

リカレントニューラルネットワークを用いた動画像生成モデル

Motion image generation model using Recurrent Neural Networks

1W143149-6 横川 奨悟 指導教員 尾形 哲也 教授

YOKOKAWA Syogo

Prof. OGATA Tetsuya

概要: 本研究ではリカレントニューラルネットワークを用いた動画像生成モデルの提案・実装し, モデルの学習結果に対する評価を行った. 自然画像や絵, 自然言語を生成するモデルは提案されているが, 動画像生成を行うモデルはまだ少ない. Goodfellow et al.[1]の提案した GAN では画像生成を行う効果的な手法ではあったが, ステップ数の多いモーションデータの生成はできなかった. そのため本研究では GAN のモデルを参考にリカレントニューラルネットワークを利用することで動画像を生成できるモデルの設計を行った. またモデルによる学習を行い, 有効性を確認した.

キーワード: 動画像生成, リカレントニューラルネットワーク, 人工知能, 学習

Keywords: Generation Motion Images, Recurrent Neural Networks,

1. はじめに

これまで, 動画を制作することは人間にしかできないと考えられてきた. 以前は自然画像や絵も同様に考えられてきたが, 機械による生成が実現している. 動画においても機械により生成したいと考えた. 人による動画の制作は人や機材等のコストに加え, 短い動画制作に対し膨大な時間を要するものもある. また機械が制作することでGoogleのDeep Dreamのような, 人が想像していないような映像を作り出せる可能性を持つと考える. Goodfellow et al. [1]やBojanowski et al. [2]の研究ではConvolutional Neural Networksを用いていたが, 動画のようなステップ数の多いデータへの対応は難しい.

このような問題点を解決するため, 本研究は時系列データに適している Recurrent Neural Networks[3]を用い, 時系列データを学習・生成できるモデルを提案する. また Long short-term memory を用いることで長期依存の時系列データへも対応できるようにした.

本研究では提案モデルにより学習を行った. リサージュカーブの学習によりモデルの有効性を確認し, 動画像データの学習を行った. また, 得られた結果よりモデルに対する評価, 考察を行った.

2. 動画像生成モデル

本研究で提案するモデルは, 画像生成モデル GAN の特性を持ち, 学習を不安定にしていた潜在変数を制限した. また, コストがかからず, 生成したデータが複数のデータを生成できるモデルを目指した. [1]

提案したモデルの目的関数は以下である:

$$\min_{\theta} \left[\sum_{t=1}^T L(y_t(x_{t=0}, z), x_t) + R(z) \right]$$

まず, 時系列データ X を用意し, 時間ごと ($t=1, \dots, T$) で区切ったもの $(x_0, \dots, x_T) \{x_t \in X\}$ を用意する. 次に x_0 と潜在変数 z を入力として, 学習によって変更可能なパラメータ θ を持つモデルの前向き計算で逐次的に y_t を得る. その結果生成されたデータが y となる. 図1が式をモデル化したものである.

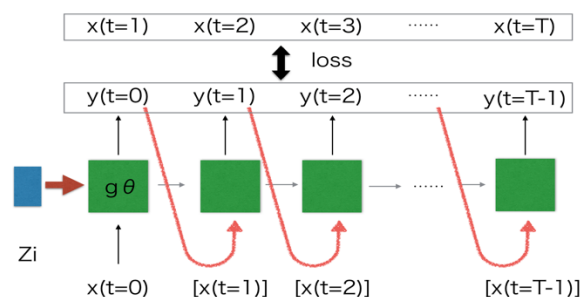


図1 提案モデル

潜在変数 z の正規化にはノルムを用い, ハイパーパラメータ a, p を変動させることで潜在変数 z に対して制限をかけることができる. これを式で表

すと以下になる：

$$R(z) = a\|z\|^p$$

3. 学習結果

3.1 リサーチカーブ

図2は36本のカーブを学習させた結果である。赤線が学習したカーブ，青線が教師データである。図3は生成したカーブ間の補間である。

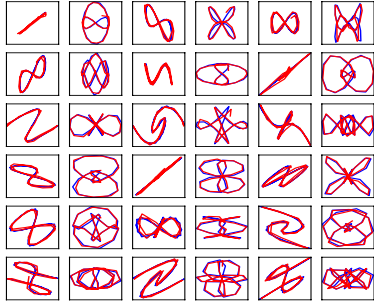


図2 36本の学習結果

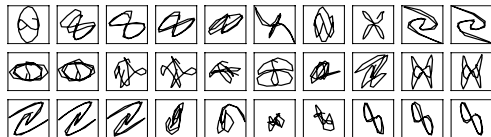


図3 図2の補間結果

学習自体は成功したが補間は十分な結果が得られなかった。そのためカーブの本数を256本に増やした学習結果が図4，その補間結果が図5である。

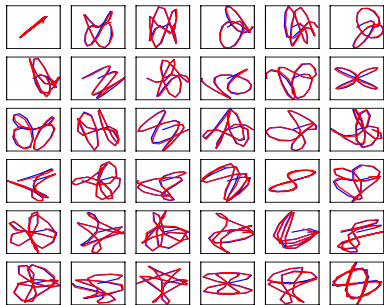


図4 256本の学習結果

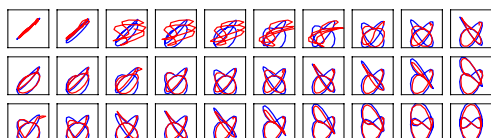


図5 図4の補間結果

補間は条件によるが学習は成功したため動画の学習を行った。

3.2 動画シーケンス学習

学習結果とデータセットを比較すると大きな誤差は見られず，モデルの学習は成功していた。潜在変数の分布においても，中心から見

て各色が混在せず近いカラーが分布していた。補間においても定量的な判断ではないが，同画像の生成が確認できモデルの有効性を確認することができた。

4 考察

3.1において補間が成功・失敗が生じた理由は潜在変数の位置と考える。図6はカーブを256本学習させた時の潜在変数の分布図である。

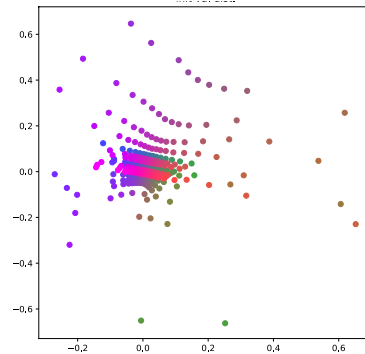


図6 潜在変数の分布図

補間の対象とするカーブの潜在変数が同じカラー上に存在している場合のみ成功すると考えられる。また，シーケンス学習においては同様のフレームワークで動画生成ができ，未知のデータ生成もできた。

5. 結論

本研究では動画生成モデルの提案を行った。提案モデルの動画生成により動画生成に成功した。また補間や新たに獲得した潜在変数により未知のデータ生成ができた。モデルによる2つのデータ学習の成功により，動画の生成モデルとしての有効性を確認することができた。

注：

[1]Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio: Generative Adversarial Nets

[2]Piotr Bojanowski, Armand Joulin, David Lopez-Paz, Arthur Szlam (Facebook AI Research): Optimizing the Latent Space of Generative Networks

[3]Alex Graves: Supervised Sequence Labelling with Recurrent Neural Networks