Deep Clustering of Compressed Variational Embeddings

Suya Wu*, Enmao Diao*, Jie Ding† and Vahid Tarokh*

*Duke University
Durham, NC, 27701, USA
suya.wu@duke.edu
enmao.diao@duke.edu
vahid.tarokh@duke.edu

†University of Minnesota-Twin Cities
Minneapolis, MN 55455, USA
dingj@umn.edu

Motivated by the ever-increasing demands for limited communication bandwidth and low-power consumption, we propose a new methodology, named joint Variational Autoencoders with Bernoulli mixture models (VAB), that performs deep clustering on binary representations of data with state-of-the-art performance at a high-compression regime. The idea is to reduce the data dimension by Variational Autoencoders (VAEs) and group data representations by Bernoulli mixture models (BMMs). The model is trained in two steps: First, VAEs are jointly trained with BMMs, where a mixture of Bernoulli models provides a probabilistic distribution of latent representations. Subsequently, the classifier is updated by Bernoulli distributed samples produced in last step. This is optimized by the loss function consists of a reconstruction loss and a clustering loss.

Once jointly trained for compression and clustering, the model can be decomposed into two parts: a data vendor that encodes the raw data into compressed data, and a data consumer that classifies the received (compressed) data. In this way, the data vendor benefits from data security and communication bandwidth, while the data consumer benefits from low computational complexity. To enable training using the gradient descent algorithm, we propose to use the Gumbel-Softmax distribution, which provides a differentiable sampling mechanism that trains the neural network with a categorical reparameterization trick to resolve the infeasibility of the back-propagation algorithm when assessing categorical samples.

This framework explores the connection between directed probabilistic models and compressed data representations, therefore making it possible to consider interpretable and computationally efficient binary code. Experimental results showed that in the low Bits Per Pixel (BPP) regime, the clustering accuracy of VAB is comparable with other deep clustering methods, and its compression performance is much better than others. Through an approximate mixture of discrete probability models, the proposed solution requires less storage complexity and has the potential to reduce transmission bandwidths. To the best of our knowledge, what we present is the first methodology for simultaneous data compression and clustering in compressed domains.

This work was mostly done when Suya Wu was a student at the University of Minnesota. She is now with Duke University.

This work was supported in part by Office of Naval Research Grant No. N00014-18-1-2244.

2375-0359/20/$31.00 ©2020 IEEE
DOI 10.1109/DCC47342.2020.00051