Unsupervised 3D Shape Classification Using Graph Neural Networks

Yu Hou, Won-Yong Shin
Yonsei University

houyu@yonsei.ac.kr, wy.shin@yonsei.ac.kr

Abstract

Point clouds consist of a discrete set of points irregularly sampled from continuous 3D objects. The representation quality of point clouds is crucial to the downstream task of classification in 3D vision. In this paper, we propose a novel solution for learning the representation of point cloud classification by multi-scale graph contrastive learning. The contributions are two-fold. First, we consider the multi-scale point clouds, which are sampled by different ratios for generating the positive samples in contrastive learning. Second, we adopt the feature of point clouds from different views for generating multi-view features. We conduct experiments on point cloud datasets for showing the potential gain of the proposed method.

Ⅰ. Introduction

3D point clouds serve as an efficient representation of 3D objects or natural scenes, which consist of irregularly sampled 3D points associated with multiple modalities. There is an observation that, as a representation of 3D objects or scenes, a point cloud usually exhibits nonlocal self-similarity [1]. Based on this observation, we propose to learn the representation from multi-scale point clouds, where the point clouds with different sample ratios are used as the positive samples. Alongside multi-view features of point clouds, we adopt a graph neural network (GNN) model to extract the shape-level embedding and the GNN model is trained in the sense of minimizing a contrastive loss function [2].

Ⅱ. Methodology

For each anchor point cloud P_i, we first select p multi-scale point clouds as positive samples, which are sampled from P_i with different sample ratios and denoted as $\{P_{i1}^{+}, \cdots, P_{ip}^{+}\}$. Then, q negative point cloud samples are selected from other point clouds, which are denoted as $\{P_{i1}^{-}, \cdots, P_{iq}^{-}\}$. The graph structure G_i can be generated by adding edges between the target point and its neighbors using k-nearest neighbors based on the Euclidean distance for each point cloud. We concatenate the coordinates of point clouds from different views as multi-view features of a 3D object. The shape-level embedding E_i of point cloud P_i is learned by the GNN encoder, which is formulated as $E_i = GNN_{\theta}(G_i, X_i)$, where θ is the trainable weights in the GNN model. The contrastive learning loss function is used for training the GNN encoder to make the positive samples close to the anchor and push the negative samples away from the anchor in the embedding space, which is formulated as follows:

$$L(P_i, P_j) = (1 - Y)\frac{1}{2}D_{\theta}^{2} + (Y)\frac{1}{2}\{max(0, m - D_{\theta})\}^{2},$$

where $Y = 1$ for negative pairs and $Y = 0$ for positive pairs, m is the so-called "margin" hyperparameter [2], and $D_{\theta} = \|E_i - E_j\|^2$ is the Euclidean distance between the representations of i-th and j-th point clouds.

Ⅲ. Experimental Results

We conduct experiments on ModelNet 10, which is a 3D shape real-world dataset collected from the CAD model. We quantify the accuracy of classification with the linear evaluation protocol, which is the portion of correct prediction accounting for the total number of samples. For the evaluation of classification, we use a linear protocol. Figure 1 is the loss and accuracy of the linear classifier. It is observed that the training accuracy is increased with epochs—the accuracy of training data is up to 84% and the accuracy of test data is up to 80%.

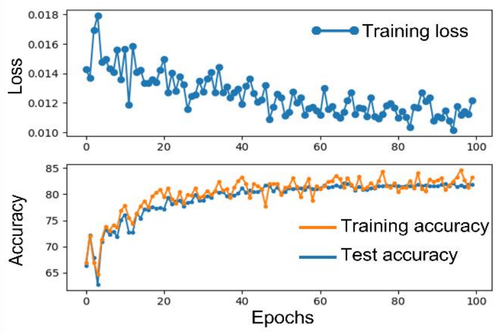

Figure 1. The training loss and the accuracy in train and test data.

ACKNOWLEDGMENT

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1A2C3004345) and by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.2021-0-00347, 6G Post-MAC (POsitioning- & Spectrum-aware intelligenT MAC for Computing & Communication Convergence)).

REFERENCES

[1] B. Du, X. Gao, W. Hu W, and X. Li, "Self-contrastive learning with hard negative sampling for self-supervised point cloud learning," ACM MM 2021.

[2] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," IEEE/CVF CVPR 2006.