Conditional Generative Adversarial Nets

Report

Title: Conditional Generative Adversarial Nets
Authors: Mehdi Mirza, Simon Osindero
Institutions: Universite de Montreal, Flickr/Yahoo Inc.
Published: 2014

Introduction

This paper extends the Generative Adversarial Networks (GANs) framework by introducing Conditional GANs (cGANs), which allow the model to be conditioned on auxiliary information such as class labels. This conditioning enables control over the modes of the data being generated, providing more specific and directed data generation capabilities.

Methodology

Generative Adversarial Networks (GANs):

- GANs consist of two neural networks: a generative model G that captures the data distribution and a discriminative model D that estimates the probability that a sample came from the training data rather than G.
- The models are trained simultaneously in a minimax two-player game, where G tries to maximize the probability of D making a mistake.

Conditional GANs (cGANs):

- cGANs extend GANs by conditioning both the generator and discriminator on additional information y.
- This additional information y can be any kind of auxiliary data, such as class labels or other modalities.
- The objective function for cGANs is:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x|y)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z|y)))]$$

Architecture:

- The generator combines the prior input noise pz(z) with the auxiliary information y to form a joint hidden representation.

- The discriminator receives both the data sample x and the auxiliary information y as inputs.

Experiments and Results

MNIST Dataset:

- The authors trained a cGAN on the MNIST dataset, conditioning the generation of digits on their class labels.
- The generator used a noise prior z of dimensionality 100 and class labels encoded as one-hot vectors.
- The model demonstrated the capability to generate MNIST digits conditioned on class labels, achieving comparable log-likelihood estimates to other network-based approaches.

Multi-Modal Learning on Flickr Data:

- The authors used the MIR Flickr 25,000 dataset for multi-modal learning, conditioning on image features to generate descriptive tags.
- They pre-trained a convolutional model on the ImageNet dataset and used a skip-gram model to learn word vectors from user-generated metadata.
- The cGAN was able to generate plausible tags for images, demonstrating the potential of conditional models for multi-modal learning.

Relevance to Synthetic Data Creation for OCR

Context

- **Monlam AI's Data**: Contains lines of Tibetan text images and their transcription, not individual glyphs.

Key Insights for Synthetic Data Creation

1. **Controlled Data Generation**:
    - cGANs allow the generation of synthetic text images conditioned on specific transcriptions, which is highly relevant for creating labeled data for OCR training. This control over data generation can help produce a diverse and representative dataset.
2. **Multi-Modal Learning**:
    - By conditioning on different types of data (e.g., textual descriptions, transcriptions), cGANs can generate synthetic text that closely matches the characteristics of real-world data. This multi-modal capability can improve the robustness and accuracy of OCR models.
3. **Augmenting Data for Rare Classes**:

    - cGANs can be particularly useful for generating synthetic data for rare or underrepresented classes in the dataset, ensuring a more balanced training set for OCR models.

Conclusion

The conditional generative adversarial nets framework provides a powerful method for creating high-quality synthetic data with specific characteristics. By leveraging cGANs, Monlam AI can generate realistic and diverse Tibetan text images conditioned on transcriptions, enhancing the training dataset for OCR and improving overall model performance. The insights from this paper offer valuable methodologies for controlled and targeted synthetic data generation, addressing data scarcity and boosting the robustness of OCR systems.