TELOS. Revista de Estudios Interdisciplinarios en Ciencias Sociales
UNIVERSIDAD Rafael Belloso Chacín
ISSN 1317-0570 ~ Depósito legal pp: 199702ZU31
Vol. 14 (3): 433 - 438, 2012

Los modelos ocultos de Markov, MOM

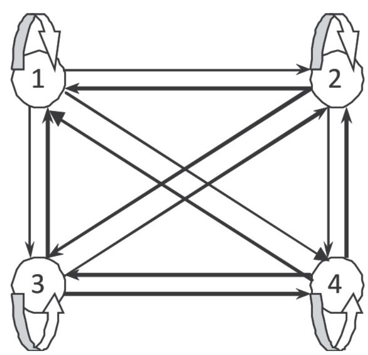
*Luciano Maldonado**

Definición de los modelos ocultos de Markov

Los Modelos Ocultos de Markov (MOM) son autómatas abstractos de estados finitos que permiten modelar procesos estocásticos, donde la ocurrencia de los estados está asociada con una distribución de probabilidad y donde las transiciones entre los estados están gobernadas por un conjunto de probabilidades llamadas probabilidades de transición de estados. En un estado particular, una observación se genera también de acuerdo a una distribución de probabilidad. Los estados no son visibles en general y su ocurrencia depende del estado en el instante anterior, de ahí el nombre de MOM (García, 2001; Peinado, 1994; Deller et al., 1993).

En la Figura 1 se muestra un MOM con una topología típica de 4 estados que admite transiciones hacia cualquier otro estado, incluso hacia sí mismo, en un instante cualquiera t.

Figura 1
Un MOM de cuatro estados completamente conectados



* Ing. de Sistemas. MSc. en Ingeniería de Control. Dr. en Ciencias Aplicadas. Profesor Titular, ULA. Facultad de Ingeniería, ULA. Director del Instituto de Estadística Aplicada y Computación, IEAC, FACES, ULA. Coordinador del Doctorado en Estadística, IEAC, FACES, ULA. Mérida, Venezuela. Correo electrónico: maldonaj@ula.ve

La selección de la topología de un MOM depende de la aplicación que se desee modelar.

Origen de los MOM

La historia de los MOM se remonta a los años cincuenta del siglo pasado, cuando un grupo de investigadores estaban tratando el problema de caracterizar procesos estocásticos para los cuales no contaban con suficientes observaciones (Deller et al., 1993). Surgió entonces la idea que implicaba modelar un proceso estocástico particular como un proceso estocástico doble, en el cual las realizaciones de un primer proceso (llamado el proceso oculto), daban origen a un segundo proceso (llamado el proceso observado). Los dos procesos se lograban caracterizar usando sólo el que se podía observar (Deller et al., 1993).

Del tratamiento de este problema surgió el algoritmo de identificación que se conoce como algoritmo de Máxima Estimación, ME (Deller et al., 1993).Luego, en los primeros años de la década de mil novecientos setenta se desarrolló un caso especial del algoritmo de ME, para estimar los parámetros de los MOM, el F-B (forward-backward) también llamado algoritmo de re-estimación Baum-Welch, en honor a sus creadores (Deller et al., 1993; Kanungo, 1998; Rabiner, 1989).

Utilidad de los MOM

Los sistemas del mundo real producen en general salidas o datos que se pueden tratar como señales. Dichas señales pueden ser de naturaleza discreta (por ejemplo, las salidas del lanzamiento sucesivo de un dado) o de naturaleza continua (por ejemplo, las medidas de la corriente eléctrica en un determinado ambiente). Éstas pueden ser estacionarias o no estacionarias según varíen o no sus propiedades estadísticas a través del tiempo y pueden estar corrompidas o no por otras señales de su entorno (Kanungo, 1998; Rabiner, 1989).En ese sentido, existe el problema fundamental de crear modelos para esas señales con la finalidad de que a partir de éstos se puedan describir teóricamente, simular, controlar y hasta construir los procesos generadores de dichas señales.

Los modelos de señales se clasifican en dos categorías: determinísticos y estocásticos. Los determinísticos, en general, explotan propiedades conocidas de las señales, mientras que en los estocásticos se intenta modelar solamente las propiedades estadísticas de la señal. Los MOM caen en esta última categoría.

Para fijar ideas respecto al uso de los MOM, se puede revisar el experimento siguiente (Deller et al., 1993): Considérese que en algún lugar de un salón hay un conjunto de cajas y bolas; y que en el salón hay un grupo de personas. Las personas se encuentran separadas del conjunto de las cajas y bolas, por una cortina, es decir, las personas no ven esos objetos. Supóngase que hay un número N de cajas numeradas y que cada caja contiene un número considerable de bolas de colores. Supóngase también que hay K distintos colores para las bolas.

El proceso físico para obtener las observaciones es como sigue: una persona, llamémosla El Mago, entra al lugar donde están los objetos mencionados y escoge una caja al azar, la caja inicial. De esta caja toma una bola también al azar y la muestra por encima de la cortina al grupo de personas; las personas anotan el color observado. El Mago vuelve a colocar la bola en la caja de donde fue extraída. Luego selecciona una nueva caja, de acuerdo al mismo proceso aleatorio con que seleccionó la primera, y toma una bola tal como lo hizo antes, y la muestra al público. Las personas vuelven a anotar el color.

Si se repite el procedimiento varias veces se tendrá una secuencia finita de observaciones de colores, que constituyen una realización del proceso observado. Aquí se puede notar que hay un proceso aleatorio oculto que da origen a ese proceso observado; ese proceso aleatorio oculto es el proceso aleatorio de la secuencia de cajas, es decir, la gente no sabe de qué caja o cajas provienen las bolas de colores, sin embargo, se dan cuenta que ocurre una realización al estilo: caja 1, caja 3, caja 6, caja 7, caja 1, ..., caja 4; que no pueden observar, pero que da origen a la realización observada de colores del tipo: azul, rojo, azul, verde, verde,...., rojo.

Un MOM permite modelar este tipo de experimento partiendo de la secuencia de salidas observadas.

Componentes de los MOM

Un MOM consiste formalmente de los siguientes elementos:

- El número de estados del modelo, **N.** En los MOM los estados están ocultos en general (son difíciles de definir), sin embargo, para algunas aplicaciones prácticas tienen significado físico. En el experimento mencionado, los estados corresponden a las cajas numeradas.
- El conjunto de los estados, E = {1,2,...,N}.
- El proceso aleatorio de los estados, $\underline{x}$. La secuencia de estados es el primero de dos procesos aleatorios asociados con un MOM.
- Las variables aleatorias asociadas al proceso aleatorio de los estados, $\underline{x}(t)$.
- El número de símbolos distintos que pueden ser observados en los estados (el número de los distintos colores de las bolas del experimento descrito), **K**.
- El conjunto de los símbolos distintos, V = {1,2,...,K} (los distintos colores de las bolas del experimento descrito).
- La longitud de la secuencia observable, **T**.
- La probabilidad de ocurrencia del estado **i** al inicio del experimento, $\prod(t)=\left[p(\underline{x}(t)=i)\right]$ en t=1.
- La matriz de probabilidades de transición de estado, A [a(i/j)].
- · La probabilidad de que ocurra el estado i en el instante t, dado que en t-1 ocurrió el estado gentlemen, a(i/j)= $P(\underline{x}(t)=i\ /\ \underline{x}(t-1)=j)$, para $1 \le i, j \le N$ y t arbitrario.

Las filas de **A** suman 1 debido a que en cualquier instante t ocurre una transición. Se asume que las a(i/j) son independientes del tiempo, es decir, no cambian durante el experimento.

La secuencia de las observaciones se modelan también como un proceso estocástico $\underline{y}$ (el segundo proceso aleatorio), con variables aleatorias, $\underline{y}(t)$, independientes e idénticamente distribuidas. Se asume que en un estado **i,** en un instante t, se genera una observación.

- La secuencia observada, y = {y(1), y(2),...,y(t),...,y(T)}(la secuencia de colores observados en el experimento descrito).
- El símbolo observado en el instante t, y(t) (el color observado en el instante t).
- La matriz de probabilidades de las observaciones, B[b(y(t)/i)].
- La probabilidad de que ocurra el símbolo u observación y(t) en el estado i, **e**n el instante t,b(y(t)/i)= p($\underline{y}$(t)= y(t)/ $\underline{x}(t)$=i).

Las probabilidades de las observaciones dependen del estado y son independientes de t. Las filas de la matriz **B** suman 1 debido a que siempre se genera una observación en el estado i en cualquier instante t.

En sentido formal un MOM comprende la siguiente estructura matemática:

$$m = \left\{ E, \prod(1), A, B \right\} \quad \textbf{(1)}$$

Bajo la condición de no preocuparnos por la historia de la secuencia de estados, dicha secuencia aleatoria (de primer orden) es un proceso markoviano (Rabiner, 1989; Juang y Rabiner S/F).

Clasificación de los MOM de acuerdo a los valores de las observaciones

Cuando el conjunto de símbolos **K** (distintos), con los que se forman las secuencias de observaciones, es muy grande (por encima de 256), se habla de secuencia de observaciones continuas. En ese caso, se usa una función de densidad de probabilidades multivariante continua, en lugar de un conjunto de probabilidades discretas para generar dichas observaciones (Rabiner, 1989; Juang y Rabiner, S/F). De esta manera se tiene la clasificación de los MOM en: MOM de observaciones discretas y MOM de observaciones continuas.

Los problemas asociados a los MOM

Una vez que se decide crear un MOM se encuentran tres problemas de interés por resolver (Rabiner, 1989; Deller et al., 1993):

- **El problema de Evaluación**: Dado un MOM m y una secuencia de observaciones y={y(1), y(2),...,y(t),...,y(T)}, cómo calcular la probabilidad de que esa secuencia sea generada por ese modelo. Se tiene el problema de calcular de manera eficiente la probabilidad de que las observaciones sean generadas por el modelo, la probabilidad P(y/m).

- **El problema de descodificación**: Dado un MOM m y una secuencia de observaciones y={y(1), y(2),...,y(t),...,y(T)}, hay que determinar la mejor secuencia de estados en el modelo, que produce a esa secuencia de observaciones. Se tiene el problema de encontrar la secuencia de estados que haga máxima a P(y/m).
- **El problema de entrenamiento**: Dado un MOM m y una secuencia de observaciones y={y(1), y(2),...,y(t),...,y(T)}, cómo ajustar los parámetros del modelo. Se está frente al problema de calcular el mejor conjunto $\left\{E, \prod(1), A, B\right\}$ con el fin de maximizar P(y/m).

Solución de los tres problemas asociados a los MOM

La construcción de un Modelo Oculto de Markov parte de un modelo inicial más o menos arbitrario, en el sentido de que la matriz de probabilidades de transición de estados y la matriz de probabilidades de producir las observaciones en los estados, contienen valores arbitrarios pero probabilísticos que cumplen con la regla de probabilidades de que la suma de las probabilidades de transición desde un estado a sí mismo y al resto debe dar uno; de la misma manera, la suma de las probabilidades de que un estado produzca, represente o modele cada observación debe dar igualmente uno.

Una vez que se tiene el MOM inicial se pueden resolver los tres problemas asociados a todo Modelo Oculto de Markov. Para resolver esos problemas existen varias técnicas algorítmicas, entre las que se destacan: los algoritmos de Viterbi y de Baum-Welch para la evaluación; un algoritmo de Viterbi para la descodificación; y para el entrenamiento otros algoritmos de Viterbi y de Baum-Welch (Rabiner, 1989; Deller et al., 1993;Kanungo, 1998; Savage, 1995).

Como se puede observar hay una serie de algoritmos que, en general, trabajan de la misma manera cuando se abordan aplicaciones de MOM, bien sean de observaciones discretas o de observaciones continuas, sin embargo, la diferencia radica en la forma de estimar la probabilidad de ocurrencia de las observaciones en los estados.

Referencias Bibliográficas

DellerJohn; Proakis, Johny Jansen, John. (1993). **Discrete-Time Processing of Speech Signals**. Macmilllan Publishing Company. USA.

García, Pedro. (2001). **Reconocimiento Automático de Voz continua con Modelos Ocultos de Markov**. Tesis doctoral, Universidad de Granada, España.

Juang, Biingy Rabiner, Lawrence. (S/F). **Issues in Using Hidden Markov Models for Speech Recognition**. Speech Research Department, AT&T Bell Laboratories.USA.

Kanungo, Tapas. (1998). **Hidden Markov Models**. University of Maryland, USA.

Peinado, Antonio. (1994). **Reconocimiento de Voz mediante Modelos Ocultos de Markov: Selección y Estimación de parámetros**. Tesis doctoral, Universidad de Granada, España.

Rabiner, Lawrence. (1989). **A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition**. Proceedings of The IEEE, Vol. 77, NO. 2, February. USA, (Pp. 257-286).

Savage, Jesús. (1995). **A hybrid System with Symbolic AI and Statistical Methods for Speech Recognition**. A dissertation for the degree of Doctor of Philosophy, University of Washington, USA.