ESE 650 Homework 3 - Gesture Recognition with HMMs

Tarik Tosun

March 4, 2014

Introduction

In this assignment, I built a hidden-markov model (HMM) classifier for six different gestures using IMU data from a mobile phone. Data preprocessing proved very important. My final method processed the data by band-pass filtering, downsampling, and then quantizing using k-means clustering. I found that a very simple cyclic left-to-right HMM with 6 symbols and 2 hidden states produced the best results. Using these techniques, I was able to achieve a LOOCV error rate of about 10% on the training set.

Data Processing

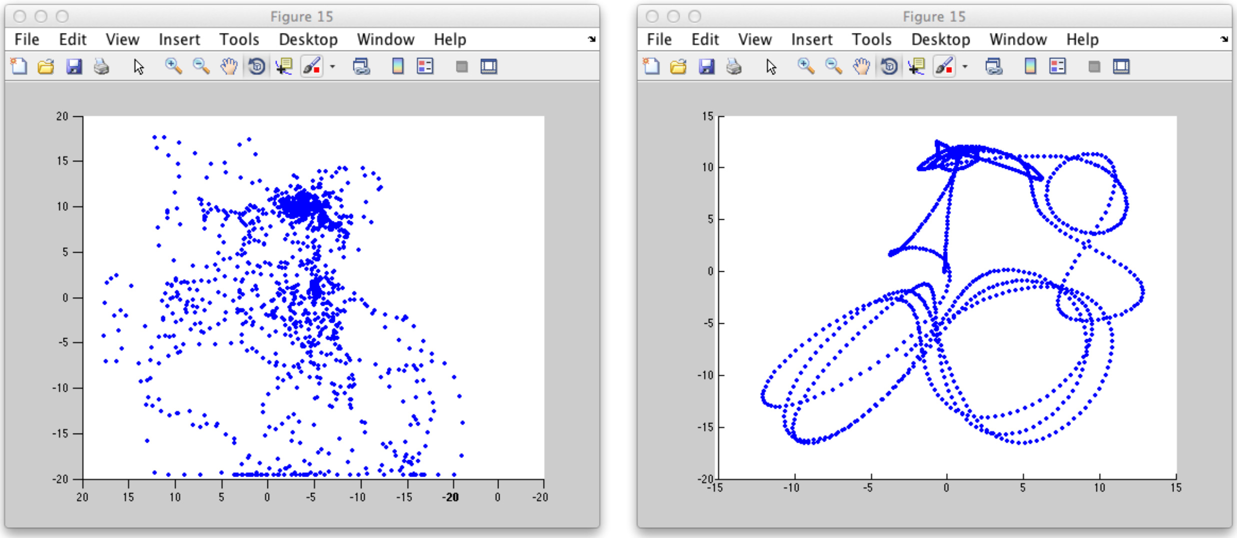
Data processing proved very important for achieving good results. My first step was to band-pass filter the data to select a range of frequencies appropriate for human gestures. I remove frequencies below 0.2 Hz and above 10Hz, thinking that 5 seconds was a reasonable upper bound on the period of a human gesture, and that it would be very difficult for a human to generate a hand-gesture at more than 10Hz. I think this filtering was important for extracting meaningful information from the IMU data. For example, Figure 1 shows the filtered and unfiltered data for a figure 8 gesture.

After filtering, the entire data set was quantized using k-means clustering. This converted the data from a 3-dimensional vector of continuous acceleration values to a 1-dimensional vector of k discrete symbols. HMM's can be run with either continuous or discrete observations, but based on my literature review it seemed that vector quantization was a popular choice for problems such as speech and gesture recognition [1, 2]. K-means and the Baum-Welch algorithm for HMM learning are both EM methods, so using them in tandem could be viewed as similar to learning a continuous-emission HMM where each emission distribution is a mixture of gaussians with shared diagonal covariances.

HMM Results

The Baum-Welch algorithm was used to train one HMM for each of the six gestures. Rather than using a left-to-right HMM, I used a cyclic model which includes a transition from the rightmost state to the leftmost state, forming a loop. My rationale was that this would make the model better-suited to modeling repeated motions and more robust to data sets of different sizes. I initialized my transition matrix with a 10% probability of transition to the next state, and my emission matrix with uniform probability over all symbols. Data was downsampled by a factor of 5 to increase HMM training speed. Using parameters $k = 6$ clusters/symbols and $n = 2$ hidden states, I was able to achieve a leave-one-out cross-validation (LOOCV) error rate of about 10%.

In my implementation, I used the matlab built-in functions `hmmtrain` and `hmmdecode` to perform Baum-Welch and Forward-Backward, respectively. I also wrote my own implementations of Baum-Welch and Forward-Backward, which can be found in `myHmmTrain.m` and `myHmmDecode.m`. I did not have enough time to thoroughly debug this code, but I believe they are fundamentally correct implementations of the algorithms. I have included the files in my submission for reference.



(a) Unfiltered

(b) Band-pass filtered (0.2-10Hz)

Figure 1: 3D plot of figure8 (#1) acceleration data. Horizontal axis is X, vertical is Z.

Analysis and Conclusions

For this problem, a simple model produced much better results than more complicated models. Through cross-validation, I found that increasing k increased the error rate slightly, and increasing n had an even larger negative effect. Decreasing k below 6 also increased the error rate. Other techniques I attempted included a more sophisticated quantization method (clustering each gesture individually to produce a different codebook for each gesture), and a larger state vector (including gyro data as well as accelerations). These both performed poorly.

From the results, it seems to me that the k-means clustering is really doing the heavy lifting in terms of differentiating gestures. Figure 2 shows clustering results for the entire data set (all five data sets for each motion are concatenated end-to-end in each picture). By inspecting the plots, we get the sense that the different gestures could probably be differentiated just by looking at the overall mix and proportion of clusters present, without even considering the ordering of the observations.

References

[1] Rabiner, L. (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proceedings of the IEEE.

[2] Mantyla, V., & Mantyjarvi, J. (2000). Hand gesture recognition of a mobile device user. In IEEE International Conference on Multimedia and Expo (Vol. 00, pp. 2–5). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=869596

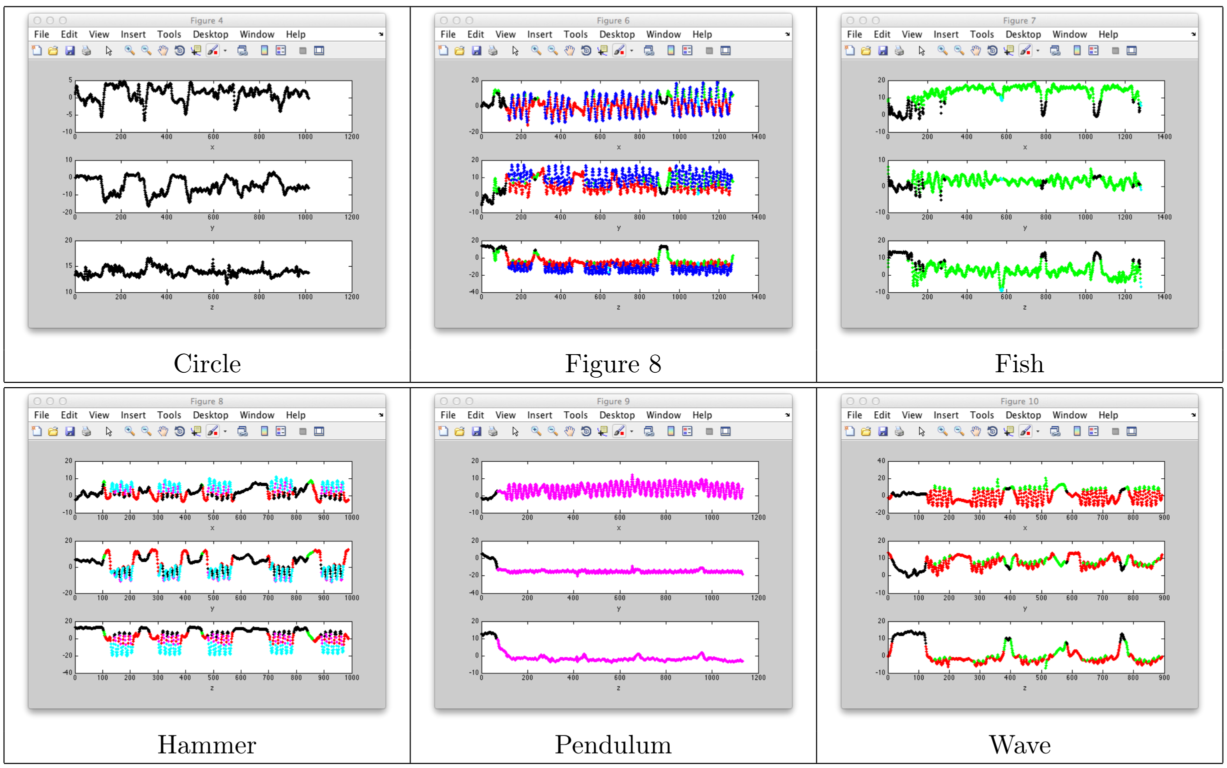


Figure 2: Clustered IMU data for the six gestures