Table Tennis Project Report

Tao Chen
chent@slxrobot.com

Shanghai LingXian Robotics

Apr 07, 2017

Outline

Background

Prediction Neural Network

Training Result

Outline

Background

Prediction Neural Network

Training Result

Mission Statement

- Mission: Given several initial dual-camera frames, predict the table tennis ball's position in future frames
- In experiments:
  - Camera sampling frequency: 30 Hz
  - Algorithm's input: 14 initial frames
  - Algorithm's output: ball's positions in 33th-38th frames

- Supervised learning → model a conditional distribution $p(\boldsymbol{t}|\boldsymbol{x})$
- Unimodal distribution:
  - $p(\boldsymbol{t}|\boldsymbol{x})$ is often chosen to be Gaussian
- Multimodal Distribution:
  - $p(\boldsymbol{t}|\boldsymbol{x})$ can be mixture density network (MDN)

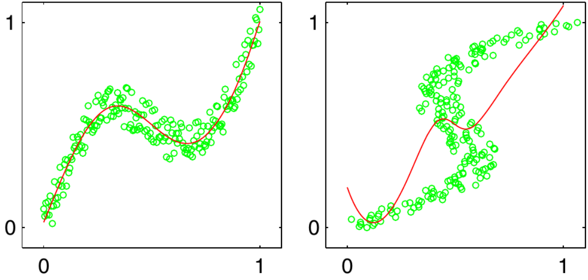


Figure : Unimodal and Multimodal

Source : *Pattern Recognition and Machine Learning, Bishop, 2006*

Basics

MDN

- MDN Formulation:

$$p(\boldsymbol{t}|\boldsymbol{x}) = \sum_{k=1}^{K} \pi_k(\boldsymbol{x}) \mathcal{N}(\boldsymbol{t}|\mu_k(\boldsymbol{x}), \sigma_k^2(\boldsymbol{x}))$$

s.t.

$$\sum_{k=1}^{K} \pi_k(\boldsymbol{x}) = 1, \quad 0 \leq \pi_k(\boldsymbol{x}) \leq 1$$

$$\sigma_k^2(\boldsymbol{x}) \geq 0$$

To satisfy the constraints:

$$\pi_k(\boldsymbol{x}) = \frac{e^{a_k^\pi}}{\sum_{\ell=1}^{K} e^{a_\ell^\pi}}, \quad \sigma_k(\boldsymbol{x}) = e^{a_k^\sigma}$$

Basics
MDN

- MDN Loss: Maximum Likelihood

$$E(\boldsymbol{w}) = -\sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k(\boldsymbol{x}_n, \boldsymbol{w}) \mathcal{N}(\boldsymbol{t}_n | \mu_k(\boldsymbol{x}_n, \boldsymbol{w}), \sigma_k^2(\boldsymbol{x}_n, \boldsymbol{w})) \right\}$$

Basics

Highway Networks

- Training deeper networks is not as straightforward as simply adding layers
- Highway Networks enables the optimization of networks with virtually arbitrary depth
- Key: **gating mechanism** (inspired by LSTM)

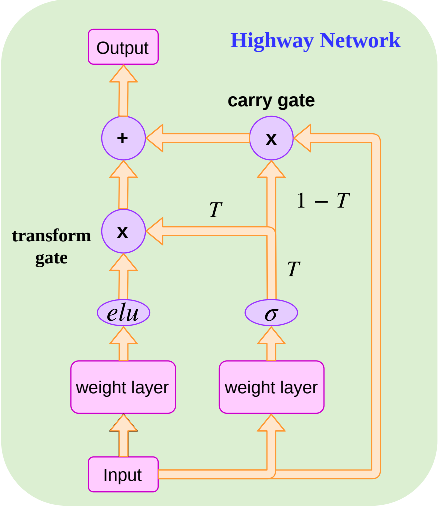
$$\boldsymbol{y} = H(\boldsymbol{x}, \boldsymbol{W_H}) \cdot T(\boldsymbol{x}, \boldsymbol{W_T}) + \boldsymbol{x} \cdot (1 - T(\boldsymbol{x}, \boldsymbol{W_T}))$$

where H can be an affine transform followed by a non-linear activation function and:

$$T(\boldsymbol{x}) = \sigma(\boldsymbol{W_T^T}\boldsymbol{x} + \boldsymbol{b_T})$$

Basics

Highway Networks



Basics

LSTM

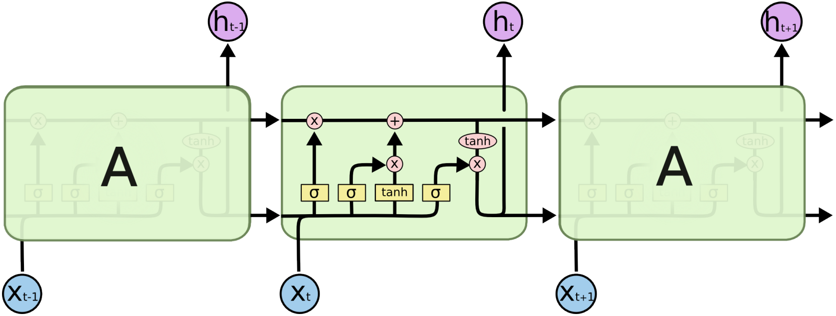


Figure : LSTM

Source : http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Outline

Background

Prediction Neural Network

Training Result

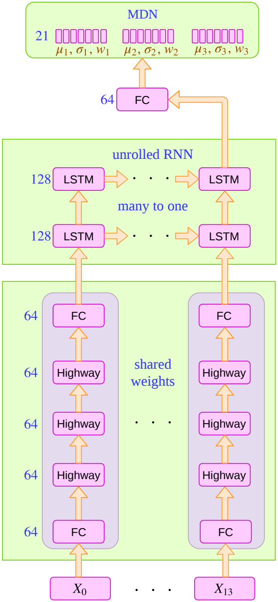
Function

- **Two** similar neural networks were designed and trained
- One is to predict the ball's position in a single future frame
- The other one is to predict the ball's positions in multiple future frames simultaneously

Prediction Neural Network

Single Frame Prediction

- Single future frame prediction
- Input: 14 initial frame data
- Output: ball's position distribution in 38th frame



Training Process

Single Frame Prediction

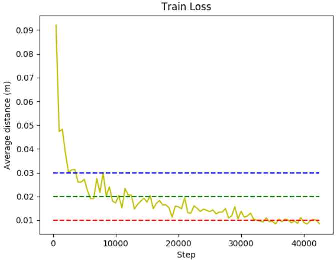


Figure : Training Loss

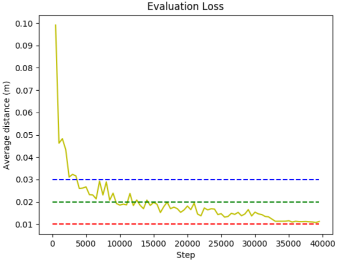
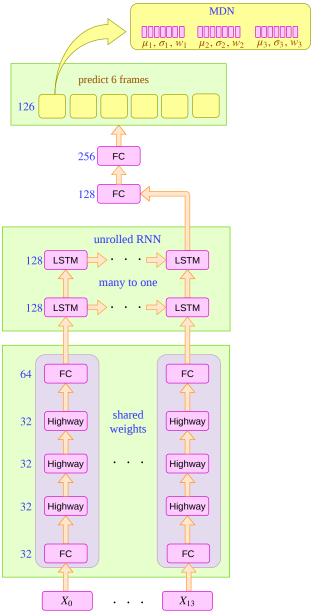


Figure : Evaluation Loss

Prediction Neural Network

Multiple Frame Predictions

- Multiple future frame predictions
- Input: 14 initial frame data
- Output: ball's position distributions in 33th-38th frames



Training Process

Multiple Frame Predictions

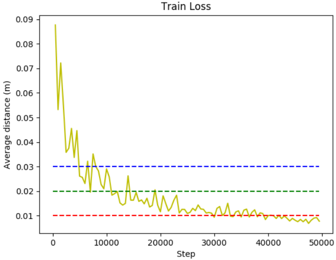


Figure : Training Loss

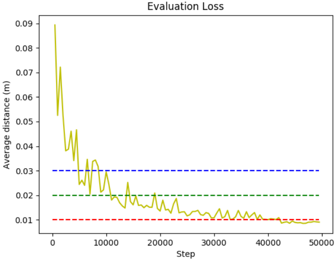


Figure : Evaluation Loss

Outline

Background

Prediction Neural Network

Training Result

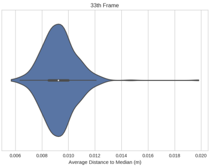
How to Measure

- To get an appropriate threshold value of the distance between true position and predicted position, the training data distribution should be considered

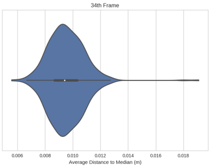
Training Data Distribution

- As the Gazebo model is not perfect yet, the table tennis ball cannot repeat its trajectory with high accuracy even under same force condition
- When collecting the training and testing data, each force condition is applied to the ball 50 times (get 50 similar trajectories)
- The degree of repeatability of each 50 trajectories is measured by the average distance from each trajectory to the median trajectory at each time step
- **480** force conditions were applied and collected, resulting in **24000** trajectories

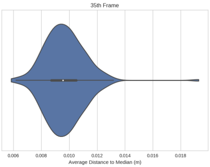
Training Data Distribution



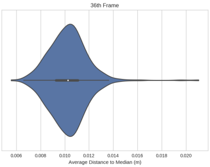
(a) 33th frame



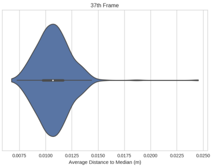
(b) 34th frame



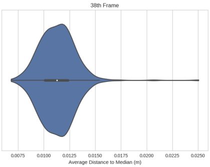
(c) 35th frame



(d) 36th frame



(e) 37th frame



(f) 38th frame

Figure : Violin-plots of Average Distance to Median Trajectory

Training Result

- Most trajectories from the training data are close to each other by the upper bound: 1.5 cm $\times 2 = 3$ cm
- Single Frame Prediction

Data Source	1 cm error	2 cm error	3 cm error
Training	58.94 %	85.05 %	93.11 %
Testing	57.22 %	82.62 %	91.17 %

Training Result

- Multiple Frame Prediction
  Training Data:

Data Source	1 cm error	2 cm error	3 cm error
33th frame	69.15 %	89.86 %	97.04 %
34th frame	65.62 %	89.11 %	96.95 %
35th frame	65.05 %	88.09 %	96.01 %
36th frame	63.00 %	86.77 %	95.24 %
37th frame	62.07 %	86.23 %	94.73 %
38th frame	56.97 %	84.46 %	94.33 %

Training Result

- Multiple Frame Prediction
  Testing Data:

Data Source	1 cm error	2 cm error	3 cm error
33th frame	68.72 %	89.21 %	96.33 %
34th frame	65.21 %	87.85 %	96.33 %
35th frame	64.35 %	87.36 %	95.64 %
36th frame	61.85 %	86.14 %	94.64 %
37th frame	60.36 %	85.56 %	94.32 %
38th frame	56.74 %	83.32 %	93.57 %

Training Result

Single Frame Prediction

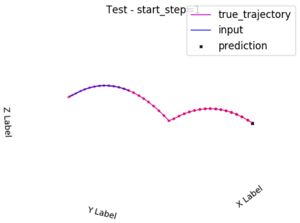


Figure : Offline Testing Data Test

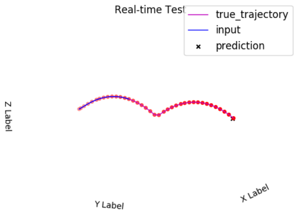


Figure : Gazebo Real-time Test

Training Result

Multiple Frame Predictions

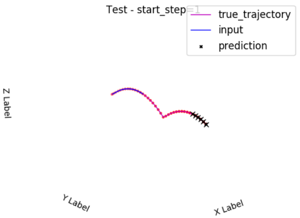


Figure : Offline Testing Data Test

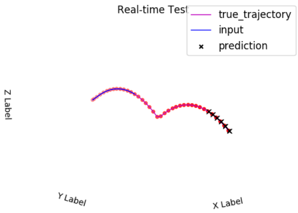


Figure : Gazebo Real-time Test