Style Transfer from Non-Parallel Text by Cross-Alignment

1 Idea

The authors aim to perform style transfer on language using non-parallel corpora by separating content from style. They re-align the latent spaces to perform three tasks: sentiment modification, decipherment of word-substitution ciphers, and recovery of word order.

2 Method

The authors' method involves learning an encoder that takes a sentence and its original style indicator as input, and maps it to a content representation devoid of style. This representation is then decoded by a style-dependent decoder.

2.1 Notation

$y \rightarrow$ latent style variable
$z \rightarrow$ latent content variable
$x \rightarrow$ data point generated from the conditional distribution $P(x|y,z)$

There are two non-parallel corpora $X_1 = \{x_1^{(1)}...x_1^{(n)}\}$, drawn from $p(x_1|y_1)$ and $X_2 = \{x_2^{(1)}...x_2^{(n)}\}$, drawn from $p(x_2|y_2)$

We want to estimate the style transferred distributions $p(x_1|x_2;y_1,y2)$ and $p(x_2|x_1;y_1,y2)$

The authors propose a constraint that x_1 and x_2's marginal distributions can only be recovered if for any different styles $y, y' \in Y$, distributions $p(x|y)$ and $p(x|y')$ are different, which is a fair assumption to make because if $p(x|y)$

$= p(x|y')$, then the style changes would be indiscernible. They also prove that if the content z is sampled from a centered isotropic distribution, the styles cannot be recovered from x, but in the case of z being a more complex distribution like a Gaussian mixture, then the affine transformation that converts y, z into x can be recovered.

Instead of the KL divergence loss, the authors propose aligning the distributions $E(x_1, y_2)$ and $E(x_2, y_2)$ where E is the encoder function.

3 Observations

- Despite the corpora being non-parallel, the content of both corpora is mostly homogenous.
- The authors cite the reason for not using VAEs for this task as the utility of having rich and unperturbed representations, which VAEs do not possess, because of the ELBO objective which forces the latent representation to be consistent with a prior distribution.
- The sentiment transfer model succeeds in retaining content 41.5% of the time.

References