Continual Variational Autoencoder Learning via Online Cooperative Memorization

Fei Ye and Adrian G. Bors
Department of Computer Science, University of York, York YO10 5GH, UK

Abstract

Due to their inference, data representation and reconstruction properties, Variational Autoencoders (VAE) have been successfully used in continual learning classification tasks. However, their ability to generate images with specifications corresponding to the classes and databases learned during Continual Learning (CL) is not well understood. In this paper, we propose a new theoretical framework to analyze the forgetting behaviour of VAE models. Inspired by the theoretical analysis, we propose a new approach for continual learning.

Introduction

In this paper, we mainly study training VAE models under task-free continual learning (TFCL). We summarize our contributions as follows:

1. Our work is the first to provide theory insights for the forgetting behaviour of VAE under TFCL.

2, We propose the Online Cooperative Memorization (OCM) that can be used in any VAE variant with minimal modification and can also be extended to a dynamic expansion mixture approach to further enhance performance.

3.We propose a new sample selection approach for dynamically transferring selected samples from the STM to LTM without requiring any supervised signal. To our best knowledge, this is the first work to explore the kernel-based distance for the sample selection under TFCL.

4.The proposed sample selection approach can be used in both supervised and unsupervised learning without modifying the selection strategy

Methodology

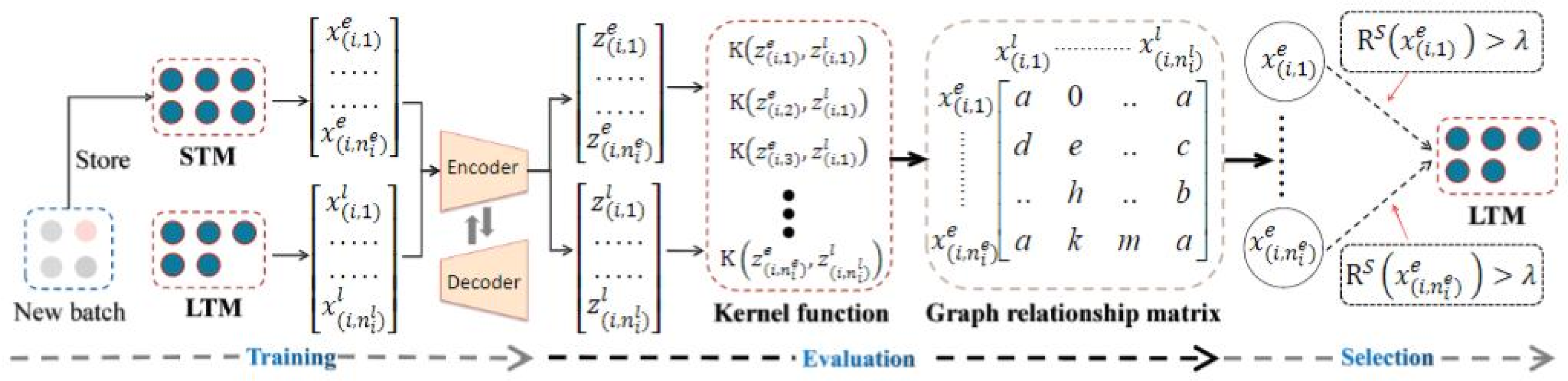

Fig.1 The training of OCM consists of three stages: the training, evaluation and sample selection.

The detailed algorithm implementation consists of three steps:

Step 1. Learning : At the training step, STM stores a new batch of samples into its memory buffer, while the model, consisting of a single VAE, is trained to update both STM and LTM using the VAE loss. Once the training is finished, we perform the next step.

Step 2. Evaluation : We use a kernel-based criterion which evaluates the correlection on a pair of samples:

$$\mathrm{K}(\mathbf{x}^e_{i,j}, \mathbf{x}^l_{i,u}) = \exp\left(-\frac{\left\|\mathbf{z}^e_{i,j} - \mathbf{z}^l_{i,u}\right\|^2}{2\alpha^2}\right) ,$$

We can accelerate the kernel-based criterion by the the matrix operation:

$$\mathbf{S}_i = \mathrm{F}_{\exp}\Big(-(\mathbf{Z}^e_i(-\mathbf{Z}^l_i)^{\mathrm{T}}) \odot (\mathbf{Z}^e_i(-\mathbf{Z}^l_i)^{\mathrm{T}})/2\alpha^2\Big),$$

where $\mathbf{S}_i(j,u) = \mathrm{K}(\mathbf{x}^e_{i,j}, \mathbf{x}^l_{i,u})$ describes the correlection between the j th sample of STM and u th sample of LTM. Based on this measure, we propose to evaluate the similarity between j th sample of STM and all samples of LTM:

$$\mathrm{R}^S(\mathbf{x}^e_{i,j}) = \frac{1}{n^l_i}\sum\nolimits_{k=1}^{n^l_i} \mathbf{S}_i(j,k)$$

Step 3. Selection : Then we can define a sample selection criterion:

$$\mathrm{R}^S(\mathbf{x}^e_{i,j}) > \lambda \Rightarrow \mathcal{M}^l_i = \mathcal{M}^l_i \cup \mathbf{x}^e_{i,j} .$$

Experiment

Methods	Split MNIST			Split Fashion			Split MNIST-Fashion		
	Log	Memory	N	Log	Memory	N	Log	Memory	N
VAE-ELBO-Random	-150.79	3.0K	1	-280.54	3.0K	1	-247.46	3.0K	1
LIMix [64]	-146.23	2.0K	30	-262.52	2.0K	30	-238.63	2.0K	30
CNDPM [35]	-120.71	2.0K	30	-257.56	2.0K	30	-236.79	2.0K	30
VAE-ELBO-OCM	-132.07	1.6K	1	-250.74	1.6K	1	-215.62	2.0K	1
VAE-IWVAE50-OCM	-127.11	1.6K	1	-247.90	1.6K	1	-224.34	2.0K	1
Dynamic-ELBO-OCM	**-115.89**	1.1K	5	**-237.69**	1.3K	10	**-187.49**	1.4K	10

Tab.1 The estimation of log-likelihood on all testing samples by using the IWVAE bound with 1000 importance samples.

Methods	*IS*	*FID*	Memory	N
VAE-ELBO-Random	3.84	116.26	1.0K	1
CNDPM [35]	4.12	95.23	1.0K	30
LIMix [64]	3.02	156.46	1.0K	30
VAE-ELBO-OCM	4.13	98.76	0.5K	1
Dynamic-ELBO-OCM	**4.16**	**92.99**	0.4K	3

Tab.2 IS and FID scores under Split CIFAR10.

Methods	Log	Memory	N
VAE-ELBO-Random	-239.71	3.0K	1
LIMix [64]	-226.63	2.0K	30
CNDPM [35]	-218.15	2.0K	30
VAE-ELBO-OCM	-201.31	2.0K	1
VAE-IWVAE50-OCM	-204.35	2.0K	1
Dynamic-ELBO-OCM	**-177.29**	1.5K	11

Tab.3 The estimation of log-likelihood on "Cross domain".

Methods	Split MNIST	Split CIFAR10	Split CIFAR100
finetune*	19.75 ± 0.05	18.55 ± 0.34	3.53 ± 0.04
GEM* [39]	93.25 ± 0.36	24.13 ± 2.46	11.12 ± 2.48
iCARL* [46]	83.95 ± 0.21	37.32 ± 2.66	10.80 ± 0.37
reservoir* [55]	92.16 ± 0.75	42.48 ± 3.04	19.57 ± 1.79
MIR* [4]	93.20 ± 0.36	42.80 ± 2.22	20.00 ± 0.57
GSS* [3]	92.47 ± 0.92	38.45 ± 1.41	13.10 ± 0.94
CoPE-CE* [13]	91.77 ± 0.87	39.73 ± 2.26	18.33 ± 1.52
CoPE* [13]	93.94 ± 0.20	48.92 ± 1.32	21.62 ± 0.69
CURL* [45]	92.59 ± 0.66	-	-
CNDPM* [35]	93.23 ± 0.09	45.21 ± 0.18	20.10 ± 0.12
Dynamic-OCM	**94.02 ± 0.23**	**49.16 ± 1.52**	**21.79 ± 0.68**

Tab.5 The classification accuracy of five indepdnent runs for various models on three datasets.

Conclusion

We introduce a new theoretical framework for providing insights into the forgetting behaviour of deep models based on VAEs under TFCL. Inspired by this result, we propose the Online Cooperative Memorization (OCM) that does not require any supervised signals and, therefore can be used to unsupervised learning. The empirical results demonstrate the effectiveness of the proposed OCM method.