Whale Song Unit Classification

Some Preliminary Exploration
Using Linear Prediction Vector Quantization
and Hidden Markov Modeling

Carlos Rueda

September 30, 2018

The Problem

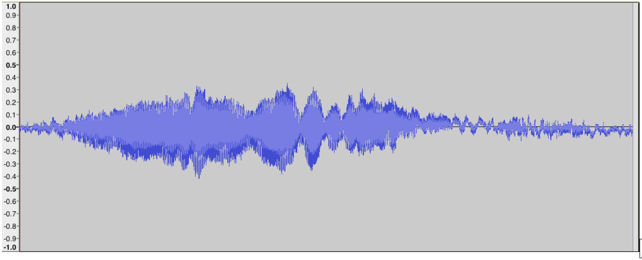
Following Mitchell (1997), we state the problem as follows:

- Task:
  Classify whale song unit instances
  according to a given vocabulary of whale song unit types

- Performance measure:
  Percent of instances correctly classified

- Training experience:
  A database of labelled whale song unit instances

Whale Song Units



Play

A "modulated cry" instance, from `HBSe_20151207T070326.wav, 124.5sec-126.5sec`

- Acoustic signal:

$$\mathbf{x} = \langle x_1, x_2, \ldots, x_N \rangle \qquad (N = 66,283)$$

Linear Predictive Coding

- Acoustic signal:

$$\mathbf{x} = \langle x_1, x_2, \ldots, x_N \rangle$$

- Transformed into a sequence of predictor vectors:

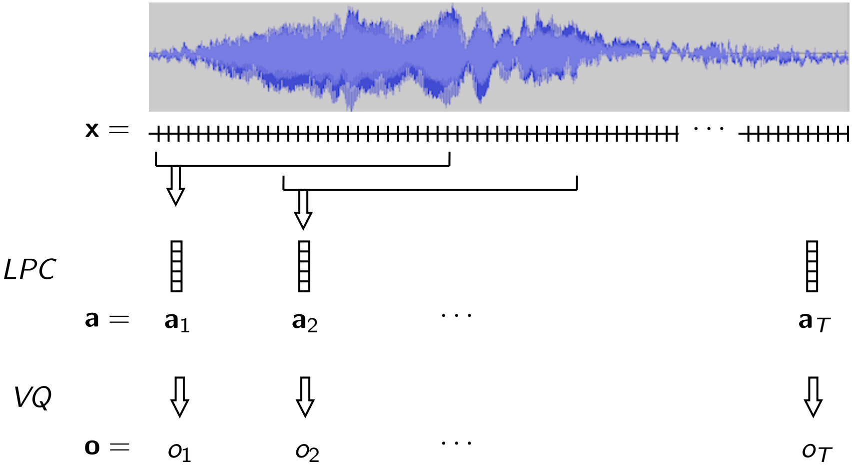
$$\mathbf{a} = \langle \mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_T \rangle$$

- Transformed into a sequence of symbols:

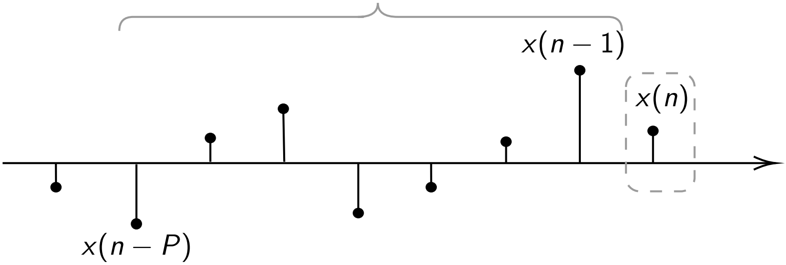
$$\mathbf{o} = \langle o_1, o_2, \ldots, o_T \rangle$$

where $o_t \in \{1, 2, \ldots, M\}$

Linear Predictive Coding



Linear Prediction



- Estimate $x(n)$ as a linear combination of P previous samples:

$$\hat{x}(n) = -\sum_{i=1}^{P} a(i)x(n-i)$$

Linear Prediction

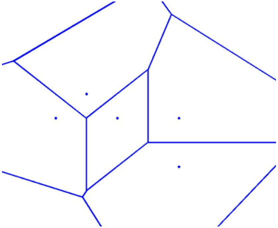
- Error:

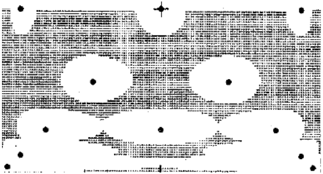
$$\begin{aligned} e(n) &= x(n) - \hat{x}(n) \\ &= \sum_{i=0}^{P} a(i)x(n-i) \end{aligned}$$

- Find prediction coefficients by minimizing the sum of the squared error over the signal interval:

$$\begin{aligned} E &= \sum_{n} e^2(n) \\ &= \sum_{n} \left(\sum_{i=0}^{P} a(i)x(n-i) \right)^2 \end{aligned}$$

Vector Quantization





- Partition the P-dimensional space into M regions that "best" represent the predictor vectors arising from song units instances
- "Best" in terms of minimizing some overall distortion measure
- Let $\mathbf{C} \equiv \langle C_1, C_2, \ldots, C_M \rangle$ be the set of centroids that best represent the space

Vector Quantization

- Quantization:
  With $\mathbf{C}$, we can map each $\mathbf{a}^{(t)}$ to a symbol o_t:

$$o_t \leftarrow \operatorname*{argmin}_{k=1}^{M} d(\mathbf{a}^{(t)}, C_k)$$

Hidden Markov Model

- π, initial state probability distribution:

$$\pi = (\pi_1, \pi_2, \ldots, \pi_N)$$

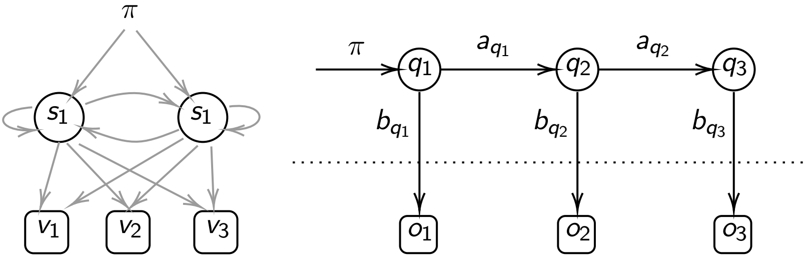
- A, state transition distributions:

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1N} \\ a_{21} & a_{22} & \ldots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \ldots & a_{NN} \end{bmatrix}$$

- B, observation symbol probabilities:

$$B = \begin{bmatrix} b_{11} & b_{12} & \ldots & b_{1M} \\ b_{21} & b_{22} & \ldots & b_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ b_{N1} & b_{N2} & \ldots & b_{NM} \end{bmatrix}$$

Hidden Markov Modeling



- $\mathbf{q} = \langle q_1, q_2, \ldots, q_T \rangle$: hidden state sequence
- $\mathbf{o} = \langle o_1, o_2, \ldots, o_T \rangle$: our acoustic signal

HMM Operations

With $\lambda = (\pi, A, B)$ denoting an HMM

- For given $\mathbf{o}$ and λ, compute $P[\mathbf{o}|\lambda]$
- For given $\mathbf{o}$ and λ, estimate a most likely state sequence $\mathbf{q}^*$
- Learn a model $\lambda^* \equiv (\pi^*, A^*, B^*)$
  such that $P[\mathbf{o}|\lambda^*]$ is maximized for a given observation sequence $\mathbf{o}$

Statistical Pattern Classification

- Vocabulary of R song unit types:

$$U = \{u_1, u_2, \ldots, u_R\}$$

- Assume we have a model h_r for each unit type u_r.
  We are interested in computing:

$$P[h_r|\mathbf{x}] \equiv \text{Probability of } h_r \text{ given } \mathbf{x}$$

- Classification of an observation sequence $\mathbf{x}$
  is based on a maximum likelihood model selection:

$$h^* \equiv \underset{h \in H}{\operatorname{argmax}} \, P[h|\mathbf{x}]$$

  where $H = \{h_1, \ldots, h_R\}$.

Maximum Likelihood model selection

$$
\begin{aligned}
h^* &\equiv \operatorname*{argmax}_{h \in H} P[h|\mathbf{x}] \\
&= \operatorname*{argmax}_{h \in H} \frac{P[\mathbf{x}|h]P[h]}{P[\mathbf{x}]} && \text{(Bayes' theorem)} \\
&= \operatorname*{argmax}_{h \in H} P[\mathbf{x}|h]P[h] && \text{(}P[\mathbf{x}]\text{ is constant wrt }h\text{)} \\
&= \operatorname*{argmax}_{h \in H} P[\mathbf{x}|h] && \text{(Assuming models are equiprobable)}
\end{aligned}
$$