Using Hidden Markov Models to analyse time series data

Ragnhild Noven

September 9, 2011
Background

- Want to analyse time series data coming from accelerometer measurements.
- 19 different datasets corresponding to different individuals.

Figure 1: Typical dataset
Background

- Want to analyse time series data coming from accelerometer measurements.
- 19 different datasets corresponding to different individuals
- Aim: classify each datapoint as belonging to some state

**Figure 1:** Typical dataset
Background

- Want to analyse time series data coming from accelerometer measurements.
- 19 different datasets corresponding to different individuals
- Aim: classify each datapoint as belonging to some state
- Compared models with 2-6 states

Figure 1: Typical dataset
Exploratory data analysis

Underlying process creating the jumps - use Hidden Markov Model (HMM).

PACF supports Markov assumption.
Exploratory data analysis

- Underlying process creating the jumps - use Hidden Markov Model (HMM)
Exploratory data analysis

- Underlying process creating the jumps - use Hidden Markov Model (HMM)
- PACF supports Markov assumption
Hidden Markov Models

- Assume data depends on a hidden Markov process
Hidden Markov Models

- Assume data depends on a hidden Markov process
- Each observation $O_t$ corresponds to a state $q_t$. 

![Diagram of Hidden Markov Models]

- Observations
- Hidden states

A → 1

B → 2

C → 3

$a_{12}$  

$a_{23}$  

$a_{13}$
Hidden Markov Models

- Assume data depends on a hidden Markov process
- Each observation $O_t$ corresponds to a state $q_t$.
- Markov process transitions between states according to transition matrix $A$. 

![Diagram of Hidden Markov Models]
Hidden Markov Models

- Assume data depends on a hidden Markov process
- Each observation $O_t$ corresponds to a state $q_t$.
- Markov process transitions between states according to transition matrix $A$.

Each state has a different probability distribution
Transforming the data

- Model assumes normal distribution in each state

Use Haar transform to decompose vector

\[ v = (v_1, v_2, \ldots, v_N) \]

into averages and differences:

Averages:

\[ a_1^m = v_{2m} - v_{2m-1} + \sqrt{2} \]

Differences:

\[ d_1^m = v_{2m} - v_{2m-1} - v_{2m} \sqrt{2} \]
Transforming the data

- Model assumes normal distribution in each state
- Use Haar transform to decompose vector $\mathbf{v} = (v_1, v_2, \ldots, v_N)$ into averages and differences:

\[
\begin{align*}
Averages: & \quad a_1^m = v_2^m - v_1^m + v_2^m \sqrt{2} \\
Differences: & \quad d_1^m = v_2^m - v_1^m - v_2^m \sqrt{2}
\end{align*}
\]
Transforming the data

- Model assumes normal distribution in each state
- Use Haar transform to decompose vector \( v = (v_1, v_2, \ldots, v_N) \) into averages and differences:
  - Averages:
    \[
    a^1_m = \frac{v_{2m-1} + v_{2m}}{\sqrt{2}}
    \]
  - Differences:
    \[
    d^1_m = \frac{v_{2m-1} - v_{2m}}{\sqrt{2}}
    \]
Fisz transform

- After decomposing into averages and differences, want to normalize data

\[ f_i^m = d_i^m \sqrt{a_i^m} \]
Fisz transform

- After decomposing into averages and differences, want to normalize data
- Use Fisz transform:
  \[ f_m^i = \frac{d_m^i}{\sqrt{a_m^i}} \]
Fisz transform

- After decomposing into averages and differences, want to normalize data
- Use Fisz transform:
  \[
  f_m^i = \frac{d_m^i}{\sqrt{a_m^i}}
  \]
- Finally perform inverse Haar transform
Fitting the HMM

- Model is described by parameters $\lambda = (A, \{b_i\}, \{\pi_i\})$
Fitting the HMM

- Model is described by parameters $\lambda = (A, \{b_i\}, \{\pi_i\})$
- Want to find the sequence of states that maximises the probability of the observations $O_1, O_2, ..., O_t$. 
Fitting the HMM

- Model is described by parameters $\lambda = (A, \{b_i\}, \{\pi_i\})$
- Want to find the sequence of states that maximises the probability of the observations $O_1, O_2, ..., O_t$.
- 2 problems in fitting the model:
  
  1. How to choose the optimal sequence of states given $O$ and $\lambda$.
  2. How to adjust the parameters $\lambda = (A, \{b_i\}, \{\pi_i\})$ to maximise $P(O|\lambda)$.
Problem 1

Want to find the optimal path of states corresponding to observations

\[ \delta_1(i) = \pi(i) b_i(O_1) \]

General case:

\[ \delta_{t+1}(j) = \max_i \delta_t(i) a_{ij} b_j(O_{t+1}) \]
Problem 1

- Want to find the optimal path of states corresponding to observations

\[ \delta_1(1) \quad \delta_2(1) \quad \delta_2(1)a_{12} \]

\[ \delta_1(2) \quad \delta_2(2) \quad \delta_2(2)a_{22} \]

\[ \delta_1(3) \quad \delta_2(3) \quad \delta_2(3)a_{32} \]

\[ 2 \]

- Initialise: \( \delta_1(i) = \pi(i)b_i(O_1) \).
Problem 1

- Want to find the optimal path of states corresponding to observations

- Initialise: \( \delta_1(i) = \pi(i)b_i(O_1) \).

- General case: \( \delta_{t+1}(j) = \left[ \max_i \delta_t(i)a_{ij} \right] b_j(O_{t+1}) \)
Problem 2

- Probabilities of observations depend on the path through the parameters $\lambda$. Use these probabilities to estimate parameters by an iterative method:
Problem 2

- Probabilities of observations depend on the path through the parameters $\lambda$. Use these probabilities to estimate parameters by an iterative method:

1. Set initial parameters $\lambda = (A, \{b_i\}, \{\pi_i\})$
Problem 2

- Probabilities of observations depend on the path through the parameters $\lambda$. Use these probabilities to estimate parameters by an iterative method:

1. Set initial parameters $\lambda = (A, \{b_i\}, \{\pi_i\})$

2. Calculate the probability of the observations and transitions given $\lambda$. 

Repeat steps 2 and 3 until convergence.
Problem 2

- Probabilities of observations depend on the path through the parameters $\lambda$. Use these probabilities to estimate parameters by an iterative method:

1. Set initial parameters $\lambda = (A, \{b_i\}, \{\pi_i\})$

2. Calculate the probability of the observations and transitions given $\lambda$.

3. Use these probabilities to reestimate the model parameters, e.g.

$$\bar{a}_{ij} = \frac{E(\# \text{ transitions from state } i \text{ to state } j)}{E(\# \text{ transitions from state } i)}$$
Problem 2

- Probabilities of observations depend on the path through the parameters $\lambda$. Use these probabilities to estimate parameters by an iterative method:

1. Set initial parameters $\lambda = (A, \{b_i\}, \{\pi_i\})$

2. Calculate the probability of the observations and transitions given $\lambda$.

3. Use these probabilities to reestimate the model parameters, e.g.

   $$\bar{a}_{ij} = \frac{E(\# \text{ transitions from state } i \text{ to state } j)}{E(\# \text{ transitions from state } i)}$$

4. Repeat steps 2 and 3 until convergence.
Results

Dataset 1

<table>
<thead>
<tr>
<th>State</th>
<th>Likelihood</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>7930</td>
<td>-15800</td>
</tr>
<tr>
<td>5</td>
<td>8360</td>
<td>-16600</td>
</tr>
<tr>
<td>4</td>
<td>6630</td>
<td>-13200</td>
</tr>
<tr>
<td>3</td>
<td>6990</td>
<td>-14000</td>
</tr>
<tr>
<td>2</td>
<td>6370</td>
<td>-12700</td>
</tr>
</tbody>
</table>
Results

<table>
<thead>
<tr>
<th>State</th>
<th>Likelihood</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>9180</td>
<td>-18300</td>
</tr>
<tr>
<td>5</td>
<td>9180</td>
<td>-18300</td>
</tr>
<tr>
<td>4</td>
<td>9170</td>
<td>-18300</td>
</tr>
<tr>
<td>3</td>
<td>9030</td>
<td>-18000</td>
</tr>
<tr>
<td>2</td>
<td>8970</td>
<td>-17900</td>
</tr>
</tbody>
</table>

Ragnhild Noven
Conclusions

The HMM describes the shape of the data fairly well, although better for some of the datasets than others. Using 5 states for the model gives the highest likelihood and lowest AIC for both datasets. The fitting procedure is very sensitive to the initial parameters. Some datasets display very rapid fluctuations in the fitted states which do not seem consistent with the data. Could require a minimum time spent in each state.
Conclusions

- The HMM describes the shape of the data fairly well, although better for some of the datasets than others.
Conclusions

- The HMM describes the shape of the data fairly well, although better for some of the datasets than others.

- Using 5 states for the model gives the highest likelihood and lowest AIC for both datasets.
Conclusions

- The HMM describes the shape of the data fairly well, although better for some of the datasets than others.

- Using 5 states for the model gives the highest likelihood and lowest AIC for both datasets.

- The fitting procedure is very sensitive to the initial parameters.
Conclusions

- The HMM describes the shape of the data fairly well, although better for some of the datasets than others.

- Using 5 states for the model gives the highest likelihood and lowest AIC for both datasets.

- The fitting procedure is very sensitive to the initial parameters.

- Some datasets display very rapid fluctuations in the fitted states which do not seem consistent with the data. Could require a minimum time spent in each state.
Thank you for listening!