Comparing Graph Architectures for Deep Metric Learning

Moritz Schüler
dept. of Informatics
Technical University Munich
Munich, Germany
moritz.schueler@tum.de

Luca Eyring
dept. of Informatics
Technical University Munich
Munich, Germany
luca.eyring@tum.de

Abstract

We compare different graph neural networks for the task of deep metric learning on the datasets CUB-200-2011, Cars196, Stanford Online Products, and In-Shop Clothes. Special focus is on different realizations of the attention mechanism. Furthermore, we try to enhance the graph construction by only connecting nodes with similar embeddings.

Keywords— deep metric learning, similarity learning, graph neural network, attention

1. Introduction

Deep metric learning tries to solve the task of attributing a similarity score between different data samples. In particular, it tries to find a lower dimensional mapping where the distance between data points depicts their similarity.

1.1. Related Work

The first works on similarity learning / deep metric learning compared embeddings with a distance function [3]. Improvements were made by adopting new loss functions. [6] introduced the Triplet loss, which compares not only two images, but provides a positive as well as a negative sample.

Furthermore, [5] not only uses three data points, but the whole mini-batch to improve the "clustering", which is called Group loss. Whilst considering the global structure the refinement step of the Group loss is fixed and cannot be adapted according to the samples of the current mini-batch. To fight this issue [7] leveraged the power of graph neural networks. By constructing a graph and running several message passing steps the whole structure of a mini-batch can be exploited, whilst also allowing for fine-tuning the refinement step. Another important part of the work was to use attention, as presented by [8].

1.2. Research Objective

In this project, we want to build upon [7]. First we want to compare its performance against graph attention models [2, 9] on the datasets of CUB-200-2011, Cars196, Stanford Online Products, and In-Shop Clothes. Specifically, the different implementations of the attention mechanism will be investigated, as work by [4] raised questions about the performance of attention without added linear layers.

Moreover, we want to improve on the sampling / construction of the graphs. Currently, a fully-connected graph is constructed for each mini batch. Our idea is to instead build the graph with limited connections between nodes. Here, we want to explore and compare different approaches starting with computing a similarity score between feature embeddings to decide which nodes to connect.

1.3. Further ideas

When improving the graph construction by leveraging node attributes, it might also make sense to introduce edge embeddings. [1] presented a general framework to divide the message passing step into note-to-edge and edge-to-node updates which could then be used to effectively train the network.

Another interesting continuation would be to investigate the transfer abilities of the trained networks from one dataset to another. To effectively analyze this we plan the following experiments:

1. Baseline results: Train the model on an evaluation dataset.
2. How does training between datasets impact performance: Train the model on a different dataset and afterwards use the trained model with a new final linear layer to fine-tune on the evaluation dataset.
3. How does the model generalize to other datasets: Same approach as *2*. However, freeze all layer except the last linear layer while fine-tuning on the evaluation dataset.

Additionally, we would like to look into how many samples are needed to effectively train the network. This could be done by iteratively reducing the amount of training samples and then comparing the performance to earlier results.

References

[1] Peter W. Battaglia, Jessica B. Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, Caglar Gulcehre, Francis Song, Andrew Ballard, Justin Gilmer, George Dahl, Ashish Vaswani, Kelsey

Allen, Charles Nash, Victoria Langston, Chris Dyer, Nicolas Heess, Daan Wierstra, Pushmeet Kohli, Matt Botvinick, Oriol Vinyals, Yujia Li, and Razvan Pascanu. Relational inductive biases, deep learning, and graph networks, 2018. 1

[2] Shaked Brody, Uri Alon, and Eran Yahav. How attentive are graph attention networks?, 2021. 1

[3] Jane Bromley, Isabelle Guyon, Yann LeCun, Eduard Säckinger, and Roopak Shah. Signature verification using a ”siamese” time delay neural network. In J. Cowan, G. Tesauro, and J. Alspector, editors, *Advances in Neural Information Processing Systems*, volume 6. Morgan-Kaufmann, 1994. 1

[4] Yihe Dong, Jean-Baptiste Cordonnier, and Andreas Loukas. Attention is not all you need: Pure attention loses rank doubly exponentially with depth, 2021. 1

[5] Ismail Elezi, Sebastiano Vascon, Alessandro Torcinovich, Marcello Pelillo, and Laura Leal-Taixe. The group loss for deep metric learning, 2020. 1

[6] Matthew Schultz and Thorsten Joachims. Learning a distance metric from relative comparisons. In S. Thrun, L. Saul, and B. Schölkopf, editors, *Advances in Neural Information Processing Systems*, volume 16. MIT Press, 2004. 1

[7] Jenny Seidenschwarz, Ismail Elezi, and Laura Leal-Taix. Learning intra-batch connections for deep metric learning, 2021. 1

[8] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017. 1

[9] Petar Velikovi, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Li, and Yoshua Bengio. Graph attention networks, 2018. 1