Dynamic Gesture Recognition with Hidden Markov Model

Qing Chen
Discover Lab
University of Ottawa

January 18, 2005

Outline

- 1.Introduction
- 2.Dynamic Data
- 3.HMM in a Nutshell
- 4.Data Processing
- 5.Model Training
- 6.Recognition
- 7.Conclusions & Future Work

1.Introduction

- Objective: Classify dynamic gesture with Discrete HMM (DHMM);
- Motivation: VERGINA Project
- Tasks involved:
  - data analysis & process
  - model training
  - classification

2.Dynamic Data

- A group of raw data which varies with time;
- For CyberGlove, data collected from 20 joint angles at about 10Hz;
- Example: (1) Raw data for the ‘Grabbing’ gesture:

Finger joints →

Time ↓

5.43E-01	3.50E-01	7.06E-01	6.42E-01	7.43E-01	6.34E-01	4.82E-01	7.64E-01	6.80E-01	3.40E-01	5.74E-01	7.59E-01	5.08E-01	5.44E-01	8.46E-01	6.42E-01	4.82E-01
5.36E-01	3.56E-01	7.13E-01	6.32E-01	7.49E-01	6.40E-01	4.40E-01	7.59E-01	6.65E-01	3.41E-01	5.61E-01	7.47E-01	5.44E-01	5.24E-01	8.40E-01	6.32E-01	4.40E-01
5.00E-01	3.50E-01	5.37E-01	5.63E-01	6.07E-01	5.43E-01	3.70E-01	6.83E-01	5.63E-01	3.40E-01	5.10E-01	6.47E-01	6.07E-01	4.63E-01	7.49E-01	5.63E-01	3.70E-01
5.00E-01	3.68E-01	3.01E-01	4.15E-01	5.48E-01	5.33E-01	3.31E-01	5.76E-01	4.71E-01	3.41E-01	4.64E-01	5.16E-01	6.41E-01	4.43E-01	6.13E-01	4.15E-01	3.31E-01
4.79E-01	3.68E-01	2.67E-01	3.73E-01	5.34E-01	5.32E-01	3.25E-01	5.26E-01	4.29E-01	3.39E-01	4.39E-01	4.60E-01	6.47E-01	4.30E-01	5.67E-01	3.73E-01	3.25E-01
4.57E-01	3.68E-01	2.36E-01	3.28E-01	5.34E-01	5.32E-01	3.62E-01	4.74E-01	4.13E-01	3.34E-01	4.13E-01	4.13E-01	6.19E-01	4.23E-01	5.32E-01	3.28E-01	3.62E-01
4.57E-01	3.68E-01	2.13E-01	2.86E-01	4.96E-01	5.35E-01	4.26E-01	4.18E-01	3.73E-01	3.06E-01	3.82E-01	3.75E-01	5.81E-01	3.84E-01	4.97E-01	2.86E-01	4.26E-01
4.50E-01	3.62E-01	1.97E-01	2.31E-01	4.52E-01	5.45E-01	5.20E-01	3.46E-01	3.27E-01	2.71E-01	3.28E-01	3.25E-01	5.26E-01	3.35E-01	4.33E-01	2.31E-01	5.20E-01
4.86E-01	3.38E-01	1.72E-01	1.53E-01	3.86E-01	5.61E-01	6.72E-01	2.09E-01	2.78E-01	3.02E-01	1.87E-01	2.78E-01	3.67E-01	1.60E-01	4.04E-01	1.53E-01	6.72E-01
5.36E-01	3.21E-01	1.42E-01	1.27E-01	2.93E-01	4.81E-01	7.56E-01	1.43E-01	2.52E-01	3.60E-01	1.23E-01	2.41E-01	2.55E-01	1.16E-01	2.83E-01	1.27E-01	7.56E-01
5.50E-01	3.04E-01	1.34E-01	1.17E-01	2.46E-01	4.16E-01	7.90E-01	1.31E-01	2.39E-01	3.76E-01	1.12E-01	2.13E-01	2.17E-01	1.13E-01	2.34E-01	1.17E-01	7.90E-01

2.Dynamic Data (cont'd)

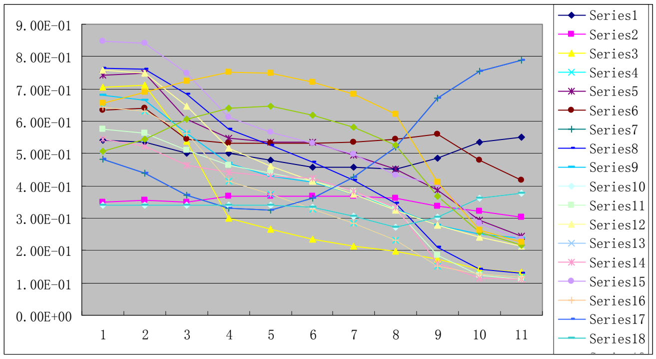


Fig.1. Dynamic data for the 'Grabbing' gesture.

2.Dynamic Data (cont'd)

- Example: (2) Raw data for the 'Quote' gesture:

Finger joints

Time

2.03E-01	2.23E-01	2.42E-01	4.97E-01	8.44E-01	7.87E-01	7.07E-01	6.66E-01	7.84E-01	2.96E-01	4.13E-01	1.44E-01	3.39E-01	3.35E-01	1.54E-01	4.97E-01	7.07E-01
1.72E-01	2.23E-01	2.08E-01	5.36E-01	2.99E-01	4.88E-01	7.57E-01	6.72E-01	2.33E-01	2.91E-01	4.01E-01	1.33E-01	2.93E-01	3.17E-01	1.39E-01	5.36E-01	7.57E-01
1.60E-01	2.23E-01	1.92E-01	5.47E-01	1.20E-01	1.48E-01	8.16E-01	6.43E-01	1.16E-01	3.27E-01	3.94E-01	1.25E-01	2.16E-01	2.94E-01	1.31E-01	5.47E-01	8.16E-01
1.57E-01	2.18E-01	1.87E-01	5.30E-01	1.03E-01	1.06E-01	8.27E-01	5.95E-01	1.03E-01	3.45E-01	3.88E-01	1.22E-01	1.96E-01	2.77E-01	1.26E-01	5.30E-01	8.27E-01
1.60E-01	2.37E-01	2.02E-01	6.01E-01	2.24E-01	3.80E-01	7.66E-01	7.15E-01	1.94E-01	2.63E-01	4.13E-01	1.22E-01	3.03E-01	3.29E-01	1.31E-01	6.01E-01	7.66E-01
1.85E-01	2.32E-01	2.19E-01	5.47E-01	7.70E-01	6.67E-01	6.85E-01	6.94E-01	7.09E-01	2.66E-01	4.13E-01	1.33E-01	3.81E-01	3.41E-01	1.34E-01	5.47E-01	6.85E-01
1.98E-01	2.23E-01	2.36E-01	4.92E-01	8.66E-01	8.31E-01	7.09E-01	6.37E-01	8.31E-01	2.88E-01	4.07E-01	1.40E-01	3.42E-01	3.41E-01	1.42E-01	4.92E-01	7.09E-01
1.98E-01	2.23E-01	2.36E-01	4.75E-01	8.52E-01	8.02E-01	7.21E-01	6.31E-01	8.11E-01	3.09E-01	4.07E-01	1.40E-01	3.17E-01	3.41E-01	1.42E-01	4.75E-01	7.21E-01
1.68E-01	2.23E-01	1.87E-01	5.85E-01	2.63E-01	4.42E-01	7.50E-01	6.66E-01	2.33E-01	2.44E-01	3.94E-01	1.22E-01	3.33E-01	3.11E-01	1.26E-01	5.85E-01	7.50E-01
1.68E-01	2.18E-01	1.68E-01	5.63E-01	1.00E-01	1.00E-01	8.06E-01	6.13E-01	1.06E-01	3.47E-01	3.88E-01	1.15E-01	2.16E-01	2.72E-01	1.21E-01	5.63E-01	8.06E-01
1.64E-01	2.14E-01	1.63E-01	5.52E-01	9.47E-02	8.84E-02	8.17E-01	5.88E-01	1.00E-01	3.60E-01	3.82E-01	1.12E-01	2.00E-01	2.67E-01	1.18E-01	5.52E-01	8.17E-01

2.Dynamic Data (cont'd)

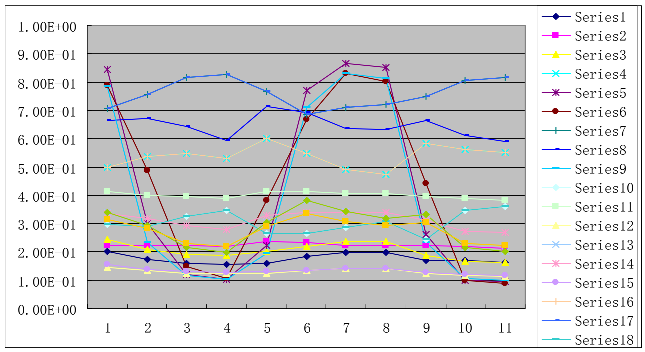


Fig.2. Dynamic data for the 'Quote' gesture.

2.Dynamic Data (cont'd)

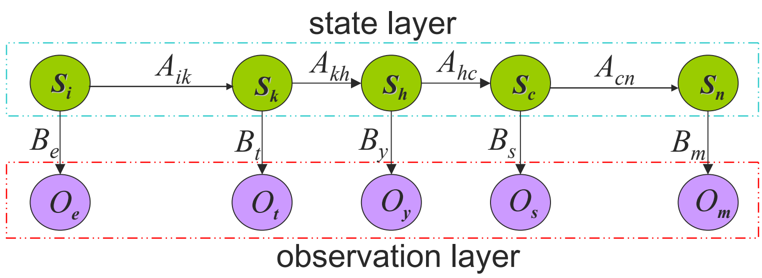
- We collected another group of data:

5.21E-01	3.68E-01	7.75E-01	6.42E-01	7.65E-01	6.60E-01	4.67E-01	7.55E-01	7.16E-01	3.47E-01	5.42E-01	7.76E-01	5.17E-01	5.44E-01	8.55E-01	6.42E-01	4.67E-01
5.21E-01	3.74E-01	7.75E-01	6.42E-01	7.65E-01	6.60E-01	4.54E-01	7.55E-01	7.16E-01	3.44E-01	5.42E-01	7.76E-01	5.31E-01	5.44E-01	8.55E-01	6.42E-01	4.54E-01
5.00E-01	3.44E-01	7.26E-01	5.90E-01	7.01E-01	5.94E-01	3.51E-01	7.15E-01	6.43E-01	3.65E-01	5.23E-01	7.22E-01	6.08E-01	4.97E-01	8.23E-01	5.90E-01	3.51E-01
4.86E-01	3.32E-01	6.63E-01	4.97E-01	6.55E-01	5.64E-01	2.94E-01	6.60E-01	5.63E-01	3.85E-01	5.10E-01	5.71E-01	6.48E-01	4.83E-01	7.53E-01	4.97E-01	2.94E-01
4.93E-01	3.27E-01	5.78E-01	3.78E-01	6.62E-01	5.68E-01	3.52E-01	5.82E-01	5.04E-01	3.85E-01	4.84E-01	5.16E-01	5.94E-01	4.50E-01	6.92E-01	3.78E-01	3.52E-01
4.79E-01	3.50E-01	5.29E-01	3.18E-01	6.42E-01	5.57E-01	3.77E-01	5.38E-01	4.46E-01	3.91E-01	4.64E-01	4.60E-01	5.67E-01	4.36E-01	6.29E-01	3.18E-01	3.77E-01
4.50E-01	3.81E-01	4.55E-01	2.64E-01	6.00E-01	5.41E-01	4.37E-01	4.74E-01	3.97E-01	3.80E-01	4.32E-01	3.98E-01	5.21E-01	3.97E-01	5.32E-01	2.64E-01	4.37E-01
4.29E-01	4.25E-01	3.67E-01	1.97E-01	5.56E-01	5.33E-01	5.39E-01	3.81E-01	3.35E-01	3.77E-01	3.52E-01	3.32E-01	4.34E-01	3.00E-01	4.85E-01	1.97E-01	5.39E-01
4.01E-01	4.44E-01	3.22E-01	1.68E-01	5.04E-01	5.34E-01	6.34E-01	2.80E-01	3.05E-01	3.78E-01	2.59E-01	3.04E-01	3.52E-01	1.83E-01	4.91E-01	1.68E-01	6.34E-01
3.68E-01	4.44E-01	2.94E-01	1.45E-01	4.15E-01	5.59E-01	7.09E-01	2.09E-01	2.84E-01	3.97E-01	1.87E-01	2.78E-01	2.76E-01	1.39E-01	4.16E-01	1.45E-01	7.09E-01
3.74E-01	4.57E-01	2.13E-01	1.31E-01	3.38E-01	5.26E-01	7.49E-01	1.67E-01	2.52E-01	4.19E-01	1.47E-01	2.41E-01	2.32E-01	1.21E-01	2.74E-01	1.31E-01	7.49E-01

- What kind of gesture is it?

3.HMM in a Nutshell

- A doubly embedded stochastic process;
- Two layers: state & observation



3.HMM in a Nutshell (cont'd)

- Only the observation layer is observable;
- The state sequence can only be exposed through the observation sequence;
- Three basic HMM problems:
  (1) The Evaluation Problem: Given the observation sequence $O=O_1O_2O\ldots O_T$, and a HMM $\lambda=(A,B,\pi)$, how to compute $P(O|\lambda)$?
  (2) The Decoding Problem: Given the observation sequence $O=O_1O_2\ldots O_T$, and the HMM λ, how to choose a corresponding optimal state sequence $Q=q_1q_2\ldots q_T$?
  (3) The Learning Problem: Given the observation sequence $O=O_1O_2\ldots O_T$, how to adjust the model parameters $\lambda=(A,B,\pi)$ to maximize $P(O|\lambda)$?
- The HMM we will use is 'Discrete' HMM.

4.Data Processing

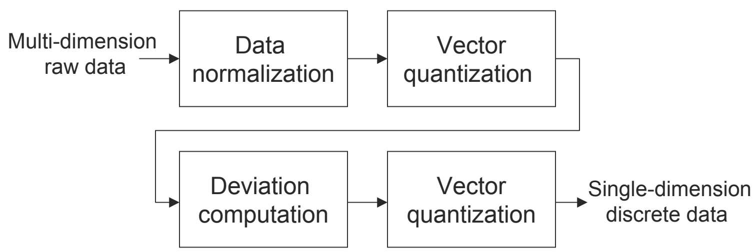
- Objective: transform multi-dimensional raw data to single-dimensional discrete data suitable for DHMM.
- Key problems:
  (1) What does state correspond to?
  (2) What does observation correspond to?
  (3) How to preprocess the raw data from the gesture input device into a sequence of discrete symbols?

4.Data Processing (cont'd)

- Proposed method:
  - states: finger joints
  - observation: joint angles
  - one-dimensional DHMM
- Advantages:
  - state layer is exposed so that the probability calculation is eliminated;
  - classification accuracy can be improved;
  - more straightforward than Multi-dimensional DHMM.

4.Data Processing (cont'd)

- Multi-dimension ⇒ Single-dimension observation sequence transformation:



4.Data Processing (cont'd)

- Example:

2.03E-01	2.23E-01	2.42E-01	4.97E-01	8.44E-01	7.87E-01	7.07E-01	6.66E-01	7.84E-01	2.96E-01	4.13E-01	1.44E-01	3.39E-01	3.35E-01	1.54E-01	4.97E-01	7.07E-01
1.72E-01	2.23E-01	2.08E-01	5.36E-01	2.99E-01	4.88E-01	7.57E-01	6.72E-01	2.33E-01	2.91E-01	4.01E-01	1.33E-01	2.93E-01	3.17E-01	1.39E-01	5.36E-01	7.57E-01
1.60E-01	2.23E-01	1.92E-01	5.47E-01	1.20E-01	1.48E-01	8.16E-01	6.43E-01	1.16E-01	3.27E-01	3.94E-01	1.25E-01	2.16E-01	2.94E-01	1.31E-01	5.47E-01	8.16E-01
1.57E-01	2.18E-01	1.87E-01	5.30E-01	1.03E-01	1.06E-01	8.27E-01	5.95E-01	1.03E-01	3.45E-01	3.88E-01	1.22E-01	1.96E-01	2.77E-01	1.26E-01	5.30E-01	8.27E-01
1.60E-01	2.37E-01	2.02E-01	6.01E-01	2.24E-01	3.80E-01	7.66E-01	7.15E-01	1.94E-01	2.63E-01	4.13E-01	1.22E-01	3.03E-01	3.29E-01	1.31E-01	6.01E-01	7.66E-01
1.85E-01	2.32E-01	2.19E-01	5.47E-01	7.70E-01	6.67E-01	6.85E-01	6.94E-01	7.09E-01	2.66E-01	4.13E-01	1.33E-01	3.81E-01	3.41E-01	1.34E-01	5.47E-01	6.85E-01
1.98E-01	2.23E-01	2.36E-01	4.92E-01	8.66E-01	8.31E-01	7.09E-01	6.37E-01	8.31E-01	2.88E-01	4.07E-01	1.40E-01	3.42E-01	3.41E-01	1.42E-01	4.92E-01	7.09E-01
1.98E-01	2.23E-01	2.36E-01	4.75E-01	8.52E-01	8.02E-01	7.21E-01	6.31E-01	8.11E-01	3.09E-01	4.07E-01	1.40E-01	3.17E-01	3.41E-01	1.42E-01	4.75E-01	7.21E-01
1.68E-01	2.23E-01	1.87E-01	5.85E-01	2.63E-01	4.42E-01	7.50E-01	6.66E-01	2.33E-01	2.44E-01	3.94E-01	1.22E-01	3.33E-01	3.11E-01	1.26E-01	5.85E-01	7.50E-01
1.68E-01	2.18E-01	1.68E-01	5.63E-01	1.00E-01	1.00E-01	8.06E-01	6.13E-01	1.06E-01	3.47E-01	3.88E-01	1.15E-01	2.16E-01	2.72E-01	1.21E-01	5.63E-01	8.06E-01
1.64E-01	2.14E-01	1.63E-01	5.52E-01	9.47E-02	8.84E-02	8.17E-01	5.88E-01	1.00E-01	3.60E-01	3.82E-01	1.12E-01	2.00E-01	2.67E-01	1.18E-01	5.52E-01	8.17E-01

Raw data processing

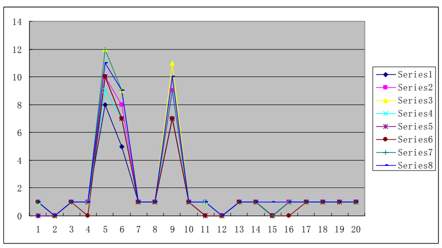
1	0	1	1	12	9	1	1	11	1	1	0	1	1	0	1	1	1	1	1

4.Data Processing (cont'd)

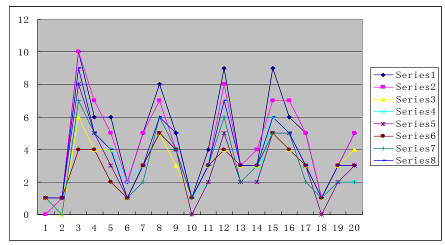
- 2 examples for 8 groups of raw data processing results:

```
List varInt array:
 0  0  1  1  8  5  1  1  7  1  1  0  1  1  0  1  1  1  1  1
 1  0  1  1 10  8  1  1  9  1  1  0  1  1  0  1  1  1  1  1
 1  0  1  1 12  9  1  1 11  1  1  0  1  1  0  1  1  1  1  1
 1  0  1  1  9  7  1  1  7  1  1  0  1  1  0  1  1  1  1  1
 0  0  1  1 10  7  1  1  7  1  0  0  1  1  0  1  1  1  1  1
 1  0  1  0 10  7  1  1  7  1  0  0  1  1  0  0  1  1  1  1
 1  0  1  1 12  9  1  1  9  1  1  0  1  1  0  1  1  1  1  1
 1  0  1  1 11  9  1  1 10  1  1  0  1  1  1  1  1  1  1  1
```

```
List varInt array:
 1  1 10  6  6  2  5  8  5  1  4  9  3  3  9  6  5  1  3  5
 0  1 10  7  5  2  5  7  4  1  3  8  3  4  7  7  5  1  3  5
 1  0  6  4  4  1  3  5  3  1  3  4  3  3  5  4  3  1  3  4
 1  1  8  5  4  2  3  5  4  1  2  5  2  2  6  5  3  1  2  2
 1  1  8  5  3  1  3  5  4  0  2  5  2  2  5  5  3  0  2  3
 1  1  4  4  2  1  3  5  4  1  3  4  3  3  5  4  3  1  3  3
 1  0  7  5  4  1  2  6  4  1  3  6  2  3  5  5  2  1  2  2
 1  1  9  5  4  1  3  6  5  1  3  7  3  3  6  5  3  1  3  3
```



'Quote' gesture



'Grab' gesture

5.Model Training

- The Learning Problem: Given the observation sequence $O=O_1O_2\ldots O_T$, how to adjust the model parameters $\lambda = (A,B, \pi)$ to maximize $P(O|\lambda)$?
- EM(Expectation-Modification), or equivalently Baum-Welch method, can locally maximize $P(O|\lambda)$;
- EM uses an iterative expectation/maximization procedure to find an HMM which is most likely to produce the training observation sequences locally;
- The accuracy of estimation improves with the number and length of the training sequences.

5.Model Training

- HMM training is a batch process;
- 8 groups of raw data for the same gesture are employed to train HMM;
- State transition matrix is decided by the data collection sequence:

joint1 → joint2 → joint3 → … … → joint19 → joint20

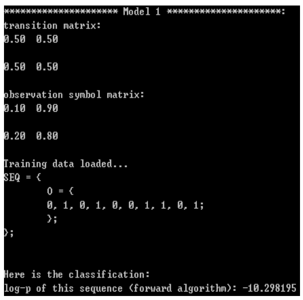
- The HMM for each dynamic gesture is trained offline.

6.Classification

- The Evaluation Problem: Given the observation sequence $O=O_1O_2O\ldots O_T$, and a HMM $\lambda=(A,B,\pi)$, how to compute $P(O|\lambda)$?
- Forward-backward procedure efficiently resolve the problem!
- Based on the trellis structure which can reduce the computation cost from $2TN^T$ to N^2T, where N is the number of states, and T is the number of the observations.

6.Classification (cont'd)

- An HMM Classification example with 2 states, 2 observation symbols :



- Model 3 produced the biggest log-p value, which verified the effectiveness of forward-backward procedure.

7.Conclusions & Future Work

- HMM can effectively resolve the classification problem for dynamic gesture data;
- The flexibility of HMM make it a generic mathematic model for recognition problems;
- Proposed data processing method effectively transformed multi-dimensional data to single-dimensional data while reserving the dynamic property;
- For dynamic gesture data, segmentation still need to be effectively resolved in order to meet the HMM requirements.