Elastic InfoGAN

(draft)

Bingbing Hu

November 27, 2019

SIST, ShanghaiTech

Outline

1. Intro
2. Problem Formulating
3. Experiments

Intro

Elastic-InfoGAN learns to disentangle object identity from other low-level aspects in class-imbalanced datasets.

Experiments show that

- better disentanglement of object identity;
- better approximation of class imbalance in the data.

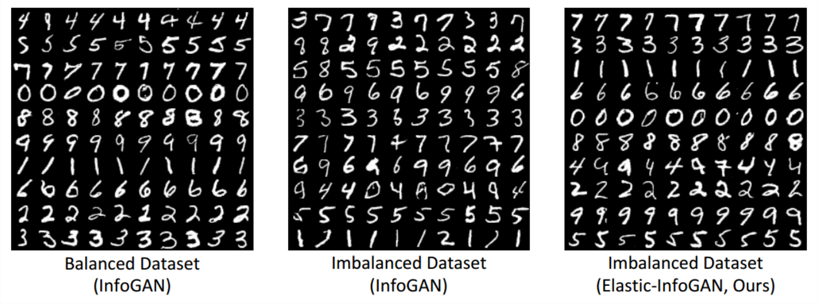

Figure 1: A comparison of results.

There are literatures that deal with disentangled representation learning, also learning from imbalanced data. But we do the both!

In other words, we do generative, disentangled representation learning in a unsupervised setting based on imbalanced data.

Problem Formulating

N unlabeled images from k different classes:

$$\mathcal{X} = \{x_1, \cdots, x_N\}$$

Our goal:

- Learn a generative model G that can disentangle *object category* from other aspects;
- recover the data distribution (maybe imbalanced).

Recall of InfoGAN,

$$\min_{G,Q} \max_{D} V_{\mathsf{InfoGAN}}(D, G, Q) = V_{\mathsf{GAN}}(D, G) - \lambda_1 L_1(G, Q)$$

$$L_1(G, Q) = \mathbb{E}_{c \sim P(c), x \sim G(z,c)}[\log Q(c|x)] + H(c)$$

Elastic InfoGAN

Two augmentations to InfoGAN,

1. learning the prori distribution,
2. learning object identities.

Elastic InfoGAN

Two augmentations to InfoGAN,

1. learning the prori distribution,
2. learning object identities.

For the 2nd, we enforce *identity-preserving transformation* invariance in the learned latent variables so that the resulting disentanglement favors groups that coincide with object identities.

Learning prori distribution

Replace fixed categorical distribution by Gumbel-Softmax distribution. For a categorical distribution $(p_1, \cdots, p_k)$, sample k-dimensional vector $\boldsymbol{c}$ where

$$c_i = \frac{\exp((\log(p_i) + g_i)/\tau)}{\sum_{j=1}^{k} \exp((\log(p_j) + g_j)/\tau)} \quad \text{for } i = 1, \ldots, k.$$

- g_i, g_j: samples from $Gumbel(0, 1)$,
- τ: softmax temperature, controls the degree to which samples from Gumbel-Softmax resemble the categorical distribution.

Learning prori distribution

This should be enough to handle the imbalance, once the true imbalance gets reflected in the class probabilities, the category code should disentangle object category. But experiments don't show that well!

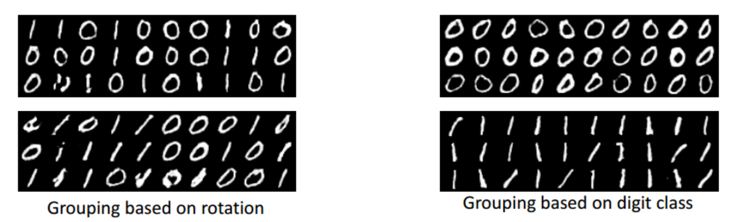

Figure 2: Unsupervised grouping can focus on non-categorical attributes such as rotation of the digit.

Learning object identities

Here comes the second augmentation: learning object identities.

We want our model

- focus on high level object identity,
- invariant to low level factors (rotation, thickness, etc.).

What we do

For any real image $x \sim P_{\text{data}}(x)$, we apply a set of transformations δ to obtain a transformed image $x' = \delta(x)$. Then we introduce a loss

$$L_{\text{trans}}(Q) = \mathsf{d}(Q(c_x|x), Q(c_{x'}|x')),$$

where $\mathsf{d}(\cdot)$ is a distance metric.

Note that

- these transformations are not learned,
- the object identity won't change after these transformations under a human view.

One more thing

$Q(c|x)$ should have low entropy (peaky class distribution) either for $x \sim P_{\text{data}}$ or $\tilde{x} \sim P_G$.

- For $\tilde{x} \sim P_G$, this is ensured by InfoGAN framework.
- For real images $x \sim P_{\text{data}}$, L_{trans} isn't sufficient to ensure this.

We hence add an additional entropy loss which forces c_x and $c_{x'}$ to have low entropy (s) class distributions:

$$L_{\text{ent}}(Q) = \mathsf{s}(Q(c_x|x)) + \mathsf{s}(Q(c_{x'}|x')).$$

The whole loss becomes

$$\min_{G,Q} \max_{D} L_{\mathsf{final}} = V_{\mathsf{InfoGAN}}(D, G, Q) + \lambda_2 L_{\mathsf{trans}}(Q) + \lambda_3 L_{\mathsf{ent}}(Q)$$

$$V_{\mathsf{InfoGAN}}(D, G, Q) = V_{\mathsf{GAN}}(D, G) - \lambda_1 L_1(G, Q).$$

- V_{InfoGAN} plays the role of generating realistic images and associating the latent variables to correspond to some factor of variation in the data,
- L_{trans} will push the discovered factor of variation to be close to object identity,
- L_{ent} ensures Q behaves similarly for real and fake image distributions.

The whole architecture looks like this:

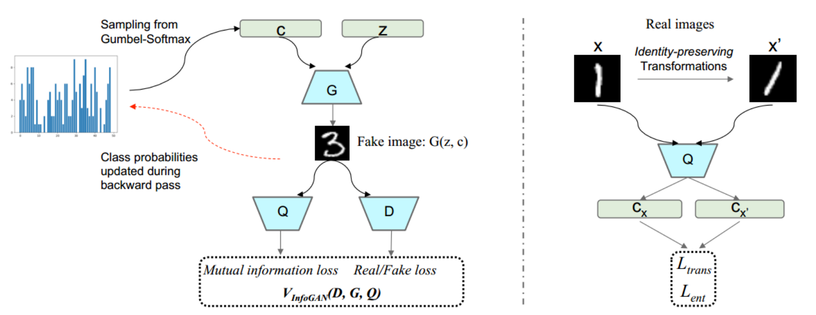

Figure 3: Elastic InfoGAN architecture.

Experiments

Baselines and evaluation metric

Different settings:

- *Uniform InfoGAN*: fixed and uniform categorical distribution.
- *Ground-truth InfoGAN*: fixed, imbalanced categorical distribution that reflect the true imbalance.
- *Ground-truth InfoGAN + Transformation constraint*: ditto but add L_{trans}.
- *Gumbel-softmax*: learnable categorical distribution.
- *Gumbel-softmax + Transformation constraint*: ditto but add L_{trans}.
- *Gumbel-softmax + Transformation constraint + Entropy loss (Elastic-InfoGAN)*: ditto but add L_{ent}.
- *JointVAE*: another paper.

Baselines and evaluation metric

Evaluation metrics:

- *Average entropy*
- *Normalized Mutual Information*
- *Root Mean Square Error*

More on paper. . .

Thank you!