Learning Flexible Latent Representations via Encapsulated Variational Encoders

Wenjun Bai, Changqin Quan, Zhi-Wei Luo
Department of Computational Science
Kobe University, Japan
bwj@cs11.cs.kobe-u.ac.jp, quanchqin@gold.kobe-u.ac.jp, luo@gold.kobe-u.ac.jp,

Introduction

Representation learning – aims to capture certain aspects of the observed data – has fuelled majority of downstream AI applications. As an emerging technique, the usage of variational encoder [1] is a celebrated probabilistic approach to learn efficient latent representations in a pure unsupervised manner. However, based on the current structure of a variational encoder, learning of flexible latent representations is still a challenge task. To this end, we propose a novel form of variational encoder: encapsulated variational encoder (EVE) that allows grouping of two encoders in single scaffold to exploit their relations in representation learning.

Imposing certain constraints on this newly derived encapsulated variational encoder, e.g., the independence and equivalence constraints, it is capable of learning **diverged** and **converged** latent representations, respectively. We format the remaining article as follows. We chiefly render out our proposed EVE in the following *Technical Background* section. Then we demonstrate that via tuning a single hyper-parameter in our proposed EVE, the diverged and converged representations can be learned. Validated on MNIST and CIFAR10(4K) datasets, we show that these learned diverged and converged latent representations elevate the discriminative and generative modelling performance respectively.

Learning flexible Latent Representations

Technical Background

Different to a conventional variational encoder (Kingma and Welling 2013), in our proposed encapsulated variational encoders (EVE), we deliberately incorporate two variational encoders, e.g., the base and scaffolding one $q_{\phi_b}(z_b|x)$ and $q_{\phi_s}(z_s|x)$ to derive the analytic expression of our EVE as: $q((z_b, z_s)|x; \phi_b, \phi_s, \alpha)$. To measure the relations between two encoders, we adopt a discrepancy function with a positive defined hyper-parameter α, i.e., $\alpha \in \Re_+$ to quantify the difference between two encoded latent representation as: $\mathcal{R}_e(q(z_b|x), q(z_s|x)) = \alpha \cdot \frac{1}{L}\sum_{l=1}^{L} \exp\{-||z_b^l - z_s^l||^2\}$, where z_b^l and z_s^l are Monte-Carlo sampled encoding representations from two encoders respectively, i.e., $z_b^l \sim q_{\phi_b}(z_b|x)$ and $z_s^l \sim q_{\phi_s}(z_s|x)$. The graphical model presentation of this derived EVE is depicted in Figure 1(a).

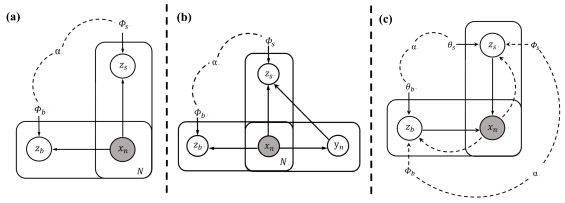

Figure 1: Graphical models of (a) the proposed encapsulated variational encoders (EVE); (b) the semi-supervised EVE, and (c) our derived encapsulated variational auto-encoder (EVAE). In (c), solid arrows denote probabilistic decoders, whereas dash arrows represent variational encoders.

Learning Converged Latent Representations

Armed with our proposed EVE, it is clear that the included hyper-parameter α exert the direct control over learned latent representations. In specific, as $\alpha \to 0$, it implies the equivalence constraint on two encoders in EVE. With sufficient training and a pre-defined small αs, latent representations from two encoders in EVE are learned to coincide with each other. As a result, driving $\alpha \to 0$ allows our derived EVE to learn **converged latent representations**.

These learned converged latent representations are featured in production of regularised latent representations to improve the discriminative performance of a semi-supervised model. To empirically validate this hypothesis, we construct a semi-supervised EVE, i.e., $q_{\phi_b,\phi_s}(z_b, z_s|x, y, \alpha)$, to see how these converged representations benefit the model in performing a designated classification task on CIFAR-10(4k) dataset. The graphical model presentation of this semi-supervised EVE can be found in Figure 1(b).

Demonstrated in Table 1, it demonstrates a negative correlation between the imposed α value and its induced discriminative performance. With a diminished α, the discriminative performance of our derived semi-supervised EVE achieved competitive empirical performance even in the face of state-of-the-art approaches.

[1] Here, we intentionly separate the usage of variational encoder that can be used in dependent of commonly assumed decoder.

Model	Test Error(%)
S3C (Goodfellow, Courville, and Bengio 2012)	28.1 ± 0.3
Ladder networks (Rasmus et al. 2015)	16.5 ± 0.3
VAT (Miyato et al. 2018)	24.1 ± 1.2
Semi-supervised EVAE ($\alpha = 1$)	46.4 ± 0.2
Semi-supervised EVAE ($\alpha = 0.1$)	21.7 ± 0.4
Semi-supervised EVAE ($\alpha = 0.01$)	19.5 ± 0.2
Semi-supervised EVAE ($\alpha = 0.001$)	18.2 ± 0.1
Benchmark:a supervised BNN	24.1 ± 0.2

Table 1: Test error rates obtained by various state-of-the-art semi-supervised models and our proposed semi-supervised EVE with varied α on CIFAR-10(4k) dataset. We ran each model five times, then averaged the performance. S3C, VAT and BNN refer to spike-and-slab sparse coding approach, virtual adversarial training(with $\epsilon = 1.0 \& \zeta = 1e^{-4}$) and a Bayesian neural network respectively.

Learning Diverged Latent Representations

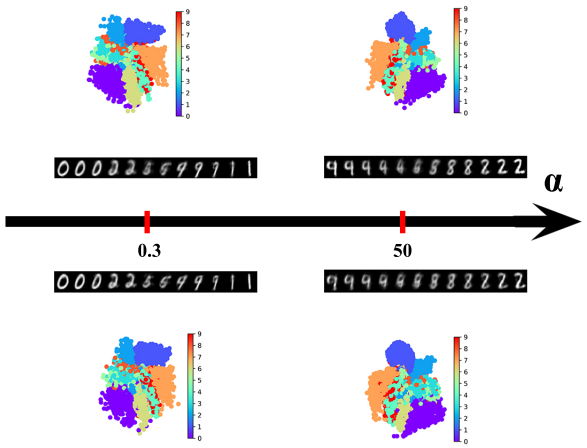

Figure 2: Qualitative results of EVAE with varied α on the binarised MNIST. Besides the rendered visualisations on generated pixel spaces (one that are positioned close to the α scale), we also render out the visualisations on latent spaces (ones that are positioned close to the α scale). The visualisations that are based on the base encoder in EVAE are grouped above the α scale, whereas these lower panel visualisations are based on the scaffolding encoder in EVAE.

In opposite to the preceding case, when $\alpha \to \infty$, it implies the independence constraint on two encoders in EVE to enforce the production of **diverged latent representations**. These learned, statistically independent latent representations are hypothesised to bring improvements on the generative modelling performance of a generative model.

To construct this generative model, we merge our proposed EVE with a corresponding probabilistic decoder, i.e., $p_{\theta_b,\theta_s}(x|z_b, z_s; \alpha)$, and a simply factorised joint prior, i.e., $p_{\theta_b}(z_b) \cdot p_{\theta_s}(z_s)$, it allows the coinage of a new variant of Variational Auto-Encoder: encapsulated Variational Auto-Encoder (EVAE). The graphical model demonstration of EVAE is depicted in Figure 1(c).

Reflected on Figure 2, with a smaller $\alpha(0.3)$, the generative modelling performance is largely constrained, rendering high similarity between generated digits from two encoders respectively. With a higher $\alpha(50)$, encoded latent representations from two encoders began to diverge, contributing to the diversification of generated digits, i.e., varied writing styles of the same digit.

Conclusion

In this abstract, we propose a novel form of variational encoder: encapsulated variational encoders(EVE). Rely on a tuneable hyper-parameter α, this derived EVE is able to learn **converged** and **diverged** latent representations of the observed data. These learned flexible latent representations can greatly improve the discriminative and generative modelling performance.

Moving forward, to extend the current formation of EVE to a non-parametric form, where the number of incorporated variational encoders reaches infinity, leads to a new avenue to explore in future studies.

Acknowledgements

This study is partially supported by National Natural Science Foundation of China under Grant No.61472117.

References

Goodfellow, I.; Courville, A.; and Bengio, Y. 2012. Large-scale feature learning with spike-and-slab sparse coding. *arXiv preprint arXiv:1206.6407*.

Kingma, D. P., and Welling, M. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Miyato, T.; Maeda, S.-i.; Ishii, S.; and Koyama, M. 2018. Virtual adversarial training: a regularization method for supervised and semi-supervised learning. *IEEE transactions on pattern analysis and machine intelligence*.

Rasmus, A.; Berglund, M.; Honkala, M.; Valpola, H.; and Raiko, T. 2015. Semi-supervised learning with ladder networks. In *Advances in Neural Information Processing Systems*, 3546–3554.