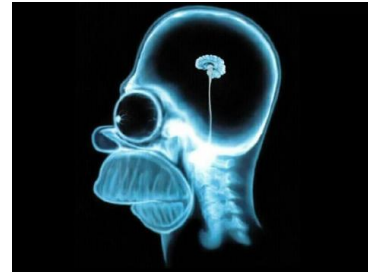


CI-1453: Investigación de Operaciones

**Segunda Tarea
Programada**



El objetivo de la siguiente tarea es la de evaluar el conocimiento adquirido sobre la codificación paramétrica efectiva de procesos estocásticos, y su medición probabilística de muestras, mediante la valoración de la medida: $P(O, q_{i=1..T} | \lambda)$. Esta valoración produce un modelo y método de clasificación/verificación mediante un Modelo Oculto de Markov (HMM).

Esta tarea consiste en la implementación de un sistema capaz de reconocer una firma manuscrita, obtenida por medio de una tableta electrónica. Por reconocer/clasificar entendemos la operación, mediante la cual valorando el trazo de una firma manuscrita, se puede determinar al individuo que efectuó dicha firma. Verificación, sería el proceso que sabiendo quién es el autor, se certifica su autoría.

Para efectuar este trabajo, deben de entregarse los siguientes documentos:

- 🏆 Debe presentarse un documento PDF, donde se responde los diferentes aspectos de la investigación. En la carátula del documento, primera página, deben de mencionarse los integrantes del grupo, carné y grupo del curso donde se encuentra matriculado.
 - Deben de responderse: estadísticamente justificando que se llega al resultado obtenido.
 - Las respuestas deben de presentarse claramente y apoyando los resultados obtenidos.
- 🏆 Debe de presentarse los programas desarrollados para responder a cada investigación solicitada: el HMM_DEF.m, las salidas de resultados, y el programa de codificación paramétrica efectiva, del trazo de la escritura manuscrita de la firma. Los programas deben de efectuarse en MatLab, todos desarrollados en la versión estándar de programas .m (NO SE PERMITEN VERSIONES GUI).
- 🏆 Finalmente todos los archivos digitales que componen el examen deben de comprimirse en conjunto en un RAR o ZIP que debe de ser “subido” al sistema: Mediación Virtual en la página del curso antes de la fecha límite. El nombre del archivo RAR o ZIP debe de ser el cané del encargado de la entrega.

¹ Escuela de Ciencias de la Computación e Informática
Facultad de Ingenierías, Universidad de Costa Rica, Prof. Briceño J.C.

²Este trabajo se puede efectuar en grupo, no mayor de tres miembros. Los integrantes pueden pertenecer a cualquiera de los dos grupos a cargo del profesor.

La tarea está compuesta de 2 estudios de investigación:

- 1- El modelo paramétrico del trazo de la firma manuscrita, para ser procesado por un Modelo Oculto de Markov discreto, HMM.
- 2- La configuración del clasificador: por un HMM, y la interpretación de qué representan los estados con respecto a esa codificación.

1. En el desarrollo del modelo paramétrico de la firma manuscrita, desarrolle los tres esquemas propuestos. Observe una descomposición del seguimiento expresado por:

1. Coordenadas angulares: ángulos de triangulación que excluyen el escalado.
2. Coordenadas polares: ángulo y amplitud que toma en cuenta el escalado y la orientación.
3. Finalmente sólo variación escalar que omite completamente la orientación angular.

Estas codificaciones expresan la idea de si dar o no importancia al tamaño de la firma. En la primera, se omite toda importancia al tamaño. En la segunda se hace un compromiso en ese sentido: un parámetro descriptivo de orientación angular y otro que expresa el tamaño en ese punto. Finalmente la tercera codificación sólo hace referencia al tamaño de la firma y no a su orientación.

Primero considere que las descripciones de puntos, se establecen de acuerdo a una ventana de definición de captura en la tableta. Es adecuado normalizar los datos de tal forma que se expresen abarcando solo la dimensión que estos puntos expresan.

Es decir del vector descriptivo VECTOR, presentado sólo con sus primeros 10 puntos por:

Línea: $x = 456 \ 458 \ 462 \ 462 \ 453 \ 449 \ 425 \ 403 \ 388 \ 375$
Columna: $y = 309 \ 307 \ 303 \ 302 \ 309 \ 311 \ 326 \ 341 \ 354 \ 367$

Efectúe las operaciones de encajado:

$$\begin{aligned}\text{VECTOR}(1,:) &= \text{VECTOR}(1,:) - \min(\text{VECTOR}(1,:)) + 1; \\ \text{VECTOR}(2,:) &= \text{VECTOR}(2,:) - \min(\text{VECTOR}(2,:)) + 1;\end{aligned}$$

El 1 al final de ambas ecuaciones, permite evitar hacer referencia a vectores con primera posición en 0, un error en MatLab.

² Escuela de Ciencias de la Computación e Informática
Facultad de Ingenierías, Universidad de Costa Rica, *Prof. Briceño J.C.*

³Observamos también que todas estas codificaciones son dependientes de la traslación, en la descripción de coordenadas, el tamaño de la ventana de captura y el tamaño de la firma.

Para evitar sesgos en ese sentido:

- a. Opte por fijar el centro de gravedad como origen descriptivo de los puntos. Entonces, cambie estas coordenadas cartesianas suministradas, que hacen referencia al origen de imágenes: punto superior izquierdo (0,0) como su origen descriptivo, por las suministradas por el cambio de origen.
- b. Para evitar problemas de dependencia con respecto al largo del vector de puntos (x,y) (tamaño de la firma) elija de forma regular un mismo número de puntos para las diferentes firmas por ejemplo 100, 150, 200, 250, 300 o 400 puntos.
- c. En lugar de considerar los valores de distancia de separación entre puntos, regularice con respecto al máximo de esas distancias (dividiendo por ese máximo la magnitud considerada). De esta forma usted obtiene una secuencia de variación relativa con respecto al tamaño general del trazo.

2- En la configuración del HMM no olvide de interpretar el significado de los estados del HMM con respecto a la estructura de la imagen del trazo de la firma, **justificándolo estadísticamente**.

Para efectuar la siguiente tarea se le suministra un Espacio de Trabajo MatLab que contiene 24 muestras de firmas de 14 individuos diferentes. Cada firma fue preprocesada de manera a obtener un trazo manuscrito compuesto de coordenadas con respecto a una ventana gráfica de captura de las muestras. Luego codificada como un vector de puntos (x,y) cartesianos.

Con respecto al vector de trazos dado en ejemplo, significa que el primer punto del trazo se encuentra en la línea 456 y en la columna 309 y así de seguido.

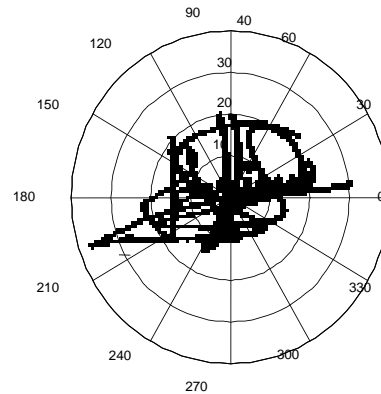
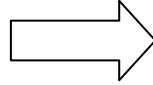
Note que el trazo no está compuesto de puntos continuos sino más bien de forma discontinua. Esto se debe a la frecuencia de muestro con que se capturó la información originalmente de la tableta. Esta varía según, si el trazo se expresa horizontal o verticalmente.

Los diferentes esquemas de codificación se ilustran y explican por medio de las figuras 1, 2 y 3.

³ Escuela de Ciencias de la Computación e Informática
Facultad de Ingenierías, Universidad de Costa Rica, *Prof. Briceño J.C.*



Imagen original



Descripción con origen en el centro de gravedad

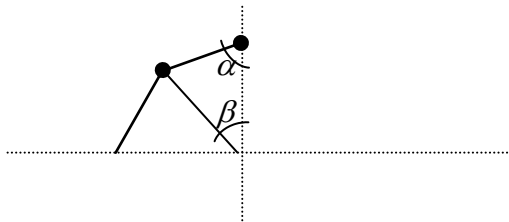


Figura 1. Cálculo del trazo de la firma por medio de ángulos.

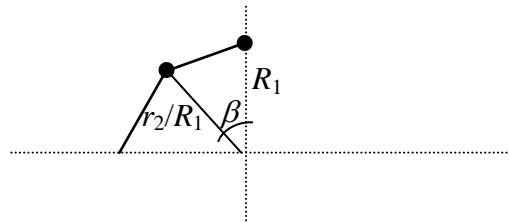


Figura 2. Cálculo del trazo de la firma por medio de coordenadas polares. Expresándose por diferencias con respecto al punto anterior

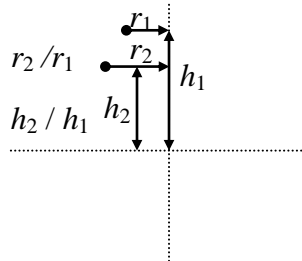


Figura 3. Cálculo de altos y anchos regularizados.

En la codificación de anchos y largos, la coordenada x le da la variación de largos, en positivos y negativos según su posición relativa con respecto al punto de origen y la coordenada y le da los altos también en positivos y negativos. Conserve los signos, esto le da una secuenciación de altos y largos, que refleja realmente la posición relativa del punto del trazo con respecto al punto de origen.

Recuerde que el algoritmo de k -medias que asigna etiquetas a los números multidimensionales, le convierte su problema de 2D en un problema 1D para poderlo procesar en un HMM.

⁴ Escuela de Ciencias de la Computación e Informática
Facultad de Ingenierías, Universidad de Costa Rica, Prof. Briceño J.C.