Learning Time-series Interaction Data by Association

Geun-Tae Park, Yoonsang Lee
Department of Computer Science
Hanyang University
Seoul, Korea
{qkrrmsxo01, yoonsanglee}@hanyang.ac.kr

***Abstract*—In this paper, we propose a novel learning method adopted associative learning that using association to train interaction between characters. Before using 3D interaction dataset, we use a simple 2D point motion dataset to verify the proposed method is available to 3D motion data.**

***Keywords*—data-driven animation, neural networks**

I. Introduction

Generating natural interactions between characters is one of the major challenges in computer graphics. Since a motion that response to changes of surrounding characters and environment increases reality and immersion about the character, it is especially important in areas where user experience is important, such as games. However, compared to their importance, it is not easy to obtain motion datasets about interactions between characters under general conditions and the number of such datasets available is much less than that of single motion. Because many of the previously proposed data-driven based motion generation methods require a large number of training datasets, it's hard to use these methods to generate motions with interactions using relatively small number of interacting motions.

We propose a novel training method based on associative learning [1] which has been previously proposed as semi-supervised learning in the image classification. The network uses a small number of interaction motion dataset that contains any motion A and the corresponding counterpart's motion B as a supervised dataset and a large number of single motion dataset that has only motion A' without its counterpart's motion as an unsupervised dataset for training. By training association between them, the network can infer the correlation between input motion and output motion, which is lacking in the supervised dataset.

In this paper, before using 3D motion data, we use simple time series data of 2D points that represent the motion on 2D plane. By using these data, we aim to verify the feasibility of our method and its applicability to the 3D motion data.

II. Learning by Association

A. Data For Training

Two corresponding points are created to represent interaction in 2D plane. Both points have same starting position. To move from the starting to ending position, one follows a circle trajectory and the other follows a straight line.

The center position and radius of the circle are randomly sampled from Gaussian distribution. The starting point is randomly selected as one of 20 equally divided points on the circle.

In our work, we use 100000 pairs of 2D movement. The mean and deviation for the center position are set to 0 and 1. The mean and deviation for the radius are set to 4 and 1. In association learning step, 10% of the movement pairs are used as supervised dataset pair (A, B) and the rest as unsupervised dataset A' without using corresponding movement B. A and A' can be circular or linear movement, which is randomly selected. For B, the movement corresponding to the movement of A is sampled. (i.e. When A is a circular movement, B is the corresponding linear movement, and vice versa.)

B. Network

The network model used in our work consists of two auto-encoder $\mathrm{AE_A}$ and $\mathrm{AE_B}$. Both auto-encoder $\mathrm{AE_A}$ and $\mathrm{AE_B}$ are simple network which consist of encoder and decoder composed of 3 fully connected layers. $\mathrm{AE_A}$ is an auto-encoder that generates motion B corresponding with input motion A. Encoder $\mathrm{E_A}$ of $\mathrm{AE_A}$ generates emb^A and $emb^{A'}$, feature of A and A'. Decoder $\mathrm{D_A}$ aims to reconstruct B from emb^A and ultimately to predict B' from $emb^{A'}$ about any A' which is not included in A. $\mathrm{AE_B}$ is an auto-encoder which restores motion B from input motion B. $\mathrm{AE_B}$'s encoder $\mathrm{E_B}$ makes a latent vector emb^B, feature of B and decoder $\mathrm{D_B}$ restores B from emb^B.

C. Loss Function

The network is trained by three types of loss term.

Restoration Loss. As a loss to learn $\mathrm{AE_B}$, its goal is to restore B from input B. Through this loss, $\mathrm{E_B}$ learns a latent space that can well represent the features of B. The restoration loss is calculated by MSE between input and output of $\mathrm{AE_B}$.

Reconstruction Loss. As a loss to learn $\mathrm{AE_A}$, its goal is to reconstruct B from input A. To calculate this loss, ground truth B is required, so it is calculated only for supervised dataset. By this loss, $\mathrm{AE_A}$ learns the way to generate output B corresponding to input A. The reconstruction loss is calculated by MSE between input and output of $\mathrm{AE_A}$.

Association Loss. Haeusser et al. [2017] [1] propose a method that learns association by aligning a round-trip probability with a target distribution. The round-trip probability represents the arrival probability of each sample of A in which starts from a sample of A and backs to a sample of A via samples of A'. The target distribution is defined as uniform distribution between samples which have same class label for each sample.

Since original method targets to image classification which has class value as ground truth, it is hard to use for motion data. Thus, instead of using class matching like original method, we firstly define the similarity S between every emb^Bs in a batch as L_2 norm.

$$S_{ij} \coloneqq \left\|\mathrm{emb}_i^B - \mathrm{emb}_j^B\right\|^2 \tag{1}$$

And then transform S into target distribution T by

$$T_{ij} \coloneqq 1 - {}^{S_{ij}}\!\Big/_{\sum_k S_{ik}} \tag{2}$$

Where *i* and *j* represent index of samples in the batch.

As Haeusser et al. [2017] [1] proposed, we define the similarity M between emb^A and $emb^{A'}$ as dot product. And then we transform M into transition probabilities $\mathrm{P}^{AA'}$ from A to A' by softmaxing M.

$$\mathrm{M}_{ij} \coloneqq emb_i^A \cdot emb_j^{A'} \tag{3}$$

$$\mathrm{P}_{ij}^{AA'} \coloneqq {}^{\exp(\mathrm{M}_{ij})}\!\Big/_{\sum_k \exp(\mathrm{M}_{ik})} \tag{4}$$

The association loss is represented as the sum of walker loss and visit loss. The walker loss aligns round trip probabilities $\mathrm{P}^{AA'A}$ with T by using cross-entropy. The visit loss makes distribution of $\mathrm{P}^{AA'}$ even so that model can visit samples of A' in the batch as many as possible by using uniform distribution U.

$$\mathrm{P}_{ij}^{AA'A} \coloneqq \left(\mathrm{P}^{AA'}\mathrm{P}^{A'A}\right)_{ij} = \sum_k \mathrm{P}_{ik}^{AA'}\mathrm{P}_{kj}^{A'A} \tag{5}$$

$$L_{walker} \coloneqq cross\_entropy(T, \mathrm{P}^{AA'A}) \tag{6}$$

$$L_{visit} \coloneqq cross\_entropy(\mathrm{U}, \mathrm{P}^{AA'}) \tag{7}$$

$$L_{assoc} = L_{walker} + L_{visit} \tag{8}$$

In (5), $\mathrm{P}^{A'A}$ is transposed form of $\mathrm{P}^{AA'}$.

III. Experiment and Results

We set batch size to 200 and train the network while 2000 epochs. In every 100 epochs, we report the parameter of model to check learning processes of the network for each case.

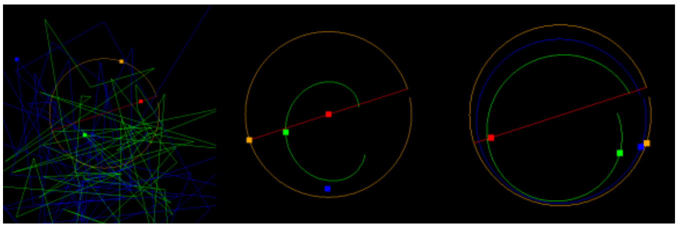

Fig. 1. Sequence of associative learning result. Red denotes input data, orange denotes ground truth, green denotes $\mathrm{AE_A}$'s output and blue denotes $\mathrm{AE_B}$'s output.

Associative Learning Result. We visualize the predicted counterpart's motion (B) from $\mathrm{AE_A}$ (green) and $\mathrm{AE_B}$ (blue). We find these two points are gradually getting closer to ground truth (orange) as learning progresses (Fig. 1).

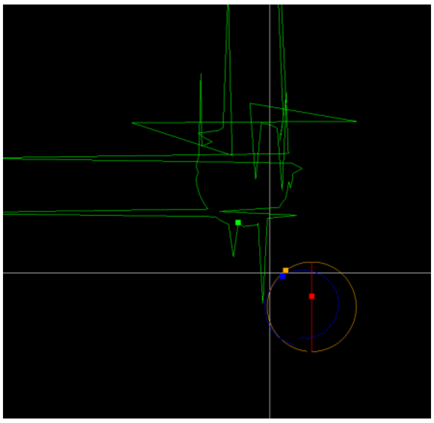

Fig. 2. Result of associative learning without reconstruction loss. Green shows circle like output.

Ablation. To verify whether association loss can really learn the feature of B, we train the network without reconstruction loss from the associative learning. The result shows that full information of feature of B cannot be obtained by the association loss alone. However, we find that association loss can extract some feature of B by the result that partially shows similar shape and size properties to the ground truth.

IV. Conclusion

By these experiments, we can see the possibility that our method can be applied to 3D motion dataset. However, when it takes to 3D space motion, new challenges may arise as dimension increase. Thus, as our next work, we will use 3D interaction and single motion dataset to associative learning.

Acknowledgment

This work was supported by the NationalResearch Foundation of Korea(NRF) grant funded by the Korea gov-ernment(MSIT) (NRF-2019R1C1C1006778, NRF-2019R1A4A1029800).

References

[1] Haeusser, Philip, Alexander Mordvintsev, and Daniel Cremers. "Learning by Association--A Versatile Semi-Supervised Training Method for Neural Networks." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017.