

Actividades Maestría, segundo semestre 2016

Manuel Felipe Pineda L 1093223607

November 22, 2016

1 Estudiar las diferentes medidas de distancia entre modelos ocultos de Markov, propuestas en la literatura.

Notación: Un modelo oculto de Markov será descrito como: $\lambda_i = \{\mathbf{u}, \mathbf{A}, \mathbf{B}\}$, donde \mathbf{u} es el vector de probabilidad del estado inicial, \mathbf{A} es la matriz de transición, y \mathbf{B} es la matriz de probabilidad de emisión.

Se analizaron algunas medidas de distancia basadas en diferentes criterios:

- Enfoque probabilístico basado en el número (o distancia) de Kullback-Leibler [1] [2] [3]: $D_{KL}(\lambda_0, \lambda)$
- Distancia euclidiana en la probabilidad de emisión [2] : $D_{ec}(\lambda_0, \lambda)$ y $D_{mec}(\lambda_0, \lambda)$
- Probabilidad de co-emisión para modelos ocultos de Markov tipo left-right [4]
- Probabilidad de error de Bayes [5]

Los dos primeros enfoques permiten establecer medidas de distancia entre modelos discretos para cadenas ocultas de Markov de primer orden, a diferencia del tercer método que presenta una medida de distancia sólo para un tipo particular de cadenas de Markov (left-right). Sin embargo, el segundo método y su generalización descrita en [2] no tienen en cuenta los estados ocultos de la cadena de Markov, lo cual hace posible encontrar dos HMMs muy diferentes cuya distancia sea muy pequeña.

En el cuarto enfoque, se define la distancia entre dos modelos basados en métodos de clasificación, estos a su vez se basan en minimizar la probabilidad de error dado un conjunto de entrenamiento.

1.1 Simulación Monte Carlo para evaluación de la distancia

La distancia D_{KL} está basada en encontrar la diferencia de la probabilidad de una observación particular de emisiones, para dos modelos diferentes, más formalmente:

$$\log P(\mathbf{O} \mid \lambda_1) - \log P(\mathbf{O} \mid \lambda_2)$$

Esta probabilidad se puede encontrar para un camino de $N + 1$ estados ocultos S como:

$$P(\mathbf{O} \mid \lambda, S) = \mathbf{u}_{s_0} \prod_{i=1}^N \mathbf{A}_{s_{i-1}, s_i} \mathbf{B}_{s_i, O_i}$$

Usando programación dinámica, es posible encontrar la probabilidad de las observaciones \mathbf{O} para todos los posibles caminos de $N + 1$ estados ocultos, lo cual será denotado como:

$$\mu(\mathbf{O} \mid \lambda)$$

Dado esto es posible evaluar $D_{KL}(\lambda_0, \lambda)$ generando diferentes secuencias de observaciones de longitud T a partir del modelo λ_0 y finalmente evaluar la función μ para obtener la distancia:

$$D_{KL}(\lambda_0, \lambda) = \frac{1}{T} (\log \mu(\mathbf{O} \mid \lambda_0) - \log \mu(\mathbf{O} \mid \lambda))$$

2 Implementar por lo menos dos de estas medidas en Python

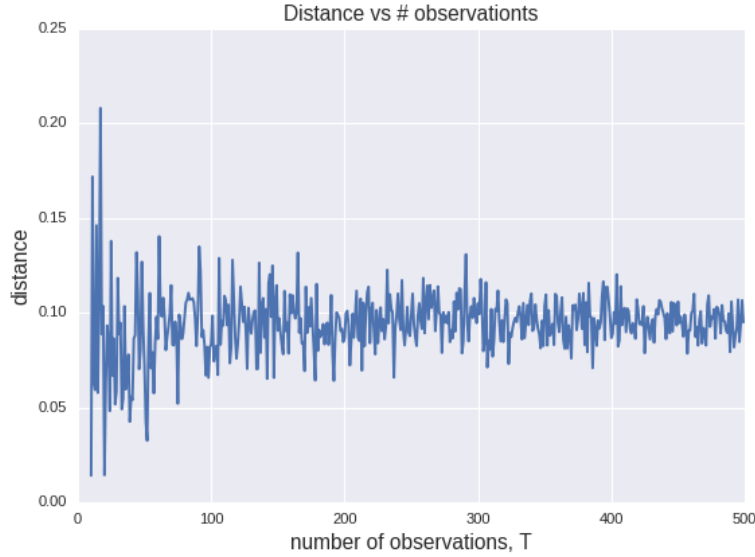


Figure 1: Distancia $D(\lambda_0, \lambda)$ vs el numero de observaciones

2.1 Distancia D_{KL}

Se implementó la distancia D_{KL} y se evaluó con los modelos provistos en [1], en la figura 1 se observa la convergencia de la distancia en la medida que incrementa

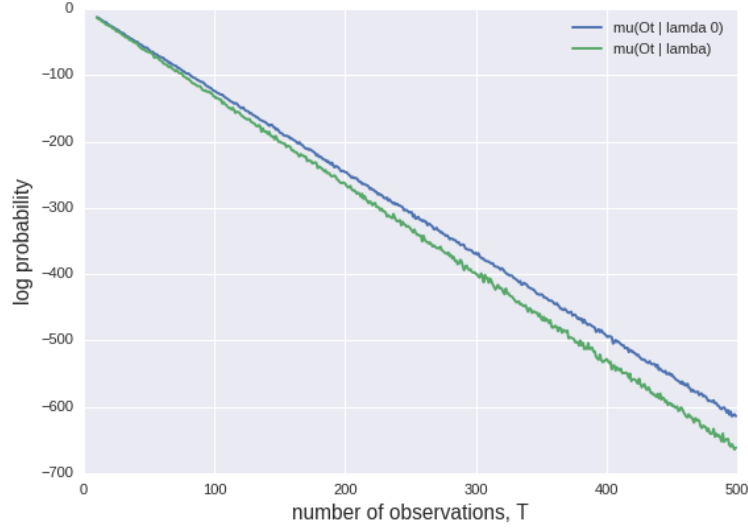


Figure 2: logaritmo de la probabilidad $\mu(O | \lambda_0)$ y $\mu(O | \lambda)$ definido en [1]

el número de observaciones.

En la figura 2 se compara el logaritmo de las probabilidades $\mu(O | \lambda_0)$ y $\mu(O | \lambda)$ obteniendo resultados similares a los de [1]

2.2 Distancia euclidiana

Se implementó también D_{ec} como se define en [2]

2.3 Distancia \tilde{D}_{vit}

En [2] se propone una variante para D_{KL} la cual no se enfoca en todos los posibles caminos para una observación, sino en el camino que maximiza la probabilidad de esa observación. Este camino es conocido como el *Viterbi path*. La implementación de esta distancia sería muy similar a D_{KL} usando una simulación por Monte Carlo y programación dinámica, debido a esto se decide implementar la aproximación propuesta, eq. (6) de [2] denotada como \tilde{D}_{vit}

2.4 Valores obtenidos

En esta sección se presentan los valores obtenidos en los métodos implementados para los siguientes modelos:

$$\mathbf{A}_0 = \begin{bmatrix} 0.8 & 0.15 & 0.05 & 0 \\ 0.07 & 0.75 & 0.12 & 0.06 \\ 0.05 & 0.14 & 0.8 & 0.01 \\ 0.001 & 0.089 & 0.11 & 0.8 \end{bmatrix}$$

$$\mathbf{B}_0 = \begin{bmatrix} 0.3 & 0.4 & 0.2 & 0.1 \\ 0.5 & 0.3 & 0.1 & 0.1 \\ 0.1 & 0.2 & 0.4 & 0.3 \\ 0.4 & 0.3 & 0.1 & 0.2 \end{bmatrix}$$

$$\mathbf{u}_0 = [0.75 \quad 0.15 \quad 0.05 \quad 0.05]$$

y

$$\mathbf{A}_1 = \begin{bmatrix} 0.4 & 0.25 & 0.15 & 0.2 \\ 0.27 & 0.45 & 0.22 & 0.06 \\ 0.35 & 0.14 & 0.4 & 0.11 \\ 0.111 & 0.119 & 0.23 & 0.54 \end{bmatrix}$$

$$\mathbf{B}_1 = \begin{bmatrix} 0.1 & 0.15 & 0.65 & 0.1 \\ 0.2 & 0.3 & 0.4 & 0.1 \\ 0.3 & 0.3 & 0.1 & 0.3 \\ 0.15 & 0.25 & 0.4 & 0.2 \end{bmatrix}$$

$$\mathbf{u}_1 = [0.4 \quad 0.25 \quad 0.15 \quad 0.2]$$

Método	Distancia	varianza
D_{KL}	0.0941689081332	0.000283555697027
D_{ec}	0.475657439761	
\tilde{D}_{vit}	-0.694147287288	

Table 1: Valores obtenidos

Vale la pena resaltar que \tilde{D}_{vit} presenta problemas de estabilidad numérica ya que su evaluación requiere calcular el logaritmo de las entradas de \mathbf{A} y de \mathbf{B} las cuales podrían ser cero, como en el ejemplo usado.

3 Implementar una medida de distancia entre modelos ocultos de Markov empleando embebimientos en Espacios de Hilbert

<http://www.cs.cmu.edu/~ggordon/song-boots-siddiqi-gordon-smola-hilbert-space-hmm.pdf>

<http://www.jmlr.org/papers/volume11/sriperumbudur10a/sriperumbudur10a.pdf>

http://www.kyb.mpg.de/fileadmin/user_upload/files/publications/attachments/ALT-2007-Gretton

https://www.cs.cmu.edu/~epxing/Class/10708-14/scribe_notes/scribe_note_lecture22.pdf

References

- [1] B-H Juang and Lawrence R Rabiner. A probabilistic distance measure for hidden markov models. *AT&T technical journal*, 64(2):391–408, 1985.

- [2] Markus Falkhausen, Herbert Reininger, and Dietrich Wolf. Calculation of distance measures between hidden markov models.
- [3] Li Xie, Valery A Ugrinovskii, and Ian R Petersen. Probabilistic distances between finite-state finite-alphabet hidden markov models. *IEEE transactions on automatic control*, 50(4):505–511, 2005.
- [4] Rune B Lyngso, CN Pedersen, and Henrik Nielsen. Metrics and similarity measures for hidden markov models. In *Proc Int Conf Intell Syst Mol Biol*, pages 178–186, 1999.
- [5] Claus Bahlmann and Hans Burkhardt. Measuring hmm similarity with the bayes probability of error and its application to online handwriting recognition.