EC601 Project1

Self-Supervised Learning for Graph Data Structure

Name: Zhang Haocheng

BUID:U94536742

1. BACKGROUND
   1. What is self-supervised learning?

Simply speaking, self-supervised learning is also a unsupervised learning method which means it doesn’t need labels or classified character in advance by human. Mainly self-supervised learning by itself. For example, for some searching engines, self-supervised learning can help the fill the blanks by understanding the context better.

Also, for some graph data, it can help to complete the whole picture even losing some parts of it.

* 1. Why we should use self-supervised learning? (Advantages)

From the perspective of cost and process, self-supervised learning can realize the goal, training a efficient module by using less points and data. Because labeling the data is a high cost comparing with collecting data. What else, in the process of reinforcement learning, we can reduce the training times by using self-supervised learning.

As for the performance of training, self-supervised learning machine can predict any part of the graph. It can predict the coming frames of a video. Every sample can provide more information than it was compared with supervised and unsupervised learning.

* 1. Why is self-supervised learning possibly available?

In the blog deriving from <https://ankeshanand.com/blog/2020/01/26/contrative-self-supervised-learning.html>, he dropped a good example for us. People cannot draw a 1 dollar cash by hands according to their memory. However, people can draw a frame with some features inside, which means they use few graphic information and distinguish 1 dollar from 20, 50, 100 dollar. So it may also works on machine learning only if it gains the semantic meaning from the information.

1. Methods of self-supervised learning

According to the research, it could be divided into two paradigms: Generative Methods and Contrastive Self-supervised learning.

* 1. Generative Methods

These methods mainly focus on the pixel space. They reconstruct the pixel and discover the Loss of pixel label. For the encoder, they want it to maintain the features of dataset as much as possible. After decoding, the similarity could be counted. However, the cost of these methods are too high because they finish all the process on pixel space which has high complexity.

* 1. Contrastive Methods

Contrastive methods do not need to reconstruct the initial input and they switch the attention on feature space, which means let machine distinguish graph from objects or patches. In some extent, this proves machines learn the semantical meanings.

* 1. Contract of these two methods

图示

描述已自动生成

Figure Source:https://ankeshanand.com/blog/2020/01/26/contrative-self-supervised-learning.html

Compared with generative methods, contrastive methods are apparently more practical and possible to realize. Here are the advantages:

* They discover the similarity and loss through the distance of feature space instead of the complex pixel space.
* There are more predictions because they can adjust the features, like rotation, augmentation as well as segmentation. Each sample can provide more information.
* More easy to optimize and deep learning.

1. Realization

Unfortunately, there are some coming problems, like how to define the positive and negative object because of lack of labels. Now based on the contrastive methods research show a positive development.

3.1Relationship between supervised and self-supervised learning

Actually there are two main paradigms: supervised learning and unsupervised learning. Now we find that “unsupervised learning” couldn’t exactly describe the meaning and we normally use the term, “self-supervised learning. In the supervised learning, we build the dataset as {(, ) (, ) ……}. Each x represents a image and y in the same group represents a tag or label. We hope that machines can optimize the output y according to the dataset from a input y. So the function of it is f () and is the parameter. While as for the self-supervised learning, the dataset has no label at all, like D={(,) ……}. To solve this problem, we need to introduce another concept of “pretext”.

3.2 Introduction of “Pretext”

Firstly, we should know what is the effect of the “pretext”. We all know that self-supervised learning is the method with no label unlike the supervised learning. However, making machine learning from nothing is too difficult to processing and we also need them to recognize the semantic characters. In this case, there is a compromise way. Pretext comes out as a method that designed by developers to achieve some efficient effect and learn some valuable characters to memory. Now a new question appears. How can we design a effective “pretext”?

3.3 Two types of “pretext” building

In this case, we find another approach to build the pretext—defining a proxy task and solving the new question. However, there is a new problem on the proxy task. It always has a gap from the mission we should achieve, which means it couldn’t be used on the practical process. This question had existed for a long time until 2016 researchers raise a solution called “Linear Probe”.

图形用户界面, 应用程序

描述已自动生成

Figure source：[arXiv:1905.09272](https://arxiv.org/abs/1905.09272)**[cs.CV]**

This figure shows a good example for us about the whole process that we just mentioned. This is from the paper [2]. Firstly, they train the machine by encoding that mutual coupling elements from one object (mostly for the image we decouple them in RBG way). Then they make the encoding function fixed and train every vector divided from the object through the encoder. After that they add a classifier to select the vector or feature which has the high-extended similarity. Then train them again and make the efficient features through the transfer and gain the Muti-task they need to check the Loss and similiarity. To prove the performance, they also put the supervised training consequence and it presents a more excellent result.

When we facing to define a proxy task for a graph, there are three normal directions to learn for reference. One is reconstruction which is used on the example above. The other two are colorization and rotation respectively. For colorization, it means machine should approximately know that sky is blue and mountains are green. That represents machine knows some semantic meanings through training. Besides, rotation is a popular thoughts for decoupling the graphic objects. Because it is simple and can easily achieve the learning goal.

图形用户界面

中度可信度描述已自动生成

Figure source: [Hankook Lee et al. 2020](https://arxiv.org/pdf/1910.05872.pdf)

These examples from paper [3] use different rotations and colorizations as transformers to expand the training set. This article points out that it could make the learning harder by training through data augmentation or multi-task learning method. It is possible to lower the performance. The paper suggest that increase the groups and sorts created by self-supervised learning and use the Loss function to calculate.

1. Supplement

4.1Two paradigms of proxy using

* Define a proxy task, proxy task is a data base for the unsupervised learning which can provide data for the supervision. After you define a proxy task, then try to solve it. However, it cannot be used for the downstream task before training with data with label because they only learn instance discrimination. The training object could be a classifier.
* Others are available for the downstream tasks, which are mostly used into prediction of video frames.

4.2Based on the context

Word2Vec: CBOW model and Skip-gram model:

CBOW model can predict the middle words according to the context. In the contrary, Skip-gram model can learn the prediction of surroundings of a word. Both of them have their own advantages and disadvantages. In paper [1] study, they proposed a method to combine the CBOW model and the skip-gram model by adjust the weight of them to realize the performance as well as the accuracy.

图示, 示意图

描述已自动生成

This model is designed mainly for the text prediction, but the thoughts of combination drop lessons for our project.

1. References

[1] T. Onishi and H. Shiina, "Distributed Representation Computation Using CBOW Model and Skip–gram Model," *2020 9th International Congress on Advanced Applied Informatics (IIAI-AAI)*, 2020, pp. 845-846, doi: 10.1109/IIAI-AAI50415.2020.00179.

[2] Henaff, Olivier. Data-Efficient Image Recognition with Contrastive Predictive Coding. 2020. Proceedings of the 37th International Conference on Machine Learning, PMLR 119:4182-4192, 2020.

[3] Lee, H., Hwang, S.J. &amp; Shin, et al. Self-supervised Label Augmentation via Input Transformations. Proceedings of the 37th International Conference on Machine Learning, 2020.119:5714-5724

[4] Zhai, Xiaohua et al. “S​L: Self-Supervised Semi-Supervised Learning.” ” ICCV 2019

Source: [Exploiting Similarities among Languages for Machine Translation](https://arxiv.org/pdf/1309.4168v1.pdf) paper.

<https://towardsdatascience.com/nlp-101-word2vec-skip-gram-and-cbow-93512ee24314>

[5]T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In NIPS, 2013. 1, 2

[arXiv:2104.07658](https://arxiv.org/abs/2104.07658) [cs.CV] C. Yang, H. Lamdouar, E. Lu, A. Zisserman, W. Xie  
Self-supervised Video Object Segmentation by Motion Grouping

[6]Chen, T., Kornblith, S., Norouzi, M. &amp; Hinton, G.. (2020). A Simple Framework for Contrastive Learning of Visual Representations.

[7]van den Oord, A., Li, Y., and Vinyals, O. Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748, 2018.

[8]Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S.,Ma, S., Huang, Z., Karpathy, A.,Khosla, A., Bernstein,M., et al. Imagenet large scale visual recognition challenge.International journal of computer vision, 115(3):211–252, 2015.

[9]YM., A., C., R., and A., V. Self-labelling via simultaneousclustering and representation learning. In InternationalConference on Learning Representations, 2020. URL https://openreview.net/forum?id=Hyx-jyBFPr.

[10]Pinto, L., Davidson, J., and Gupta, A. Supervision via competition: Robot adversaries for learning tasks. *arXiv preprint arXiv:1610.01685*, 2016.