyao. 2020. Learning with contrastive examples for data-to-text generation. In *Proceedings of the 28th International Conference on Computational Linguistics*.

Jannis Vamvas and Rico Sennrich. 2021. Contrastive conditioning for assessing disambiguation in mt: A case study of distilled bias. In *EMNLP 2021*.

Aäron van den Oord, Yazhe Li, and Oriol Vinyals. 2018. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*.

Danqing Wang, Jiaze Chen, Hao Zhou, Xipeng Qiu, and Lei Li. 2021a. Contrastive aligned joint learning for multilingual summarization. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 2739–2750.

Dong Wang, Ning Ding, Piji Li, and Hai-Tao Zheng. 2021b. Cline: Contrastive learning with semantic negative examples for natural language understanding. *arXiv preprint arXiv:2107.00440*.

Liwen Wang, Xuefeng Li, Jiachi Liu, Keqing He, Yuanmeng Yan, and Weiran Xu. 2021c. Bridge to target domain by prototypical contrastive learning and label confusion: Re-explore zero-shot learning for slot filling. *arXiv preprint arXiv:2110.03572*.

Ziqi Wang, Xiaozhi Wang, Xu Han, Yankai Lin, Lei Hou, Zhiyuan Liu, Peng Li, Juanzi Li, and Jie Zhou. 2021d. Cleve: Contrastive pre-training for event extraction. *arXiv preprint arXiv:2105.14485*.

Hanlu Wu, Tengfei Ma, Lingfei Wu, Tariro Manyumwa, and Shouling Ji. 2020a. Unsupervised reference-free summary quality evaluation via contrastive learning. *arXiv preprint arXiv:2010.01781*.

Zhuofeng Wu, Sinong Wang, Jiatao Gu, Madian Khabsa, Fei Sun, and Hao Ma. 2020b. CLEAR: contrastive learning for sentence representation. *CoRR*, abs/2012.15466.

Tete Xiao, Xiaolong Wang, Alexei A Efros, and Trevor Darrell. 2021. What should not be contrastive in contrastive learning. In *International Conference on Learning Representations*.

Lee Xiong, Chenyan Xiong, Ye Li, Kwok-Fung Tang, Jialin Liu, Paul N. Bennett, Junaid Ahmed, and Arnold Overwijk. 2021. Approximate nearest neighbor negative contrastive learning for dense text retrieval. In *International Conference on Learning Representations*.

Hu Xu, Gargi Ghosh, Po-Yao Huang, Dmytro Okhonko, Armen Aghajanyan, and Florian Metze Luke Zettlemoyer Christoph Feichtenhofer. 2021a. Videoclip: Contrastive pre-training for zero-shot video-text understanding. *arXiv preprint arXiv:2109.14084*.

Peng Xu, Xinchi Chen, Xiaofei Ma, zhiheng huang, and Bing Xiang. 2021b. Contrastive document representation learning with graph attention networks. *EMNLP-Findings 2021*.

An Yan, Zexue He, Xing Lu, Jiang Du, Eric Chang, Amilcare Gentili, Julian McAuley, and Chun-Nan Hsu. 2021a. Weakly supervised contrastive learning for chest x-ray report generation. *arXiv preprint arXiv:2109.12242*.

Yuanmeng Yan, Rumei Li, Sirui Wang, Fuzheng Zhang, Wei Wu, and Weiran Xu. 2021b. Consert: A contrastive framework for self-supervised sentence representation transfer. *arXiv preprint arXiv:2105.11741*.

Haoran Yang, Wai Lam, and Piji Li. 2021a. Contrastive representation learning for exemplar-guided paraphrase generation. *arXiv preprint arXiv:2109.01484*.

Nan Yang, Furu Wei, Binxing Jiao, Daxing Jiang, and Linjun Yang. 2021b. xmoco: Cross momentum contrastive learning for open-domain question answering. 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 6120–6129.

Seonghyeon Ye, Jiseon Kim, and Alice Oh. 2021. Efficient contrastive learning via novel data augmentation and curriculum learning. *arXiv preprint arXiv:2109.05941*.

Chenyu You, Nuo Chen, and Yuexian Zou. 2021. Self-supervised contrastive cross-modality representation learning for spoken question answering. *arXiv preprint arXiv:2109.03381*.

Yue Yu, Simiao Zuo, Haoming Jiang, Wendi Ren, Tuo Zhao, and Chao Zhang. 2020. Fine-tuning pre-trained language model with weak supervision: A contrastive-regularized self-training approach. *arXiv preprint arXiv:2010.07835*.

Zhenrui Yue, Bernhard Kratzwald, and Stefan Feuerriegel. 2021. Contrastive domain adaptation for question answering using limited text corpora. *arXiv preprint arXiv:2108.13854*.

Zhiyuan Zeng, Keqing He, Yuanmeng Yan, Zijun Liu, Yanan Wu, Hong Xu, Huixing Jiang, and Weiran Xu. 2021. Modeling discriminative representations for out-of-domain detection with supervised contrastive learning. *arXiv preprint arXiv:2105.14289*.

Dejiao Zhang, Shang-Wen Li, Wei Xiao, Henghui Zhu, Ramesh Nallapati, Andrew O Arnold, and Bing Xiang. 2021a. Pairwise supervised contrastive learning of sentence representations. *arXiv preprint arXiv:2109.05424*.

Dejiao Zhang, Feng Nan, Xiaokai Wei, Shangwen Li, Henghui Zhu, Kathleen McKeown, Ramesh Nallapati, Andrew Arnold, and Bing Xiang. 2021b. Supporting clustering with contrastive learning. *arXiv preprint arXiv:2103.12953*.

Jianguo Zhang, Trung Bui, Seunghyun Yoon, Xiang Chen, Zhiwei Liu, Congying Xia, Quan Hung Tran, Walter Chang, and Philip Yu. 2021c. Few-shot intent detection via contrastive pre-training and fine-tuning. *arXiv preprint arXiv:2109.06349*.

Wenzheng Zhang and Karl Stratos. 2021. Understanding hard negatives in noise contrastive estimation. *arXiv preprint arXiv:2104.06245*.

Yan Zhang, Ruidan He, Zuozhu Liu, Kwan Hui Lim, and Lidong Bing. 2020a. An unsupervised sentence embedding method bymutual information maximization. *CoRR*, abs/2009.12061.

Zhu Zhang, Zhou Zhao, Zhijie Lin, Xiuqiang He, et al. 2020b. Counterfactual contrastive learning for weakly-supervised vision-language grounding. *Advances in Neural Information Processing Systems*, 33:18123–18134.

Wenxuan Zhou and Muhao Chen. 2021. Contrastive out-of-distribution detection for pretrained transformers. *arXiv preprint arXiv:2104.08812*.