Un modelo de Markov para la segmentación automática de señales de audio

Rafael de Jesús Robledo Juárez
rrobledo@cimat.mx

Asesor: Dr. Salvador Ruíz Correa

Departamento de Ciencias de la Computación
Centro de Investigación en Matemáticas, Guanajuato

CIMAT

Presentación de Avance de Tesis, 2013

Outline

Motivación

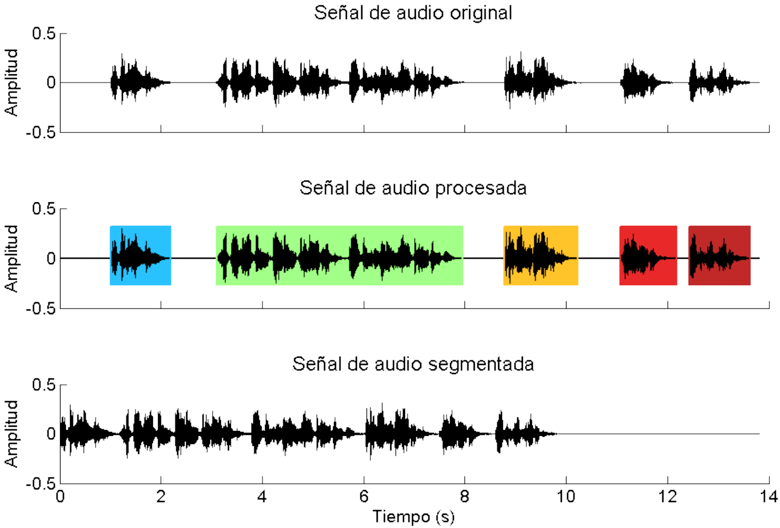
- La segmentación automática de una señal de audio en zonas homogéneas de acuerdo a las personas que hablan se conoce como *speaker diarization*.
- Es una etapa importante en el procesamiento de voz. Tanto para reconocimiento como transcripción de voz.
- Es un problema interesante para modelar, pues no se puede considerar que se tiene información a priori.

Trabajo previo

You can create overlays…

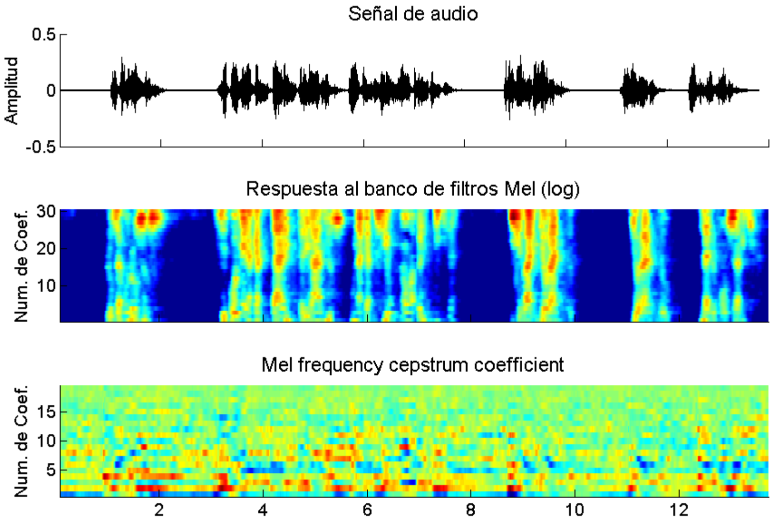
- using the `pause` command:
  - First item.
  - Second item.
- using overlay specifications:
  - First item.
  - Second item.
- using the general `uncover` command:
  - First item.
  - Second item.

Elminación de ruido / Detección de silencios



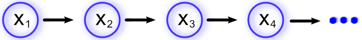
Mel Frequency Cepstrum Coefficient

- FFT (ventana) -> Banco de filtros triangular (Mel Scale) -> Log -> DCT -> MFCC



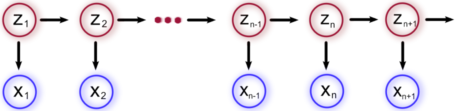
Hidden Markov Model

- Cadena de Markov de primer orden



$$p(x_1, ..., x_N) = \prod_{n=1}^{N} p(x_n \mid x_1, ..., x_{n-1}) = p(x_1) \prod_{n=2}^{N} p(x_n \mid x_{n-1}) \qquad (1)$$

- Modelo estocástico de Markov en el que el estado de la cadena es parcialmente observado.



- Modelar proceso bivariado en el tiempo. Una variable observada y una variable latente asociada.

Hidden Markov Model

- Agregar una variable latente z_n (discreta), que corresponda a cada observación x_n.

$$z_{n+1} \perp z_{n-1} \mid z_n \tag{2}$$

$$p(x_1, ..., x_N, z_1, ..., z_N) = p(z_1) \left[\prod_{n=2}^{N} p(z_n \mid z_{n-1}) \right] \prod_{n=1}^{N} p(x_n \mid z_n). \tag{3}$$

- Mezcla de distribuciones en la que la densidad tiene un distribución dada por $p(x|z)$

Parámetros del HMM

- Probabilidad de cambio entre estados dada una matriz de transición **A**

$$A_{jk} \equiv p(z_{nk} = 1 \mid z_{n-1,j} = 1) \tag{4}$$

$$p(z_n \mid z_{n-1}, \mathbf{A}) = \prod_{k=1}^{K} \prod_{j=1}^{K} A_{jk}^{z_{n-1,j} \cdot z_{n,k}} \tag{5}$$

- Vector de distribución inicial $\boldsymbol{\pi}$ para variable latente.

$$\pi_k \equiv p(z_{1k}) \tag{6}$$

$$p(z_1 \mid \pi) = \prod_{k=1}^{K} \pi_k^{z_{1k}} \tag{7}$$

- Probabilidad de emisión de una variable observada x_n dada una variable latente z_n.

$$p(x_n \mid z_n, \phi) = \prod_{k=1}^{K} p(x_n \mid \phi_k)^{z_{nk}} \tag{8}$$

HMM con EM

- Probabilidad conjunta del modelo

$$p(\mathbf{X}, \mathbf{Z} \mid \theta) = p(z_1 \mid \pi) \left[\prod_{n=2}^{N} p(z_n \mid z_{n-1}, \mathbf{A}) \right] \prod_{n=1}^{N} p(x_n \mid z_n, \phi) \tag{9}$$

- Parámetros del modelo $\theta = \{\boldsymbol{\pi}, \mathbf{A}, \phi\}$
- Función de verosimilitud completa

$$\mathcal{Q}(\theta, \theta^{old}) = \sum_{\mathbf{Z}} p(\mathbf{Z} \mid \mathbf{X}, \theta^{old}) \log p(\mathbf{X}, \mathbf{Z} \mid \theta) \tag{10}$$

- Prob. marginal de una variable latente, prob. conjunta de dos variables latentes consecutivas

$$\gamma(z_n) = p(z_n \mid \mathbf{X}, \theta^{old}) \tag{11}$$

$$\xi(z_{n-1}, z_n) = p(z_{n-1}, z_n \mid \mathbf{X}, \theta^{old}) \tag{12}$$

HMM con EM

- Prob. marginal de $z_n k = 1$, prob. conjunta de $z_{n-1,j}, z_{nk}$

$$\gamma(z_{nk}) = \mathbb{E}\left[z_{nk}\right] = \sum_{Z} \gamma(\mathbf{z}) z_{nk} \tag{13}$$

$$\xi(z_{n-1,j}, z_{nk}) = \mathbb{E}\left[z_{n-1,j} \cdot z_{nk}\right] = \sum_{Z} \gamma(\mathbf{z}) z_{n-1,j} \cdot z_{nk} \tag{14}$$

- Función de verosimilitud completa (reescrita con (13), (14))

$$\begin{aligned} \mathcal{Q}(\theta, \theta^{old}) = \sum_{k=1}^{K} \gamma(z_{1k}) \log \pi_k + \sum_{n=2}^{N} \sum_{j=1}^{K} \sum_{k=1}^{K} \xi(z_{n-1,j}, z_{nk}) \log A_{jk} + \\ \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \log p(x_n \mid \phi_k) \end{aligned} \tag{15}$$

- Parámetros estimados por EM:

$$\pi_k = \frac{\gamma(z_{1k})}{\sum_{j=1}^{K} \gamma(z_1 j)}, \quad A_{jk} = \sum_{n=2}^{N} \frac{\xi(z_{n-1,j}, z_{nk})}{\sum_{l=1}^{K} \xi(z_{n-1,j}, z_{nl})} \tag{16}$$

Algoritmo backward-forward

$$\gamma(z_n) = p(z_n \mid X) = \frac{p(X \mid z_n)p(z_n)}{p(X)} \tag{17}$$

$$\gamma(z_n) = \frac{p(x_1, ..., x_n, z_n)p(x_{n+1}, ..., x_N \mid z_n)}{p(X)} \tag{18}$$

$$\gamma(z_n) = \frac{\alpha(z_n)\beta(z_n)}{p(X)} \tag{19}$$

$$\tag{20}$$

donde

$$\alpha(z_n) \equiv p(x_1, ..., x_n, z_n) \tag{21}$$

$$\beta(z_n) \equiv p(x_{n+1}, ..., x_N \mid z_n) \tag{22}$$

$$\alpha(z_n) = p(x_n \mid z_n) \sum_{z_{n-1}} \alpha(z_n \mid z_{n-1}) \tag{23}$$

$$\alpha(z_1) = p(z_1)p(x_1 \mid z_1) = \prod_{k=1}^{K} \{\pi_k p(x_1 \mid \phi_k)\}^{z_{1k}} \tag{24}$$

Algoritmo backward-forward

$$\beta(z_n) = \sum_{z_{n+1}} \beta(z_{n+1}) p(x_{n+1} \mid z_{n+1}) p(z_{n+1} \mid z_n) \tag{25}$$

Bootstrap

Numero fijo de speakers

Numero variable de speakers

Resumen

- The first main message of your talk in one or two lines.
- The second main message of your talk in one or two lines.
- Perhaps a third message, but not more than that.

- Outlook
  - Something you haven’t solved.
  - Something else you haven’t solved.

Trabajo futuro

Referencias I

A. Author.
Handbook of Everything.
Some Press, 1990.

S. Someone.
On this and that.
Journal of This and That, 2(1):50–100, 2000.