Generative Oversampling for Imbalanced Data via Majority-Guided VAE

Ai, Q.; Wang, P.; He, L.; Wen, L.; Pan, L.; Xu, Z.

UESTC

Sep 22, 2023

Presented by Scott Sun from Duke B&B

Introduction

The authors propose a new generative oversampling approach called **Majority-guided VAE**

Goal: to address imbalance classification by oversampling a minority class(es) through a generative model. The new approach motivate the idea of learning **inter-class** relationships from the data to augment the synthesis of new data in minority class(es).

- modern advanced models are driven by large-scale, well-designed dataset; however, real-world data tend to be complex (i.e. class-imbalanced)
- by using a generative model, we are able to create fake but reasonable data to enhance the training process of downstream classification tasks; however, in some extreme cases, we only have a limited amount data for the minority class(es), which makes the training a deep generative hard
- this is highly relevant to my current project autism in children, where the event rate is only 2%.

Background: Variational Inference & VAE

Getting posterior $p_\theta(\boldsymbol{z}|\boldsymbol{x})$ is intractable, so we use $q_\phi(\boldsymbol{z})$ to approximate it by minimizing

$$\mathrm{KL}(q_\phi(\boldsymbol{z})||p_\theta(\boldsymbol{z}|\boldsymbol{x}))$$

derive evidence lower bound (ELBO) on board

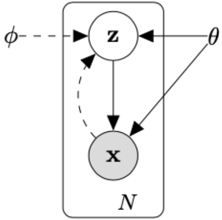


Figure: Variational Inference. Solid arrows denote an assumed data generating process; dashed arrows denote variational approximation (Kingma & Welling)

Background: Variational Inference & VAE

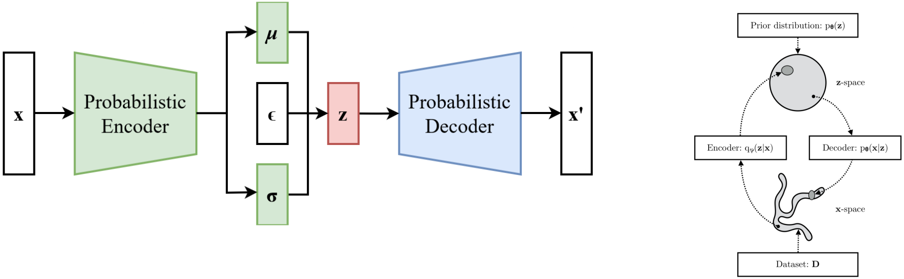


Figure: VAE & Reparameterization trick

Assume we have trained a VAE model that can successfully reconstruct $\boldsymbol{x}$ from $\boldsymbol{z}$. To synthesize new data, we can take random samples from the latent space and then apply the decoder to them.

Methods: Overview

Goal: Learn intra-class relationships and overcome the constraint of small sample size for the minority class(es) so that we can generate new minority samples under the guidance of a majority-based prior.

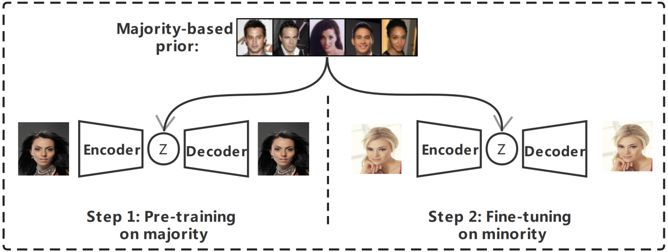


Figure: Majority-guided VAE

MGVAE achieves a one-to-one probabilistic "translation" from majority to minority. In another words, we can synthesize a minority sample from a observed majority sample.

Methods: Settup & Training Objective

Define majority class $\mathcal{X}^+ := \{x_n^+\}_{n=1}^{N^+}$ and minority class $\mathcal{X}^- := \{x_n^-\}_{n=1}^{N^-}$ s.t. $N^+ \gg N^-$. Let distributional parameter be Ψ. Given we want to borrow the knowledge from the majority class to learn the minority class, consider a conditional log-likelihood function and estimate it using transition distributions (note: think about KDE).

$$\log p(x_i^- | \mathcal{X}^+, \Psi) = \log \sum_{n=1}^{N^+} \frac{1}{N^+} T_\Psi(x_i | x_n^+)$$

Let $r_\phi(z|x^+)$ be a prior conditional on the majority, $q_\phi(z|x^-)$ be the encoder, and $p_\theta(x^-|z)$ be the decoder. An ELBO objective function (to be maximized) can be derived for each

$$\log p(x^- | \mathcal{X}^+, \Psi) \overset{\text{Jensen's}}{\geq} \boxed{\mathbb{E}_q \log p(x^-|z) - \mathbb{E}_q \log \frac{q(z|x^-)}{\sum_{n=1}^{N^+} r(z|x_n^+)/N^+}} = O(\Psi, \mathcal{X}^+; x^-)$$

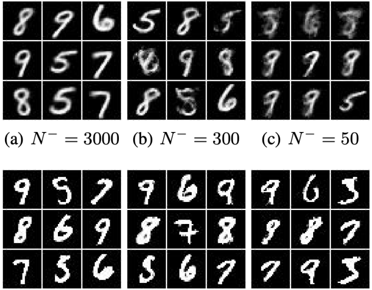
Methods: Pre-training & Fine-tuning

Pre-training: Training a model $\mathcal{M}_{\text{pre}}$ over the objective $O(\Psi, \mathcal{X}^+; x^+)$

Fine-tuning: Elastic Weight Consolidation (EWC) regularization. Update $\mathcal{M}_{\text{pre}}$ with all minority data with regularized objective function O_{EWC}. EWC preserves the important parameters by penalizing parameter change, and importance is measured by **Fisher Information** $F_\psi = \mathbb{E}\big[-\frac{\partial^2}{\partial \psi^2} L\big]$.

$$O_{\text{EWC}} = O(\Psi, \mathcal{X}^+; x^-) + \lambda \sum_{\psi_i \in \Psi} F_\psi \cdot (\psi_i - \psi_{+,i})^2$$

Experimental Results, Part 1



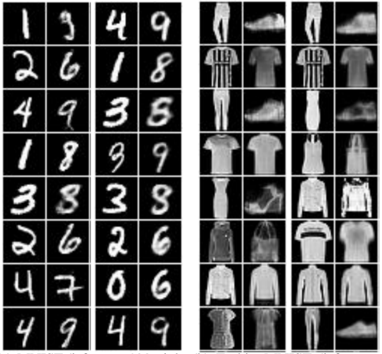
(a) $N^- = 3000$ (b) $N^- = 300$ (c) $N^- = 50$

(d) $N^- = 300$ (e) $N^- = 50$ (f) $N^- = 10$

Figure 2: The sampling results of MGVAE. (a-c): Without the Pre-training and Fine-tuning; (d-f): With the Pre-training and Fine-tuning. The label under each figure is the size of the downsampled minority samples.

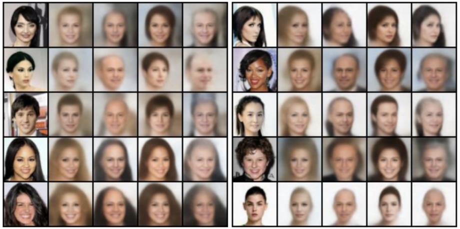
Figure: Experiment results on a modified MNIST dataset, where all 0-4 are used as the majority and downsampled 5-9 as the minority.

Experimental Results, Part 2



(a) MNIST (left: $\rho = 600$; right: $\rho = 100$)

(b) FashionMNIST (left: $\rho = 600$; right: $\rho = 100$)



(c) CelebA (from left to right: black,blonde,bald,brown,and gray)

Figure 3: Samples from MGVAE. In each plate, the first column is the reference majority sample, and the rest columns are the corresponding generated minority ones. Best viewed in color.

Figure: MG-VAE can generate meaningful outputs.

Recommendations

Is it worth reading? Yes/Maybe (depending on the research problem)

- the article gives a list of possible off-shelf oversampling approaches
- there are a few sections & figures particularly serve to illustrate the effectiveness of MGVAE
- the article covers implementing fine-tuning generative models; it worth paying attention to this important concept/technique

Is it worth implementing? Yes!

- the idea is straightforward to follow
- they released their GitHub repo: https://github.com/Aiqz/MGVAE
  (has good code design but is lack of documentation)