Deep Clustering of Compressed Variational Embeddings

Suya Wu*, Enmao Diao*, Jie Ding† and Vahid Tarokh*1

*Duke University, Durham, NC, 27701, USA
†University of Minnesota-Twin Cities, Minneapolis, MN 55455, USA

Introduction

Clustering is a fundamental task with applications in medical imaging, social network analysis, bioinformatics, computer graphics, etc. However, applying classical clustering methods directly to high dimensional data may be computational inefficient and suffer from instability. On the other hand, lossy compression can achieve a high compression ratio to reduce computation and communication costs.

Performing clustering on compressed data is a potential solution to problems arising in storage, computing, and communicating unstructured and unlabelled image collections.

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.

Contributions

We propose a new method, namely joint Variational Autoencoders and Bernoulli mixture models (VAB) to learn the data representation for both clustering and compression simultaneously. The idea is to reduce the data dimension by Variational Autoencoders (VAEs) and group data representations by Bernoulli mixture models (BMMs). The model can be decomposed into two parts:

1. a data vendor that encodes the raw data into compressed data.

2. a data consumer that classifies the received (compressed) data.

Finally, the data vendor benefits from data security and communication bandwidth, while the data consumer benefits from low computational complexity.

Method

Considering the dataset $\mathbf{x}$ with N identically independently distributed (i.i.d) samples $\{x_i\}_{i=1}^{N}$ and $x_i \in \mathbf{R}^d$, we assume that the data is generated by some random process, involving an unobserved Bernoulli random variable $\mathbf{z}$ which belongs to one of k classes c. The joint distribution is formulated as

$$p_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{z}, \mathrm{c}) = p_{\boldsymbol{\theta}}(\mathrm{c}) p_{\boldsymbol{\theta}}(\mathbf{z} \mid \mathrm{c}) p_{\boldsymbol{\theta}}(\mathbf{x} \mid \mathbf{z}, \mathrm{c}), \tag{1}$$

where $\boldsymbol{\theta}$ stands for the generative model parameters. Along with this generative process, we assume

$$p_{\boldsymbol{\theta}}(\mathbf{x} \mid \mathbf{z}, \mathrm{c}) = p_{\boldsymbol{\theta}}(\mathbf{x} \mid \mathbf{z}), \tag{2}$$

The recognition model is $q_{\boldsymbol{\phi}}(\mathbf{z}, \mathrm{c} \mid \mathbf{x})$ as the variational approximation to the true intractable posterior and $\boldsymbol{\phi}$ stands for the recognition model parameters.

To perform clustering embedded in training VAEs, we optimize the lower bound $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi} \mid \mathbf{x})$ with respect to the model parameters and assign clusters simultaneously. The value of the evidence lower bound (ELBO) for VAB is,

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\phi} \mid \mathbf{x}) = E_{q_{\boldsymbol{\phi}}(\mathbf{z}, \mathrm{c} \mid \mathbf{x})}[&\log p_{\boldsymbol{\theta}}(\mathbf{x} \mid \mathbf{z}) + \log p_{\boldsymbol{\theta}}(\mathbf{z} \mid \mathrm{c}) \\ &+ \log p_{\boldsymbol{\theta}}(\mathrm{c}) - \log q_{\boldsymbol{\phi}}(\mathbf{z} \mid \mathbf{x}) - \log q_{\boldsymbol{\phi}}(\mathrm{c} \mid \mathbf{z})].\end{aligned} \tag{3}$$

To train the recognition model $q_{\boldsymbol{\phi}}$ with reparameterization trick [4], non-differentiable categorical samples $\boldsymbol{z}$ are replaced with Gumbel-Softmax estimators $\boldsymbol{y}$ [5]. It results to approximating $\nabla_\theta \mathbf{z}$ with $\nabla_\theta \mathbf{y}$ in backpass.

For each generated sample $\boldsymbol{y}^{(i,l)}$ corresponding to each input $\mathbf{x}^{(i)}$, we update the classes by

$$q_{\boldsymbol{\phi}}(\mathrm{c} \mid \boldsymbol{y}^{(i,l)}) = \frac{p_{\boldsymbol{\theta}}(\mathrm{c}) p_{\boldsymbol{\theta}}(\boldsymbol{y}^{(i,l)} \mid \mathrm{c})}{\sum_{c=1}^{k} p_{\boldsymbol{\theta}}(\mathrm{c}) p_{\boldsymbol{\theta}}(\boldsymbol{y}^{(i,l)} \mid \mathrm{c})}. \tag{4}$$

In addition to parameters $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$,$\boldsymbol{\pi}$ in $p_{\boldsymbol{\theta}}(c)$ and $\boldsymbol{\mu}_z$ in $p_{\boldsymbol{\theta}}(\mathbf{z} \mid \mathrm{c})$ are also trained as the model parameters.

Experimental Results

The performance of our method will be evaluated with classical clustering methods K-means and Gaussian mixture models (GMMs), as well as deep clustering methods on the hand-written digit image dataset MNIST [1].

Method	K-means	GMM	VaDE [2]	VAB
Best Clustering Acuracy (%)	55.37	42.22	95.30	71.69

Table 1: The clustering performance is compared on the MNIST test data.

Table 1 shows that VAB is much better than the classical methods, K-means and GMMs on image clustering. Although it is not comparable with the performance of VaDE [2], VAB achieves this result at a much lower bits per pixel as shown in Figure 1 and Figure 2, more suitable for compressed data.

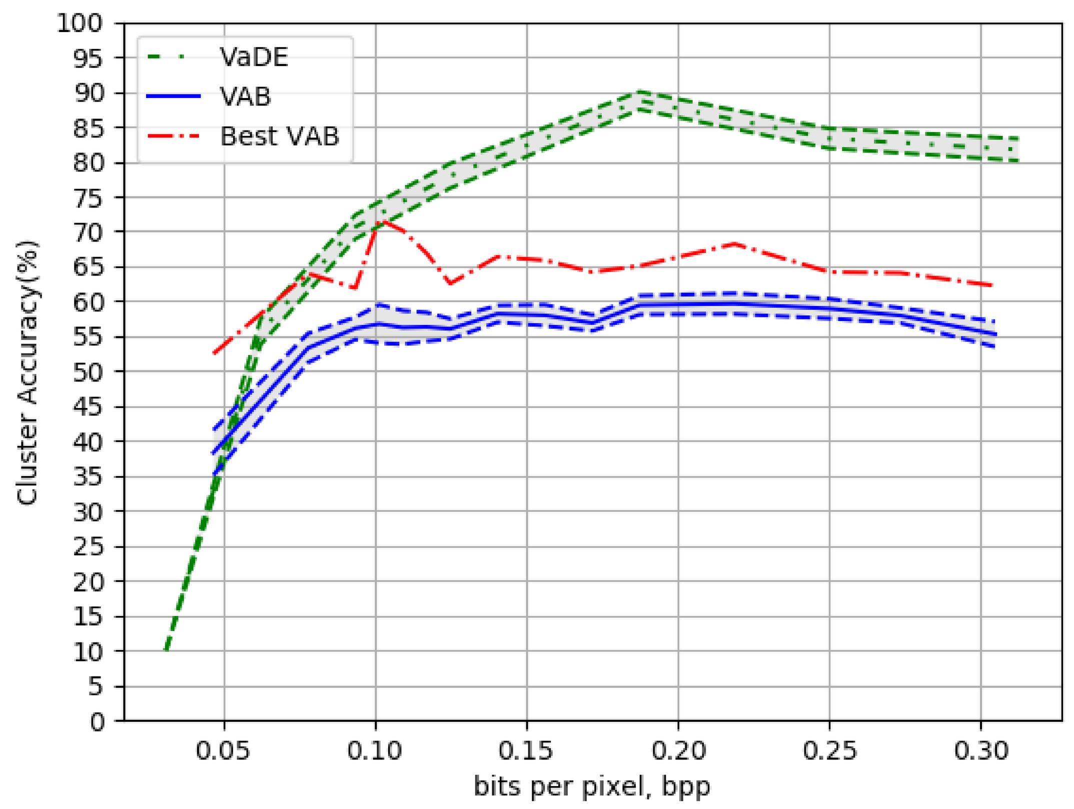

Figure 1: In the low BPP regime, the clustering accuracy of VAB is comparable with the result of VaDE [2].

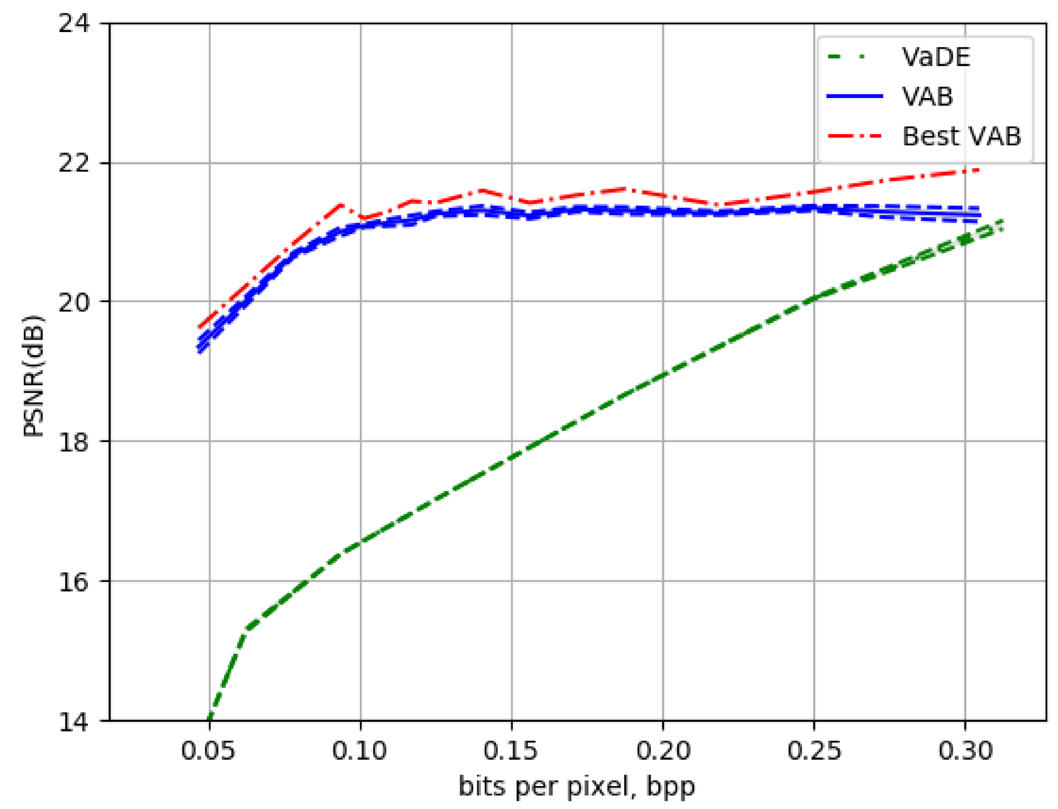

Figure 2: The compression performance of VAB is much better than VaDE [2] with lower BPP rate.

In Figure 1 and Figure 2, all results are averaged from 10 experiments and presented by the solid lines. The grey area between two dashed lines shows the standard errors of the mean from 10 replications.

Experimental Metrics

The following evaluation metric values are mentioned to perform a comparison.

1. Clustering accuracy (ACC): the clustering accuracy that varies from 0 to 1, and a higher clustering accuracy indicates a more accurate a clustering performance.

2. Peak signal-to-noise ratio (PSNR): the metric that measures the distance of the reconstructed image with the original image, and the higher the PSNR, the better the quality of the reconstruction.

3. Bits per pixel (BPP): the compressed rate, and more BPP indicates more memory required to store or display the image.

Experimental Details

The following experimental details are mentioned to perform a comparison.

1. K-means and GMMs are applied directly on raw image pixels, while the results of VaDE will be reported by re-running the code released from the original paper [2].

2. The architecture for the autoencoder is feedforward artificial neural networks. All layers are fully connected with a rectified linear unit (ReLU).

3. In training, Adam is applied to optimize the full set of parameters with $\beta = (0.9, 0.999)$ [3]. The learning rate is initialized at 0.001 and decreases every 10 epochs with a decay rate of 0.9 down to 0.0002.

References

[1] Yann LeCun and Corinna Cortes, "MNIST handwritten digit database," 2010.

[2] Zhuxi Jiang, Yin Zheng, Huachun Tan, Bangsheng Tang, and Hanning Zhou, "Variational deep embedding: A generative approach to clustering," *ArXiv*, vol. abs/1611.05148, 2016.

[3] Diederik P. Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014.

[4] Diederik P. Kingma and Max Welling, "Auto-encoding variational bayes," *CoRR*, vol. abs/1312.6114, 2013.

[5] Eric Jang, Shixiang Gu, and Ben Poole, "Categorical reparameterization with gumbel-softmax," *ArXiv*, vol. abs/1611.01144, 2016.