Introduction

We generate realistic handwriting by combining stacked LSTMs, an Attention mechanism, Mixture Density Networks (MDNs), and Professor Forcing—a recent technique for training RNNs as sequential generative models.

Dataset

We used the IAM online handwriting database to train our model [3]. It contains sequences of pen tip positions and end of stroke tokens annotated with ASCII characters. Total of 86,272 words in 13,049 lines from 221 writers. Sample line below.

A MOVE to stop Mt. Gotskel

Handwriting Generator [1]

Network models $p(x|c)$.

$c$: ASCII character seq.

Input to network.


$x$: pen tip position + <eos> token seq. Sampled.

**Loss:** [2] train to predict $x$ and fool discriminator.

$\mathbb{E}_{(c,x)\sim \text{data}} [-\log p(x|c)]$

$C_{\text{free}}(\theta_{\text{gen}}|\theta_{\text{disc}}) =
\mathbb{E}_{c\sim \text{data}, x\sim P_{\text{gen}}(x|c)} [-\log D(g(c, x, \theta_{\text{gen}}), \theta_{\text{disc}})]$

$C_{\text{gen}} = \mathbb{E}_{c\sim \text{data}, x\sim P_{\text{gen}}(x|c)} [-\log D(g(c, x, \theta_{\text{gen}}), \theta_{\text{disc}})]$

Professor Forcing [2]

Professor Forcing applies a GAN like framework to training RNNs. The similarity in the dynamics of the two sampling modes leads to longer and more robust sequences.

Discriminator

Network predicts if generator is sampling in teacher forcing mode – $D(g)$. $g$: seq. of hidden states of generator. Here, parameters of $p$. Input to network. FC weights shared across time.

**Loss:** [2]

$C_{\text{disc}}(\theta_{\text{disc}}|\theta_{\text{gen}}) =
\mathbb{E}_{(c,x)\sim \text{data}} [-\log D(g(c, x, \theta_{\text{gen}}), \theta_{\text{disc}})]$

$+ \mathbb{E}_{x\sim P_{\text{gen}}(x|c)} [-\log (1 - D(g(c, x, \theta_{\text{gen}}), \theta_{\text{disc}}))]$

Results and Discussion

Goal: Generate longer, more legible sequences.

$c$ = “Joe Moses opened up a large coloured”

Able to generate longer sequences.

$c$ = “to the nuclear tests”

References


