IEEE WCCI 2020

IEEE World Congress on Computational Intelligence

Virtual Conference – July 19-24, 2020

Unsupervised Clustering through Gaussian Mixture Variational AutoEncoder with Non-Reparameterized Variational Inference and Std Annealing

清華大學

Tsinghua University

Zhihan Li, Youjian Zhao, Haowen Xu, Wenxiao Chen, Shangqing Xu, Yilin Li, Dan Pei

SPONSORS:

IEEE Advancing Technology for Humanity

IEEE Computational Intelligence Society

IET

The International Neural Network Society (INNS)

EPS Evolutionary Programming Society

Outline

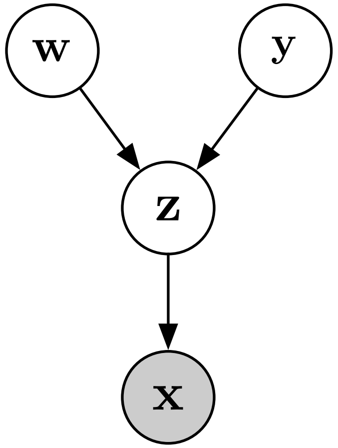
- Background
- Algorithm
- Experiments and Analysis
- Conclusion

Unsupervised Clustering

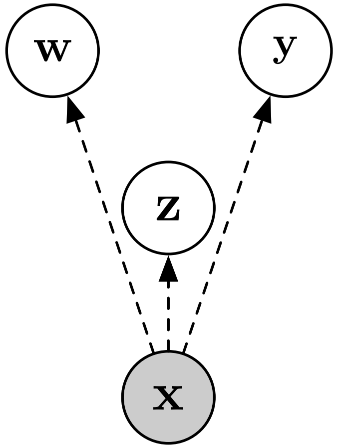
- Unsupervised clustering aims at grouping similar observations together in an unsupervised manner.
- Traditional methods: simple similarity measure and statistical distributions in data space.
- Learning good representations plays an important role in improving clustering accuracy.

Generative Model Based Methods

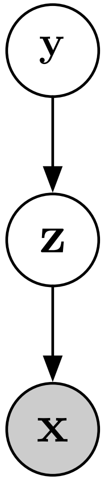
- Model the clustering assignments as a latent variable through deep generative models.



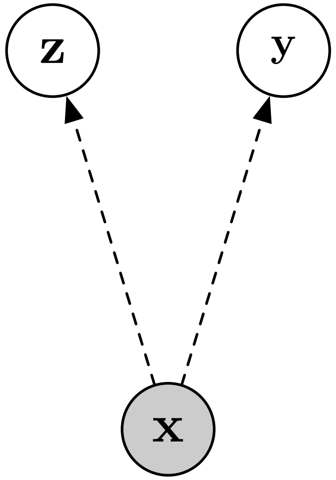
GMVAE Generative [1]



GMVAE Variational



VaDE Generative [2]



VaDE Variational

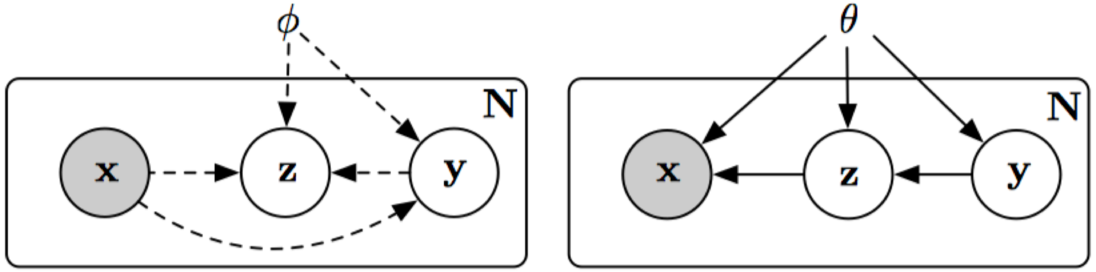
Mean-field approximation: $q(\mathbf{z}, \mathbf{y}|\mathbf{x}) = q(\mathbf{z}|\mathbf{x})q(\mathbf{y}|\mathbf{x})$ Additional Assumption: $q(\mathbf{y}|\mathbf{x}) = p(\mathbf{y}|\mathbf{z})$

Outline

- Background
- Algorithm
- Experiments and Analysis
- Conclusion

Our approach NVISA

- Directly using q(z|y,x) instead of using the mean-field approximation for unsupervised clustering.



Graphical model of NVISA. (left) Variational model (right) Generative model. Gray nodes are observations, white nodes are latent variables.

Straightforward factorization: $q(\mathbf{z}, y|\mathbf{x}) = q(\mathbf{z}|y, \mathbf{x})q(y|\mathbf{x})$

Variational Inference (VI) selection and std annealing trick

Candidate Variational Inference

EnumY

$$\mathcal{L}_{\text{ELBO}}(\mathbf{x};\theta,\phi)=\sum_{y=1}^{K} q_\phi(y|\mathbf{x})\,\mathbb{E}_{q_\phi(\mathbf{z}|y,\mathbf{x})}\Big[\log p_\theta(\mathbf{x},y,\mathbf{z})-\log q_\phi(y|\mathbf{x})-\log q_\phi(\mathbf{z}|y,\mathbf{x})\Big]$$

NVIL[3]

$$\nabla\,\mathcal{L}_{\text{ELBO}}(\mathbf{x};\theta,\phi)=\mathbb{E}_{q_\phi(\mathbf{z},y|\mathbf{x})}\Big[\nabla f(\mathbf{x},y,\mathbf{z})+\big(f(\mathbf{x},y,\mathbf{z})-C_\psi(\mathbf{x})-c\big)\nabla\log q_\phi(\mathbf{z},y|\mathbf{x})\Big]$$

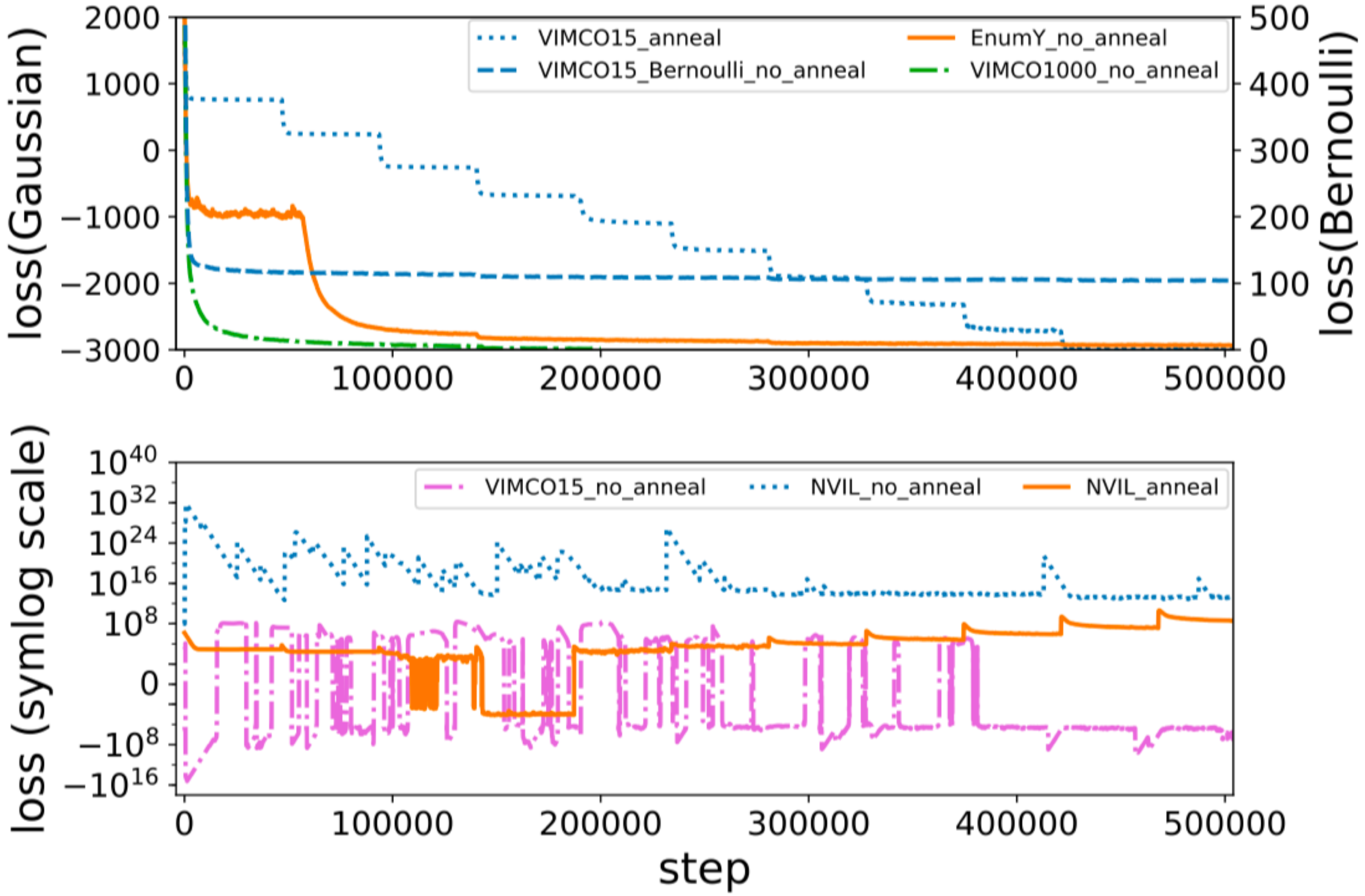
VIMCO[4]

$$\mathcal{L}^M(\mathbf{x};\theta,\phi)=\mathbb{E}_{(y,\mathbf{z})^{(1:M)}\sim q_\phi(\mathbf{z},y|\mathbf{x})}\left[\log\frac{1}{M}\sum_{m=1}^{M}\frac{p_\theta(\mathbf{x},y^{(m)},\mathbf{z}^{(m)})}{q_\phi(\mathbf{z}^{(m)},y^{(m)}|\mathbf{x})}\right]$$

VI Selection

Training loss of NVISA with different variational inference methods on MNIST.

The VIMCO method with our proposed std annealing trick below achieves the stable training and best clustering results.



Std Annealing

- Standard deviations annealing on Gaussian posteriors p(x|z) and q(z|y,x).

$$\boldsymbol{\sigma}_{\mathbf{x}}{}'(\mathbf{z};\theta) = \mathbf{SoftPlus}(\boldsymbol{\sigma}_{\mathbf{x}}(\mathbf{z};\theta)) + \epsilon_x$$
$$\boldsymbol{\sigma}_{\mathbf{z}}{}'(\mathbf{x},y;\phi) = \mathbf{SoftPlus}(\boldsymbol{\sigma}_{\mathbf{z}}(\mathbf{x},y;\phi)) + \epsilon_z$$

- Help the model converge.

- Alleviate the ``mixture model'' (trivial solutions) problem and improve the clustering performance.

Outline

- Background
- Algorithm
- Experiments and Analysis
- Conclusion

Experiments Setup

- Datasets: MNIST, fashion-MNIST, STL-10, CIFAR-10, CIFAR-100.
- Raw pixels for the first two datasets, features extracted by pretrained Resnet-50 on ImageNet for the others.
- Baseline methods: AE+GMM, VAE+GMM, DEC[5], GM-GAN[6], GMVAE[1], VaDE[2] and LTVAE[7].
- The last three methods also used VAEs with mixture priors for clustering.

Overall Performance

Method	MNIST	Fashion-MNIST	STL-10	CIFAR-10	CIFAR-100
AE+GMM	84.27	53.83	76.18	43.66	29.41
VAE+GMM	76.87	53.43	82.42	59.4	30.81
DEC [3]	84.30*	57.58	84.08	**70.31**	31.22
GM-GAN [6]	**99.24***	58.16*	-	-	-
GMVAE [7]	88.54*	50.15	65.72	49.79	14.46
VaDE [8]	94.46*	56.12	84.45*	59.69	23.58
LTVAE-GMM [21]	86.31	55.69	81.66	63.51	16.98
LTVAE-full [21]	86.32*	**61.32**	**90.00***	69.16	**38.52**
NVISA(best)	**98.40**	**66.14**	**91.22**	**76.96**	**38.79**
NVISA(avg$\pm$std)	**96.55$\pm$2.64**	**62.63$\pm$2.92**	**89.15$\pm$2.36**	**71.37$\pm$2.79**	**37.32$\pm$0.70**

Results marked by $*$ are excerpted from their original paper.
"-" means there is no published result and we didn't find any released code to produce it.

Use pretrained AE+GMM as an initialization can further improve the performance on MNIST to $98.04 \pm 0.17\%$.

Ablation Study

Unsupervised Clustering accuracy on MNIST, with different variational inference methods and other configurations.

The default settings are Gaussian posteriors, q(z|y,x) factorization and 15 VIMCO samples, if not specified.

Variational Inference	Method	ACC (%)
	GMVAE	88.54
	VaDE	94.46
EnumY	NVISA (Bernoulli)	95.12
	NVISA	78.43
	NVISA*	93.81
NVIL	NVISA	N/A
	NVISA*	N/A
VIMCO	NVISA	N/A
	NVISA (Bernoulli)	95.74
	NVISA (1000 samples)	71.42
	NVISA*(1000 samples)	91.3
	NVISA* (mean-field)	95.25
	NVISA*	**98.40**

"*" indicates the *std annealing* is applied.
"N/A" means the loss diverges and we cannot get a meaningful result.

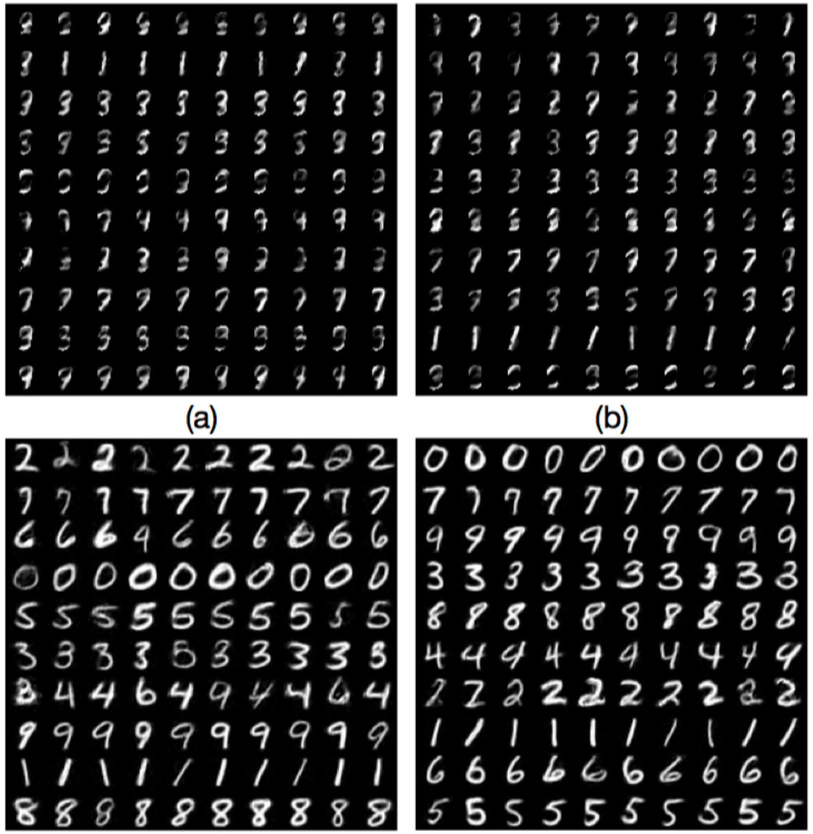
Sample Generation

Generated samples using NVISA with different VI methods. (a) EnumY (b) VIMCO with 1000 samples (c) EnumY with std annealing (d) VIMCO with 15 samples and std annealing (NVISA)

Ancestral sampling:

$$y \sim p_\theta(y), \mathbf{z} \sim p_\theta(\mathbf{z}|y), \mathbf{x} \sim p_\theta(\mathbf{x}|\mathbf{z})$$

Marginal distribution: $p_\theta(\mathbf{z}) = \sum_y p_\theta(\mathbf{z}|y)p_\theta(y)$



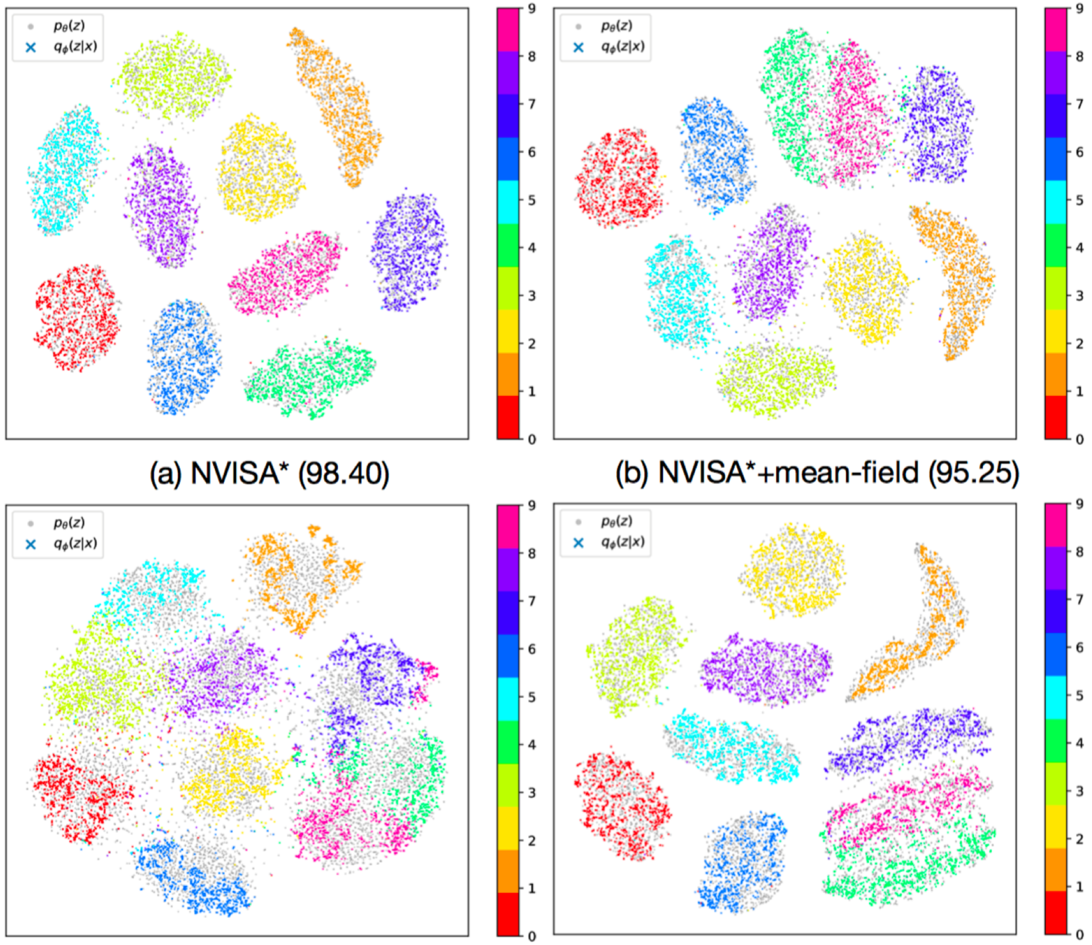
(a) (b) (c) (d)

Learned latent space

10,000 encoded test-set digits and 10,000 samples from the prior on MNIST, dimensionally reduced by t-SNE[8], colored by model predicted labels.

"*" indicates the std annealing is applied.

Values in parenthesis are the test-set clustering accuracy.



(a) NVISA* (98.40)
(b) NVISA*+mean-field (95.25)
(c) NVISA+EnumY (78.43)
(d) NVISA*+EnumY (93.81)

Outline

- Background
- Algorithm
- Experiments and Analysis
- Conclusion

Conclusion

- We propose an unsupervised clustering method NVISA based on Gaussian mixture VAE.
- We point out that directly using $q(z|y,x)$ instead of the mean-field approximation can benefit the clustering task.
- We use the non-reparameterized VI method VIMCO as well as our proposed std annealing trick to stabilize the training process and help form better embeddings in the latent space.
- NVISA overall outperforms all the baseline methods on the five benchmark datasets.
- More powerful priors can be incorporated into NVISA to enhance the representation capability. The recent self-supervised data augmentation technique can also be applied on NVISA to further improve the clustering accuracy.

IEEE WCCI 2020

IEEE World Congress on Computational Intelligence

Virtual Conference – July 19-24, 2020

Thank You!

Q&A?

lizhihan17@mails.tsinghua.edu.cn

References

[1] N. Dilokthanakul, P. A. Mediano et al., “Deep unsupervised clustering with gaussian mixture variational autoencoders,” arXiv preprint arXiv:1611.02648, 2016.

[2] Z. Jiang, Y. Zheng, H. Tan, B. Tang, and H. Zhou, “Variational deep embedding: an unsupervised and generative approach to clustering,” in Proceedings of the 26th International Joint Conference on Artificial Intelligence. AAAI Press, 2017, pp. 1965–1972.

[3] A. Mnih and K. Gregor, “Neural variational inference and learning in belief networks,” in ICML, 2014, pp. 1791–1799.

[4] A. Mnih and D. Rezende, “Variational inference for monte carlo objectives,” in ICML, 2016, pp. 2188–2196.

[5] J. Xie, R. Girshick, and A. Farhadi, “Unsupervised deep embedding for clustering analysis,” in International conference on machine learning, 2016, pp. 478–487.

References

[6] M. Ben-Yosef and D. Weinshall, "Gaussian mixture generative adversarial networks for diverse datasets, and the unsupervised clustering of images," arXiv preprint arXiv:1808.10356, 2018.

[7] X. Li, Z. Chen, L. K. M. Poon, and N. L. Zhang, "Learning latent superstructures in variational autoencoders for deep multidimensional clustering," in 7th ICLR, 2019.

[8] L. v. d. Maaten and G. Hinton, "Visualizing data using t-sne," Journal of machine learning research, vol. 9, no. Nov, pp. 2579–2605, 2008.