Modelos Ocultos de Markov Continuos

José Antonio Camarena Ibarrola

Idea fundamental

- Los Modelos Ocultos de Markov Discretos trabajan con secuencias de observaciones
- Esto requiere del uso de cuantización Vectorial
- La cuantización vectorial implica pérdida de información
- Los modelos ocultos de Markov Continuos trabajan con secuencias de vectores de características
- Es decir, vectores de números reales que pertenecen a un espacio de características continuo

Tipos de Modelos Continuos

- Los Modelos ocultos de Markov continuos de mezcla de gausianas Modelan las funciones de densidad de probabilidad de vectores característicos mediante mezcla de gaussianas
- Este enfoque tiene una excelente capacidad de modelado pero requiere de una gran cantidad de datos de entrenamiento
- Los modelos ocultos de Markov semicontinuos combinan las distribuciones de probabilidad discretas con funciones de densidad tomadas de un libro de códigos modelado mediante una familia paramétrica de mezcal de gausianas

Ventajas de los Modelos Semicontinuos (SCHMM)

- los SCHMM tienen fijas las funciones de densidad (gaussianas) para los diferentes estados
- Solo cambian los pesos de las gaussianas de un estado a otro
- Esto reduce grandemente el número de parámetros a ser estimados
- No se necesitan tantos datos de entrenamiento como en los CHMM

SCHMM

- $b_j(\mathbf{x})$ es la función de densidad de emisiones asociadas al estado j
- La cual se puede modelar mediante mezcla de gaussianas como

$$b_j(\mathbf{x}) = \sum_{k=1}^{M} c_{jk} N(\mathbf{x}, \mu_{jk}, \Sigma_{jk})$$

- c_{jk} es el peso de la k-ésima gaussiana en el estado j
- M es el número de gaussianas de la mezcla

Interpretación

- El libro de códigos utilizado para los HMM discretos puede ser cambiado por una mezcla de gaussianas
- Cada entrada del libro de códigos puede ser representado por una de las gaussianas
- Una combinación de estas (mezcla) se puede usar para modelar un vector de características
- Comparando con HMM discretos se está minimizando el error de cuantización al evitar la partición implicita reemplazando por un modelado con traslapes entre las regiones
- Comparando con HMM continuos de mezcla se están utilizando las mismas gaussianas en los diferentes estados reduciendo el número de parámetros, la complejidad y el número de datos de entrenamiento requerido

SCHMM

- Asumiendo que cada entrada del libro de códigos está representado por una gaussiana
- Asumiendo tambien que que el estado y la observación contínua **x** son independientes

$$b_{S_t}(\mathbf{x}) = \sum_{j=1}^{L} f(\mathbf{x}|O_j, S_t)P(O_j|S_t) = \sum_{j=1}^{L} f(\mathbf{x}|O_j)b_{S_t}(O_j)$$

  donde L representa el número de entradas del libro de códigos
  $b_{S_t}(O_j)$ representa la probabilidad discreta de emitir el símbolo O_j en el estado S_t
- Dado O_j, la función de densidad $f(\mathbf{x}|O_j, S_t)$ puede ser estimada mediante el algoritmo EM

Estimación simultánea de SCHMM y libro de códigos

- Considerando a $b_i(O_j)$ como los pesos de las gausianas tendremos diferentes funciones de densidad (una mezcla de gausianas por cada estado)
- Las fórmulas de re-estimación pueden ser determinadas definiendo las variables $\alpha_t(i)$ y $\beta_t(i)$ de la siguiente manera

$$\alpha_t(i) = P(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_t, S_t = i|\lambda)$$

$$\beta_t(i) = P(\mathbf{x}_{t+1}, \mathbf{x}_{t+2}, ..., \mathbf{x}_T|S_t = i, \lambda)$$

Las cuales se pueden calcular recursívamente iniciando con $\alpha_t(i) = \pi_i$ y $\beta_T(i) = 1$

Definiciones de probabilidades intermedias

$$\chi_t(i,j,k) = P(S_t = i, S_{t+1} = j, O_k|\mathbf{X},\lambda)$$

$$\chi_t(i,j,k) = \alpha_t(i)a_{ij}b_j(O_k)f(\mathbf{x}_{t+1}|O_k)\beta_{t+1}(j)/P(\mathbf{X}|\lambda)$$

$$\gamma_t(i,j) = P(S_t = i, S_{t+1} = j|\mathbf{X},\lambda)$$

$$\gamma_t(i) = P(S_t = i|\mathbf{X},\lambda)$$

$$\xi_t(i,k) = P(S_t = i, O_k|\mathbf{X},\lambda)$$

$$\xi_t(k) = P(O_k|\mathbf{X},\lambda)$$

Continuación

Todas las probabilidades intermedias pueden obtenerse a partir de $\chi_t()$

$$\gamma_t(i,j) = \sum_{k=1}^{L} \chi_t(i,j,k)$$

$$\gamma_t(i) = \sum_{j=1}^{N} \gamma_t(i,j) = \sum_{j=1}^{N} \sum_{k=1}^{L} \chi_t(i,j,k)$$

$$\xi_t(i,k) = \sum_{j=1}^{N} \chi_t(i,j,k)$$

$$\xi_t(k) = \sum_{i=1}^{N} \sum_{j=1}^{N} \chi_t(i,j,k)$$

Fórmulas de reestimación

$$\bar{\pi}_i = \gamma_1(i)$$

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \gamma_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \qquad 1 \leq i,j \leq N$$

$$\bar{b}_i(O_j) = \frac{\sum_{t=1}^{T} \xi_t(i,j)}{\sum_{t=1}^{T} \gamma_t(i)} \qquad 1 \leq i \leq N \quad 1 \leq j \leq L$$

Reestimación de los parámetros de las gaussianas

$$\bar{\mu}_j = \frac{\sum_v \left[\sum_{t=1}^{T} \xi_t(j) \mathbf{x}_t \right]}{\sum_v \left[\sum_{t=1}^{T} \xi_t(j) \right]}$$

$$\bar{\Sigma}_j = \frac{\sum_v \left[\sum_{t=1}^{T} \xi_t(j) (\mathbf{x}_t - \bar{\mu}_j)(\mathbf{x}_t - \bar{\mu}_j)^t \right]}{\sum_v \left[\sum_{t=1}^{T} \xi_t(j) \right]}$$

- donde v denota el vocabulario completo y las expresiones en [] son variables del modelo para una palabra específica del vocabulario
- Las re-estimaciones para las medias y las varianzas se utilizan no solo para todos los estados de un modelo sino para todos los modelos del vocabulario

Algortimo Viterbi modificado

- Se utiliza la función de densidad contínua que reemplaza al libro de códigos como puente entre la secuencia de observaciones no cuantizadas y los parámetros del HMM discreto
- Utilizamos la siguiente ecuación en el Algoritmo Viterbi para encontrar la trayectoria mas probable de la secuencia de observaciones

$$b_{S_t}(\mathbf{x}) = \sum_{j=1}^{L} f(\mathbf{x}|O_j, S_t)P(O_j|S_t) = \sum_{j=1}^{L} f(\mathbf{x}|O_j)b_{S_t}(O_j)$$

- Para economizar cálculos se pueden utilizar solamente los valores mas significativos de $f(\mathbf{x}|O_j)$ (En la práctica de dos a 5 valores)

Algortimo Viterbi modificado

- La complejidad computacional del Viterbi modificado es menor que la del Viterbi de los HMM contínuos de mezcla
- Esto se debe a que $f(\mathbf{x}|O_j)$ solo se tiene que calcular para cada índice del libro de códigos en lugar de por cada estado y usando diferentes funciones de densidad
- Esto es particularmente cierto para vocabularios grandes
- los requerimientos de memoria son similares

Referencias

- Huang, X.D. and Jack M.A. “Unified Techniques for vector quantization and hidden markov modeling using semicontinuous models” Proceedings of the International Conference on Acoustics Speech and Signal Processing (ICASSP). 1989 pp 639-642 Liga