Latent Motion Manifold with Sequential Auto-Encoders

Deok-Kyeong Jang and Sung-Hee Lee

Korea Advanced Institute of Science and Technology (KAIST)

Abstract

We propose the sequential autoencoders for constructing latent motion manifold. Sequential autoencoders minimize the difference between the ground truth motion space distribution and reconstructed motion space distribution sampled from the latent motion manifold. Our method is based on sequence-to-sequence model for encoding the temporal information of human motion. We also adopt Wasserstein regularizer for matching encoded training distribution to the prior distribution of motion manifold. Our experiments show that randomly sampled points from trained motion manifold distribution become natural and valid motions.

Introduction

Constructing human motion spaces has been regarded as a classical and important problem as it has a wide range of applications such as motion synthesis, prediction, and interpolation. However, constructing a concise motion space and extracting valid motions from the motion space is a challenging problem. [Law04] used Gaussian Process Latent Variable Model (GPLVM) to find a low dimensional latent space for high dimensional motion data. Recently, with the development of deep learning technology, a method of constructing a motion manifold by using Convolutional Neural Network (CNN)-based encoder was introduced by [HSKJ15].

In this paper, we introduce a novel sequential autoencoder optimized for motion data by using Recurrent Neural Nework (RNN) cell. In order to model temporal information of human motion, our method is based on sequence-to-sequence model that is commonly used in neural machine translation. In addition, we use the Wasserstein regularizer [TBGS17] which encourages the encoded training distribution to match the prior to obtain more natural motion reconstruction results. Based on this, we generate a meaningful time dependent latent motion manifold.

References

1. [HSK15] Learning motion manifolds with convolutional autoencoders.
2. [Law04] Gaussian process latent variable models for visualisation of high dimensional data.
3. [IPOS14] Human3. 6m: Large scale datasets and predictive methods for 3d human sensing in natural environments
4. [MBR17] On human motion prediction using recurrent neural networks

Method

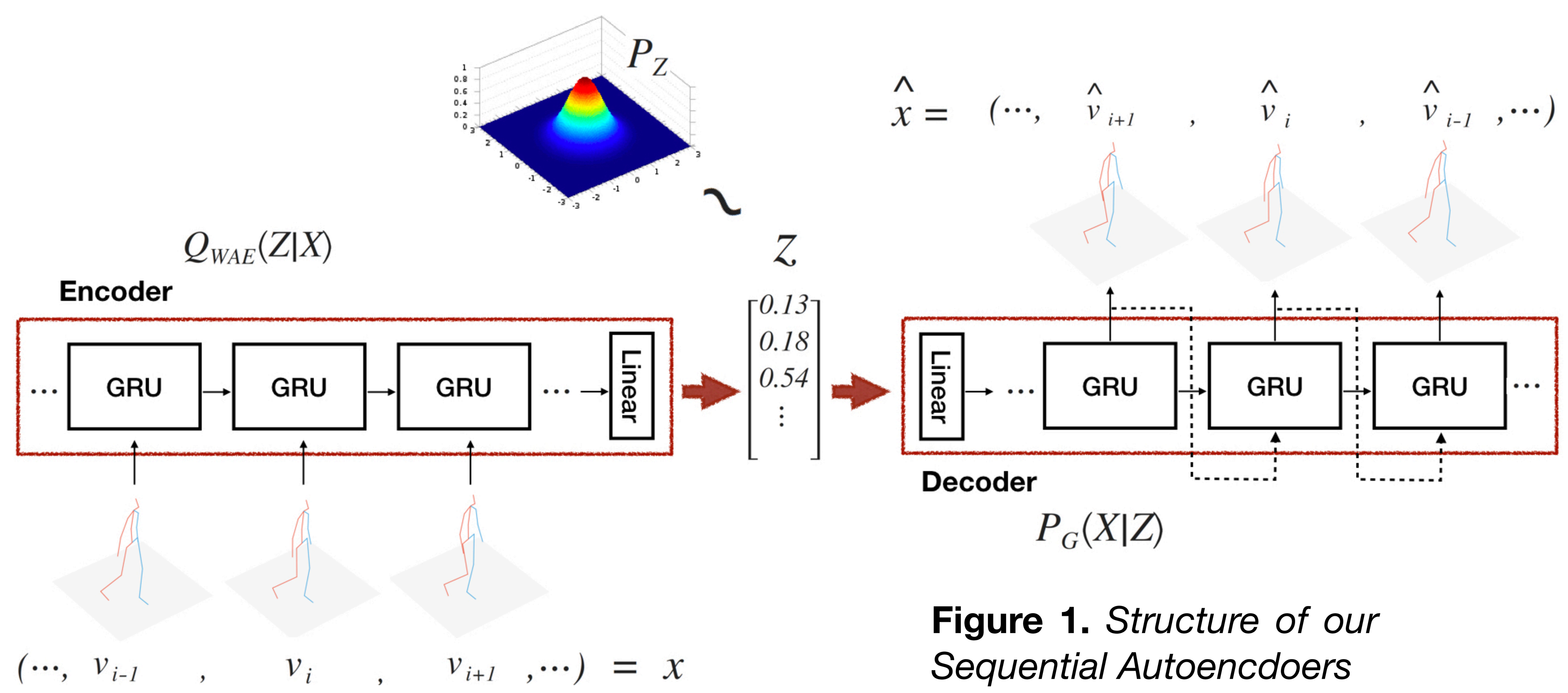

Figure 1. *Structure of our Sequential Autoencdoers*

We learn the motion data in an unsupervised way using a network of sequential autoencoders. To train and test our network, we used Human 3.6M (H3.6M) dataset by [IPOS14]. H3.6M data consist of 15 activities for 7 subjects respectively such as walking, smoking, discussion, taking pictures, and phoning. Euclidean distance was used as a metric in the joint angle space.

Sequential Auto-Encoders with Wasserstein Regularizer

Here we describe our sequential autoencoders. Our purpose is to minimize the difference between the ground truth motion space distribution and reconstructed motion space distribution sampled from the latent motion manifold. The overall structure is shown in Fig. 1. Sequence-to-sequence model first reads the source motion sequence using an encoder to build a motion manifold. A decoder, then, processes the motion manifold to emit a reconstructed motion sequence. For this, we use Gated Recurrent Unit (GRU), which is widely used for handling sequential data. We adopt Wasserstein regularizer for matching encoded training distribution to the prior distribution of motion manifold. This regularizer forces the continuous mixture motion manifold distribution to match prior distribution, and thereby promotes better reconstruction. Refer to [TBGS17] for details about Wasserstein regularizer. One frame of motion information (joint angles) was used as the input for each GRU cell. In a decoder case, we used sampling-based method [MBR17]: decoder's guess is used as next input.
The target motion sequence is set in a reversed order from the source motion sequence.

Experimental Setup

Among H3.6M actions, we performed manifold learning for walking, waiting and walking dog actions, respectively. Motion clips of 50 frames were randomly selected from the input motion sequence. Motion is represented in joint space, whose dimension is 54 made by 17 joints.

To verify whether the latent motion manifold is formed well, we randomly sampled points to reconstruct the actions from trained multinomial motion manifold distribution, and checked whether a meaningful motion was generated after passing through the decoder or not.

Results

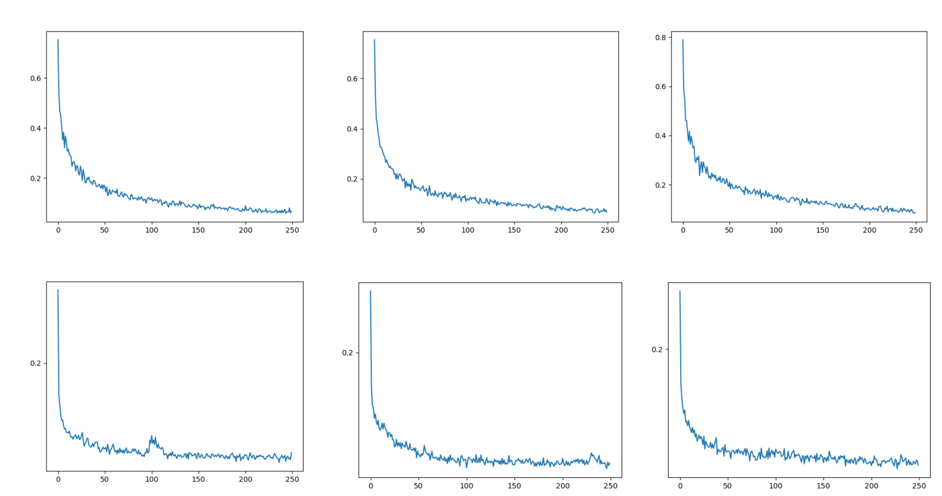

Figure 2. *Error curves for each motion about reconstructions (first row) and discrepancy between prior distribution of motion manifold and distribution induced by the encoder (second row). The leftmost column is about walking action, middle one is waiting action and last column is walking dog action.*

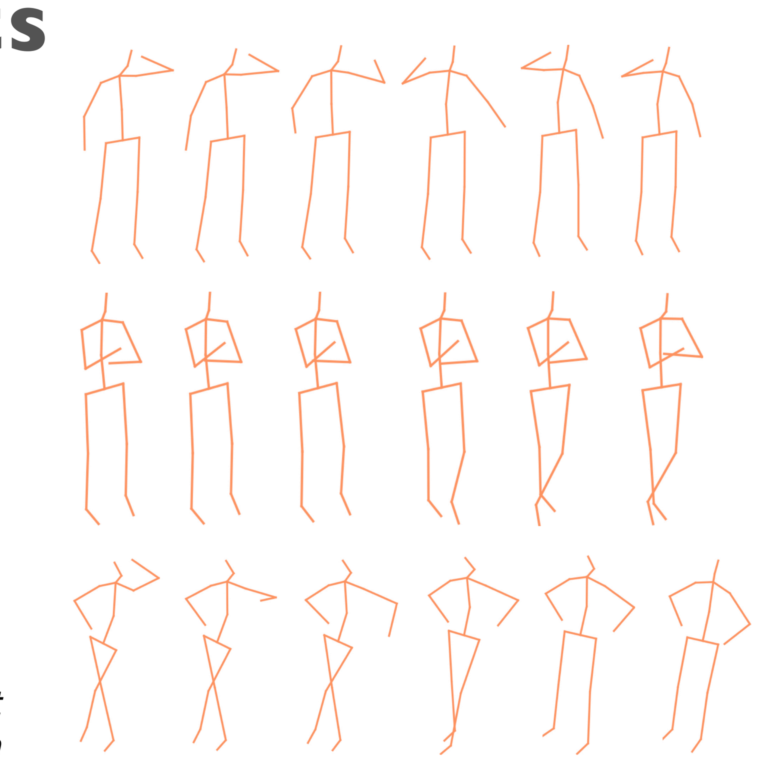

Figure 3. *Sampling motions from waiting motion manifold. First row motion seems like stretching, Second one is twisting legs and Last one is resting on the wall.*

The training results on walking, waiting, and walking dog actions are shown in Fig. 2. It can be seen that as the iteration increase, reconstruction error and discrepancy error between prior distribution of motion manifold and distribution induced by the encoder gradually decrease. We extracted 16 random samples per action and checked the reconstruction results. Because walking actions are a periodic motion, all of the 16 sampled results show similar walking pattern. Waiting and walking dog actions are non-periodic motions including various behaviors. Therefore, all 16 reconstructed motions showed different behaviors. Figure 3 is the result of sampled reconstruction for the waiting motion. One can see that stretching, crossing the legs, and resting on the wall have been reconstructed. This result suggests that our motion manifold and decoder can create various plausible behaviors.

shadofex@kaist.ac.kr