¿Qué es una Cadena de Markov Escondida?
(*Hidden Markov Model*)

Joshua Kunst

jbkunst@gmail.com

4 de Agosto, 2007

Acá introduciremos el concepto de Cadena de Markov Escondida (CME o HMM, de su sigla en inglés *Hidden Markov Models*, también se conoce como Cadena de Markov Oculta), algo de lo que he estado leyendo hace algún tiempo. También veremos la idea detrás de una CME, un poco de la teoría y algún (canónico) ejemplo. Utilizaremos en la mayoría de las definiciones como referencia la introducción del libro *Inference in Hidden Markov Models* (O. Cappé, E. Moulines y Tobias Ryden; Springer).

La idea general de una Cadena de Markov Escondida está en el hecho de que al medir o cuantificar cierto fenómeno, la lectura que se genera -la secuencia observable de símbolos- no necesariamente es el proceso real del fenómeno, ya que el instrumento de medición podría estar introduciendo un ruido en la verdadera señal. Es por lo anterior que este tipo de modelamiento introduce dos procesos: uno observable y otro oculto. A continuación daremos una definición teórica más concreta.

Definición: Una Cadena de Markov Ocula es un proceso doblemente estocástico $\{(X_k, Y_k), k \in \mathbb{Z}^+\}$ (tembien nos refiriremos al proceso $(\mathcal{X}, \mathcal{Y})$) que cumple lo siguiente:

- El proceso $\mathcal{X}$ (oculto) cumple con ser una cadena de markov sobre el algún espacio de estados E, esto es:

  $$\Pr\{X_{n+1} = i_{n+1} | X_n = i_n, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\} = \Pr\{X_{n+1} = i_{n+1} | X_n = i_n\}.$$

- Que condicionalmente del proceso $\mathcal{X}$, para el proceso $\mathcal{Y}$ (de observaciones) se tiene que para cada n la distribución de Y_n depende solamente de X_n. En lenguaje probabilístico:

  $$\Pr\{y_1, y_2, \ldots, y_n | x_1, x_2, \ldots, x_n\} = \prod_{i=1}^{n} \Pr\{y_i | x_i\}.$$

Así, dada la definición, la distribución conjunta de un HHM es:

$$\Pr\{\{x_k, y_k\}_{k=1}^{n}\} = \prod_{i=2}^{n} \Pr\{y_i | x_i\} \Pr\{x_i | x_{i-1}\} \cdot \Pr\{y_1 | x_1\} \Pr\{x_1\}.$$

Sin embargo, aunque la distribución conjunta del proceso es fácil de ser calculada y agradable a la vista, hay que tener en cuenta que toda la inferencia que se hace sobre el fenómeno es en base a lo observado, por tanto se deber marginalizar la conjunta en el proceso latente (o no observado) $\mathcal{X}$. Y, debido a lo anterior, aparece una complicación en este tipo de modelos, pues si el proceso tiene largo n y la cardinalidad del espacio de estados (finito) tiene cardinalidad m entonces existen m^n posibles recorridos de la cadena oculta, lo que para n y m no tan grandes se vuelve un número grande difícil de manejar. La marginalización se obtiene a partir del teorema de probabilidades totales para cada y_i del proceso quedando:

$$\Pr\{\{x_k, y_k\}_{k=1}^n\} = \prod_{i=2}^n \{\sum_{x_i \in E} \Pr\{y_i|x_i\} \Pr\{x_i|x_{i-1}\}\} \cdot \{\sum_{x_j \in E} \Pr\{y_1|x_j\} \Pr\{x_j\}\}.$$

Lo que es complicado de trabajar. No obstante existen varios métodos que solucionan el problema. Los cuales no veremos con detalles acá.

Aplicación: Imagine que un amigo suyo realiza cierto experimento con una(s) moneda(s), al cual usted no tiene acceso de ver que tipo de experimento es, sin embargo su amigo le dice los resultados, vale decir solamente 'caras'(c) y 'sellos'(s). Entonces usted solamente conocerá una cierta secuencia, por ejemplo $\{c, s, c, c, s, c, \dots\}$. A usted le interesa saber que tipo de experimento realiza su amigo, entonces itentará hacer inferencia sobre el tipo de experimento a partir de la secuencia, quizás no sepa en que consiste pero tal ves encuentre un modelo que se ajuste de buena forma a los datos y que se riga de manera parecida al proceso oculto, para connocer más del fenómento y predecir una próxima secuencia.

Entonces, partiendo con la modelación se puede partir postulando un modelo simple para esta situación, algo simple sería suponer que el experimento se rige por una secuencia independiente e idénticas de variables aleatorias bernoulli con cierta probabilidad p. Quizás el modelo anterior se ajuste bien a los datos observados (y que para suerte de usted ese sea el experimento), sin embargo puede ocurrir que la secuencia tengas rachas, ejemplo $\{c, s, s, s, s, \dots, s, c, c, c, c, \dots\}$ lo que desecha el supuesto de independencia y de que probabilidad sea fija en el modelo anterior. Entonces usted se pregunta: ¿Qué pasa? Y se pone a pensar que el experiemento consiste en más de una moneda, y que algunas quizás están cargadas. Para modelar las rachas se puede postular que se tienen dos monedas y que al lanzar una moneda se sigue lanzando con la misma moneda con cierta probabilidad (dependiente de la moneda), si no se pasa a lanzar con la otra moneda, que en palabras más técnicas significa que hay un modelo markoviano en la selección de monedas. O quizás haya tipos de subsecuencias dentro de la gran secuencia, y así sucesivamente se puede complejizar el modelamiento (que no es la idea) sucesivamente hasta encontrar uno modelo que se adecúe a los datos observados. Principalmente el número de monedas se trabajan con *clusters*, cosa que veremos más adelante (espero).