Abstract

Generative adversarial networks (GANs) have achieved significant success in generating real-valued data. However, the discrete nature of text hinders the application of GAN to text-generation tasks. Instead of using the standard GAN objective, we propose to improve text-generation GAN via a novel approach inspired by optimal transport. Specifically, we consider matching the latent feature distributions of real and synthetic sentences using a novel metric, termed the feature-mover’s distance (FMD). This formulation leads to a highly discriminative critic and easy-to-optimize objective, overcoming the mode-collapsing and brittle-training problems in existing methods.

Introduction

The main contributions of this paper are as follows:

• A new GAN model based on optimal transport is proposed for text generation. The proposed model is RL-free, and uses a so-called feature-mover’s distance as the objective.
• We evaluate our model comprehensively on unconditional text generation. When compared with previous methods, our model shows a substantial improvement in terms of generation quality based on the BLEU statistics and human evaluation. Further, our model also achieves good generation diversity based on the self-BLEU statistics.
• In order to demonstrate the versatility of the proposed method, we also generalize our model to conditional-generation tasks, including non-parallel text style transfer, and unsupervised cipher cracking.

Feature Mover’s distance

- **Earth-Mover’s Distance (EMD):** consider two probability distribution \( x \sim \mu \) and \( y \sim \nu \). EMD can be then defined as:
  \[ D_{EMD}(\mu, \nu) = \inf \sum_{i,j} \left| x_{ij} - y_{ij} \right| c(x_{ij}, y_{ij}), \]
  where \( \Pi(x, y) \) denotes the set of all joint distributions \( (x, y) \) with marginals \( \mu(x) \) and \( \nu(y) \), and \( c(x, y) \) is the cost function (e.g., Euclidean or cosine distance). Intuitively, EMD is the minimum cost that \( y \) has to transport from \( x \) to \( y \).

- **Feature Mover’s Distance (FMD):**
  \[ D_{FMD}(\mathbb{P}_1, \mathbb{P}_2) = \min_{\Pi} \sum_{i,j} \left| x_{ij} - y_{ij} \right| c(x_{ij}, y_{ij}), \]
  where \( \sum_{j} T_{ij} = \frac{1}{n} \) and \( \sum_{i} T_{ij} = \frac{1}{n} \) are the constraints, and \((\cdot, \cdot)\) represents the Frobenius dot-product.

We use the Inexact Proximal point method for Optimal Transport (IPOT) algorithm to compute the optimal transport matrix \( T \):

Algorithm 1 IPOT algorithm [1]

1. **Input:** batch size \( n \), \( \{f_j\}_{j=1}^{m} \), learning rate \( \eta \), maximum number of iterations \( N \).
2. for \( itr = 1 \ldots N \) do
3.  for \( j = 1 \ldots m \) do
4.     \( \sigma = \mathbb{1}_{n} \cdot T^{-1} \cdot \mathbb{1}_{n} \)
5.     for \( i = 1 \ldots n \) do
6.         \( Q = A \otimes T_{ij} \otimes \mathbb{1}_{n} \) Hadamard product
7.         \( \delta = \frac{\mu_i}{\sigma_i} \)
8.         \( \Delta = \delta \cdot Q \) (Diagonal)
9.     end for
10. end for
11. Return \( C, T \)

Adversarial distribution matching with FMD

To integrate FMD into adversarial distribution matching, we propose to solve the following minimax game:

\[
\min_{\theta} \max_{\mathbb{P}} D_{FMD}(\mathbb{P}, \mathbb{Q}) = \max_{\mathbb{P}} \min_{\theta} D_{FMD}(\mathbb{P}, \mathbb{Q})
\]

...