Backpropagating the Unsupervised Error of Self-Organizing Maps to Deep Neural Networks

Anonymous Author(s)
Affiliation
Address
email

1 Introduction

Previous research has shown the potential that Deep Neural Networks have in building representations that are useful not only for the task that the network was trained for but also for correlated tasks that take data from similar input distributions. For instance, Nanni et al. [9] showed that representations built by a Convolutional Neural Network (CNN) are better than the state-of-art handcrafted features used for object classification and Medeiros et al. [8] demonstrated that the representations learned by GoogleLeNet can be used for the task of Unsupervised Visual Object Recognition (UVOC), achieving about 75-90% of agreement with labels assigned by humans in an unseen dataset, when fed as input to a Self-Organizing Map (SOM)-based clustering method [2].

In this work, we propose an approach that combines SOM with Deep Learning in a synergic way to allow dealing with complex data structures, such as images and sound, by backpropagating the unsupervised error through layers of neurons.

2 Research Problem

A key to the success of supervised learning, especially, deep supervised learning, is the availability of sufficiently large labeled training data. Unfortunately, creating a sufficiently large amount of properly labeled data with enough examples for each class is not easy. On the other hand, unlabeled data usually can be easily obtained due to the advances in technology that have produced datasets of increasing size, not only in terms of the number of samples but also in the number of features.

SOM [6] is a biologically inspired unsupervised learning method that maps data from a higher-dimensional input space to a lower-dimensional output space, while preserving the similarities and the topological relations found between points in the input space. SOM can create abstractions and provide a simplified way of exhibiting information, being widely used in robotics, computer vision, data mining, and natural language processing.

State-of-the-art SOM-based models are suitable for clustering high-level features. In [1], a fixed topology SOM was proposed for subspace clustering by learning the relevance of each input dimension for each cluster. In [2], a time-varying structure version was proposed, in which the map only grows when new nodes are required, called LARFDSSOM. In [3, 4], this model was extended to take advantage of labeled data when it is available, thus enabling Semi-Supervised Learning.

However, despite these models are suitable for high-level features, they cannot deal with more complex data structures, such as images and sound. Therefore, combining SOM with Deep Learning, appears to be a good path to follow, and is the research problem considered in this work.

3 Motivation

DeepCluster, proposed by Caron et al. [5], provides a framework to train Deep Learning Models using K-means cluster to create pseudo-labels and adjust the weights. The work shows that good representations of visual features can be learned. Van den Oord et al. [11] introduce the VQ-VAE, a family of models that combine VAEs with vector quantization to build a discrete latent representation. Their experiments demonstrated that the discrete latent space generated by the model captures important features of the data in a completely unsupervised manner preserving the neighborhood. However, current SOM-based methods, such as LARFDSSOM [2], display far superior results in clustering than K-Means, being able to create a map that preserves the neighborhood and to learn the relevance of each input dimension for each cluster. This leads us to believe that the error signal provided by LARFDSSOM could be used to guide a Deep Neural Network to learn a smooth and disentangled representation in its latent space.

4 Technical Contribution

The main contribution of this work is the proposition of a new loss function based on the clustering error of SOM in a way that can be used to backpropagate errors to previous layers, counteracting the tendencies of the model to collapse to trivial solutions. This loss function allows learning mappings from a complex to a simple disentangled representation space that SOM can handle with smooth transitions between cluster points.

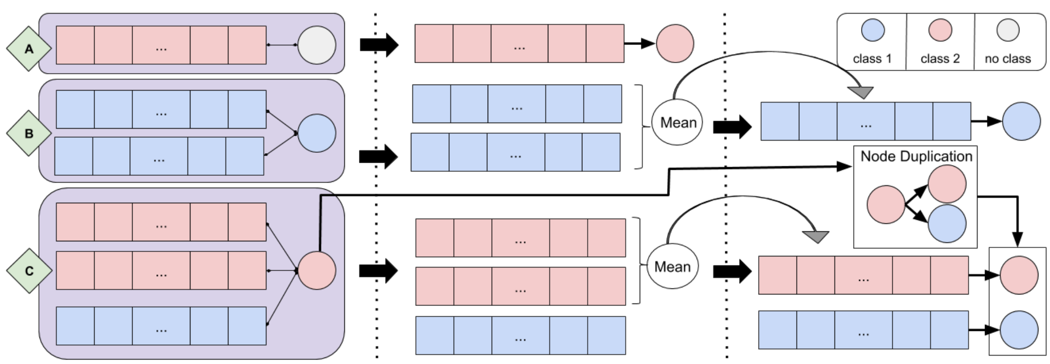

Figure 1: Each distinct situation that must be handled when a batch is given.

In order to achieve this, a PyTorch implementation of a SOM-based trainable layer was developed to take advantage of the high levels of parallelism of GPUs. This approach can be viewed as an extension of traditional shallow SOM models. It also supports semi-supervised learning by splitting the samples into two different batches. This process is illustrated by Fig. 1.

4.1 Experimental Results

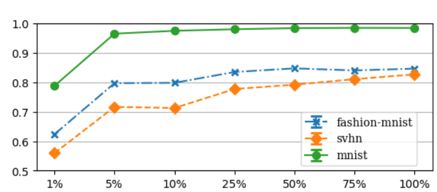

Figure 2: Results obtained varying the amount of labeled data from 1% to 100% on MNIST, Fashion-MNIST and SVHN datasets.

Preliminary results show that the proposed approach performs well in the image classification benchmark datasets, SVHN [10], MNIST [7], and Fashion-MNIST [12], even with very low amounts of labeled data (Fig. 2). However, it is necessary to carry out more detailed experiments to verify the properties of the representations learned.

References

[1] H. F. Bassani and A. F. Araújo. Dimension selective self-organizing maps for clustering high dimensional data. In *The 2012 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2012.

[2] H. F. Bassani and A. F. Araujo. Dimension selective self-organizing maps with time-varying structure for subspace and projected clustering. *IEEE transactions on neural networks and learning systems*, 26(3):458–471, 2014.

[3] P. H. Braga and H. F. Bassani. A semi-supervised self-organizing map for clustering and classification. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2018.

[4] P. H. Braga and H. F. Bassani. A semi-supervised self-organizing map with adaptive local thresholds. *arXiv preprint arXiv:1907.01086*, 2019.

[5] M. Caron, P. Bojanowski, A. Joulin, and M. Douze. Deep clustering for unsupervised learning of visual features. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 132–149, 2018.

[6] T. Kohonen. The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480, 1990.

[7] Y. LeCun. The mnist database of handwritten digits. *http://yann. lecun. com/exdb/mnist/*, 1998.

[8] H. R. Medeiros, F. D. de Oliveira, H. F. Bassani, and A. F. Araujo. Dynamic topology and relevance learning som-based algorithm for image clustering tasks. *Computer Vision and Image Understanding*, 179:19–30, 2019.

[9] L. Nanni, S. Ghidoni, and S. Brahnam. Handcrafted vs. non-handcrafted features for computer vision classification. *Pattern Recognition*, 71:158–172, 2017.

[10] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng. Reading digits in natural images with unsupervised feature learning. In *NIPS Workshop on Deep Learning and Unsupervised Feature Learning*, volume 2011, page 5, 2011.

[11] A. Van den Oord, O. Vinyals, et al. Neural discrete representation learning. In *Advances in Neural Information Processing Systems*, pages 6306–6315, 2017.

[12] H. Xiao, K. Rasul, and R. Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017.