Contrastive Learning and Self-Supervised Learning: Data Driven Problem Solving Final Report

Authors: Shreyas Krishnan and

Gautaman Asirwatham

ME371 – Data Driven Problem Solving

Professor Masoud Masoumi

The Cooper Union

Introduction

The technical report discusses the historical background, technical aspects, current applications, future directions, and many other topics about contrastive learning and self-supervised learning. Data is used to train models and data can come in two main forms, labeled and unlabeled. Labeled data has been annotated by humans. This is a time consuming and expensive process and the vast majority of real world data is messy and unlabeled [3].

Contrastive learning and self-supervised learning are two machine learning paradigms where the models are trained on unlabeled data. A few other popular learning methods include supervised learning, unsupervised learning, and semi-supervised learning. Supervised learning is done by using labeled data and can be used for object classification. Unsupervised learning is done by using unlabeled data and can be used for clustering [1].

Semi-supervised learning is a mix of supervised and unsupervised learning. Both contrastive learning and self-supervised learning are forms of unsupervised learning. In self-supervised learning, the model annotates its own data labels to the data. It uses some of the input to learn more of the input. Self-supervised learning can be broken down into either generative or contrastive methods. The main difference is where the loss is measured. In generative methods, the loss is measured at the output while the loss is measured in the representation space for contrastive methods [6]. Contrastive learning is a technique which uses comparisons between inputs to learn from the data. Today, these methods are widely used in the fields of computer vision, natural language processing, and speed processing.

Historical Background

Despite the recent popularity of self-supervised learning, this is not a new concept. The earliest usage of the term is 1978. In 2006, Geoffrey Hinton introduced the idea of using an unsupervised way to pre-train Restricted Boltzmann Machines. Even though this concept has been available for many years, the recent resurgence can be attributed to the abundant unlabeled data and the pressing need to extract data from it [10]. The term self-supervised learning was first used in the field of robotics with respect to input sensor data, but was adapted to the field of machine learning. A notable moment in contrastive self-supervised learning’s history is when the method trained on unlabeled ImageNet data was more accurate than supervised AlexNet, a convolutional neural network trained for image classification. Another moment was when the model pre-trained on ImageNet data was more capable at certain downstream tasks than supervised pre-trained counterparts [6].

Technical Concept Overview

In generative learning data is selectively removed to generate the training set. Next the model attempts to reconstruct the missing data. Finally the model compares its guess to the actual sample the training data is taken from. This allows the model to create its own loss signal and update its weights.

This technique is illustrated in BERT which was trained by google to understand search queries. BERT was trained using the BookCorpus (800M words) and a filtered version of English Wikipedia (2.5 B words). Fifteen percent of tokens are replaced with a “mask” token. In training the model attempts to replace the mask token with the appropriate word. It generates probabilities for what tokens should replace the mask token. Then the true token is used to calculate the loss between the distribution of the model’s guesses and the actual token. Finally, the model’s weights are updated to minimize this loss [13].

In contrastive learning, the model learns a representational space in which similar concepts are closer and dissimilar concepts are far apart. In training the model compares the training sample to similar images called positive pairs and dissimilar images called negative pairs. It pushes dissimilar representations away from each other and pulls similar representations together [6].

The comparing process is done by an encoder. The encoder aims to learn a score function such that any data point, which we call the “anchor”, has a higher score with a more similar point, called “positive”, than a more dissimilar point which we call “negative” [6].

$$\underline{\mathrm{score}(f(x), f(x^{+}))} >> \underline{\mathrm{score}(f(x), f(x^{-}))}$$

[6] Contrastive Self-Supervised Learning | Ankesh Anand

The loss function (shown below) takes dot products between the anchor and positive/negative samples. This is calculating the similarity in representational space. Its form resembles a softmax normalization over a set of comparisons. By minimizing this loss function we are trying to reduce the dot product of anchor and negative samples and increase the dot product of anchor and positive samples [6].

$$\mathcal{L}_N = -\mathbb{E}_X \left[\log \frac{\overline{\exp\left(f(x)^T f\left(x^{+}\right)\right)}}{\underline{\exp\left(f(x)^T f\left(x^{+}\right)\right)} + \sum_{j=1}^{N-1} \underline{\exp\left(f(x)^T f\left(x_j\right)\right)}} \right]$$

[6] Contrastive Self-Supervised Learning | Ankesh Anand

An example of how contrastive learning is used is image models. If we take a part of an image to be our anchor, another part of the same image would be our positive sample while another part of a different image would be our negative sample. We update our encoder to reduce the loss of our score function [6].

Current Applications and Use Cases

Contrastive self-supervised learning is widely used today for natural language processing and computer vision. Natural language processing was the first field that widely adopted contrastive self-supervised learning but today the method is most extensively used in the computer vision field. Natural language processing is more difficult than image augmentation since language modifications should still retain the original meaning. For example when a sentence is augmented, it can lose its meaning and the model must ensure that the sentence is still trying to convey the same information.

In the computer vision field there are many applications and established models. For video sequence prediction the VideoMoCo model uses temporally adversarial examples to change the input. For object detection, DetCo uses contrastive learning between local patches of an image and the global image. For perceptual audio similarity, CDPAM classifies audio samples based on their perceptual similarity [2]. Simple Contrastive Learning (SimCLR) is a model type that is trained to recognize images that have undergone various transformations. This can be used to learn image classification and object detection. Some examples of object transformations that this can detect are flipping, rotating, cropping, color distortion, sharpening, and random erasing [4].

In natural language processing this method of learning is used in a variety of ways. One example is back-translation. This involves the generation of augmented sentences. A sentence that has been translated to a different language is translated back to provide the translator with a better understanding of what the final translation is. Another example is for lexical edits. Lexical edits include random insertions, random swaps, random deletions, and synonym replacements. This type of model is also used for cutoffs and dropouts. Cutoffs include Feature Cutoffs, Token Cutoffs and Span Cutoffs [2].

There are a wide variety of fields that contrastive self-supervised learning is used for such as healthcare, 3D rotation, signature detection, image annotation, image editing, and robotics. In healthcare, this is used for cancer detection and x-ray segmentation. It has been used to control robot movements that rotate objects in 3 dimensions. Models have been used to detect forgeries and authenticate valid signatures. The mars rover relied heavily on unsupervised navigation since the lag between Earth and Mars makes manual control infeasible [3].

Benefits and Drawbacks

The main advantage for contrastive self-supervised learning is the ability to learn from unlabeled data. This type of data is significantly more abundant than labeled data. As a result, this method is highly scalable and can be used in far more situations than supervised methods. This method also explores a machine's ability to think independently like a human. The model needs to be able to evaluate its own generated labels and decide if it should use them in the future. Since this type of learning is used on unlabeled data, it can be used in a very wide variety of applications while other methods have more specific use cases [4]. Another key advantage of this method is its ability to be useful in transfer learning. Transfer learning is making use of a model trained on a certain task to perform on a related task. This highlights the high adaptability of the learning method [5]. This method also results in a model that can recognize a new concept after only seeing a few labeled examples. This highlights that the method allows for quick learning.

Despite the benefits, there are some drawbacks. This method requires extensive computational power and computational resources. This is due to the fact that the model needs to make sense of the unlabeled data and generate corresponding labels. Specifically for contrastive self-supervised learning, for each anchor-positive pair, multiple anchor-negative pairs need to be tested in every iteration [3]. Another major drawback is accuracy. This method of learning is not always as accurate as supervised methods. If the model predicts wrongly but is highly confident in its prediction, the model will continue to make wrong predictions and will not tune its weights accordingly. Contrastive self-supervised learning has a wide use case but there are certain applications that require supervised learning and contrastive self-supervised learning cannot be used [3]. This method also lacks standardized evaluation methods. This causes difficulties with regard to comparing different models [12]. Table 1 summarizes the advantages and disadvantages of contrastive self-supervised learning.

Table 1: Summarized Advantages and Disadvantages of Contrastive Self-Supervised Learning

Advantages	Disadvantages
Uses abundant unlabeled data,	Needs extensive computation power and resources
Highly scalable and adaptable	Can be less accurate than supervised models
Insight into human learning	Certain use cases where it is not applicable
Useful in transfer learning	Lack of standardized evaluation metrics

Future Directions and Challenges

Contrastive self-supervised learning will continue to be developed and researched. This method is used at many large tech companies for various purposes. One of the companies that makes use of this currently and wants to develop it for specific use is Meta. Meta’s goal is to produce “machines with human-level intelligence.” They want to deploy a pre-trained language model onto Facebook to proactively detect hate speech. They hope to detect hate speech in languages where there is not much training data [9]. Jonathan Boigne believes that self-supervised learning will become the “gold standard in most fields of machine learning,” and will eventually have a wider adoption. Boigne also points out the public availability and open-source nature of self-supervised models and how the mass sharing to the technique will lead to use in every application [10].

Some of the disadvantages mentioned previously are also challenges that researchers and specialists are trying to address. The main challenge to tackle is accuracy. The current way to address this issue is to use vast amounts of data or just accept the loss in accuracy. This accuracy challenge is very crucial to solve since inaccurate generated labels would just continue to propagate. The other challenge is computational efficiency. The vast amount of data needed to improve the accuracy also has a significant impact on the computational efficiency. This is also compounded with the multiple stages of training. Another challenge for prediction models is choosing the correct pretext task. This directly affects how the model will learn and perform in the future and this needs to be taken into high consideration for this type of modeling [1].

Ethical Considerations

One of the major topics in the field of machine learning and AI is ethics. Issues such as privacy, fairness, transparency, and trust are major factors. With regard to machine learning, a certain question about bias and discrimination comes up, "How can we safeguard against bias and discrimination when the training datasets can lend itself to bias?" [11]. Contrastive self-supervised learning is at risk of supporting and intensifying data biases. As a result, when the model is implemented, the implementation must ensure that the model does not promote and enhance harmful biases [12].

References

[1]Shah, D., 2023, "Self-supervised learning and its applications," neptune.ai [Online]. Available: https://neptune.ai/blog/self-supervised-learning. [Accessed: 17-Dec-2024].

[2]"The beginner's guide to contrastive learning," V7 [Online]. Available: https://www.v7labs.com/blog/contrastive-learning-guide. [Accessed: 17-Dec-2024].

[3]"Self-supervised learning: Definition, tutorial & examples," Self-Supervised Learning: Definition, Tutorial & Examples [Online]. Available: https://www.v7labs.com/blog/self-supervised-learning-guide. [Accessed: 17-Dec-2024].

[4]Boesch, G., 2024, "Self-supervised learning: Everything you need to know (2024)," viso.ai [Online]. Available: https://viso.ai/deep-learning/self-supervised-learning-for-computer-vision/#elementor-toc__heading-anchor-5. [Accessed: 17-Dec-2024].

[5]GeeksforGeeks, 2023, "Self-supervised learning (SSL)," GeeksforGeeks [Online]. Available: https://www.geeksforgeeks.org/self-supervised-learning-ssl/. [Accessed: 17-Dec-2024].

[6]Ankeshanand, 2020, "Contrastive self-supervised learning," Ankesh Anand [Online]. Available:

https://ankeshanand.com/blog/2020/01/26/contrative-self-supervised-learning.html. [Accessed: 17-Dec-2024].

[7]Yolyan, L., 2021, “Review on self-supervised contrastive learning,” Medium [Online]. Available: https://towardsdatascience.com/review-on-self-supervised-contrastive-learning-93171f695140. [Accessed: 17-Dec-2024].

[8]Ibm, 2024, “What is self-supervised learning?,” IBM [Online]. Available: https://www.ibm.com/think/topics/self-supervised-learning. [Accessed: 17-Dec-2024].

[9]“Self-supervised learning: The dark matter of intelligence,” AI at Meta [Online]. Available: https://ai.meta.com/blog/self-supervised-learning-the-dark-matter-of-intelligence/. [Accessed: 17-Dec-2024].

[10]Boigne, J., 2020, “The rise of self-supervised learning,” Jonathan Bgn [Online]. Available: https://jonathanbgn.com/2020/12/31/self-supervised-learning.html. [Accessed: 17-Dec-2024].

[11]Ibm, 2024, “What is ai ethics?,” IBM [Online]. Available: https://www.ibm.com/topics/ai-ethics#:~:text=Examples%20of%20AI%20ethics%20issues,%2C%20trust%2C%20and%20technology%20misuse. [Accessed: 17-Dec-2024].

[12]Rosidi, N., 2024, “Self-supervised learning guide: Super simple way to understand AI,” Medium [Online]. Available: https://nathanrosidi.medium.com/self-supervised-learning-guide-super-simple-way-to-understand-ai-f7a47f1a7b7a#:~:text=Ethical%20and%20Bias%20Considerations,are%20not%20enhancing%20harmful%20biases. [Accessed: 17-Dec-2024].

[13] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K., 2019, *“BERT: Pre-training of deep bidirectional transformers for language understanding,”* arXiv [Online]. Available: https://arxiv.org/abs/1810.04805. [Accessed: 17-Dec-2024].