Instituto Tecnológico y de Estudios Superiores de Monterrey

Análisis de métodos de razonamiento e incertidumbre
MA2014.101

PBL3. Localización mediante Modelos Ocultos de Markov

Karla Andrea Palma Villanueva / A01754270
Daniela Márquez Campos / A00833345
Julio Eugenio Guevara Galván / A01704733
Adrian Pineda Sánchez / A00834710
David Fernando Armendáriz Torres/ A01570813
Kevin Antonio González Díaz / A01338316

Docente: Daniel Otero Fadul

Monterrey, Nuevo León, México. 10 septiembre 2023

Índice

1. Problematización

Los modelos ocultos de Markov son autómatas abstractos de estados finitos que permiten modelar procesos aleatorios, en los cuales la presencia de los estados se asocia con una distribución de probablidad en donde las transiciones entre los estados se ven regidas por un conjunto de probabilidades denominadas como probabilidades de transición de estados. Se define de igual forma como un modelo generativo que usa una cadena de Markov para modelar la evolución de los estados ocultos y una distribución de probabilidad condicional para modelar la relación entre los estados ocultos y las observaciones. [?]

La forma en la que funciona el modelo es asumiendo que las observaciones se encuentran conectadas exclusivamente con el estado actual del sistema, sin depender de estados previos o posteriores. Estos modelos se utilizan en una diversidad de aplicaciones, como el procesamiento del lenguaje hablado, el reconocimiento de la escritura manual, la predicción meteorológica o la categorización de señales biológicas, entre otros usos. Su finalidad principal es inferir la secuencia de datos oculta, ofreciendo una poderosa herramienta matemática para modelar procesos y situaciones generales, basado en las probabilidades de ocurrencia de etapas bien definidas dentro del comportamiento de tales procesos o situaciones y en observaciones indirectas de los mismos. [?]

De igual forma los modelos ocultos de Markov se implementan en métodos estadísticos y en entrenamientos no supervisados de redes neuronales, sistema de reconocimiento de patrones con necesidad de modelar correlaciones entre eventos adyacentes en base a secuencias de tiempo.

2. Enfoque

El caso de un agente -un robot, por ejemplo- que para localizar su propia ubicación puede discernir obstáculos en sus inmediaciones se puede considerar un sistema de Markov. En este tipo de sistemas, se supone que el estado actual depende de una cantidad finita de los estados previos. En esta situación, tal supuesto toma forma considerando los estados como las posiciones posibles del robot, por lo que el razonamiento es que para determinar la posición actual, se considera cierta cantidad de posiciones previas -sobre esto se profundizará posteriormente-.

Para efectos de este ejercicio específico, se definió un espacio discreto bidimensional compuesto por espacios disponibles y otros inaccesibles -los antes mencionados obstáculos- que en conjunto crean una cuadrícula de 4 filas por 16 columnas, como se muestra a continuación en la imagen 1, donde las celdas en color blanco son las disponibles.

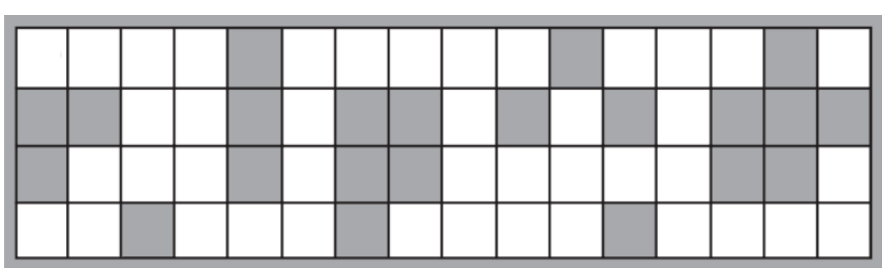

Figura 1: Entorno

3. Propósito

La aplicación de la teoría relacionada a los estados ocultos de Markov resulta aplicable y útil en situaciones en las que la información relevante para tomar decisiones se captura de manera discreta, lo cual llamamos 'evidencias', mismas que reflejan de alguna u otra forma la realidad -que no podemos observar-.

Muchas de las situaciones que cumplen con estas características se pueden encontrar en el ámbito de la navegación remota. Actualmente, una gran variedad de áreas de la ciencia y la sociedad se ven beneficiadas por el uso de robots o drones que pueden explorar lugares que serían difícil acceso o directamente inaccesibles para los seres humanos. Los ejemplos incluyen pero no se limitan a la agricultura, el rescatismo, la exploración interplanetaria y marina.

Si bien los sistemas de localización empleados para el manejo de los robots desempeñados en estas áreas pueden pueden variar significativamente, la aplicación de los modelos derivados de los procesos de Markov y el razonamiento probabilístico en el tiempo no deja de ser una alternativa tremendamente útil para numerosos casos.

4. Información

Profundizando en la base teórica sobre la cual se elabora una herramienta concreta que permita resolver situaciones como las previamente planteadas, se comienza por explorar un concepto clave en el razonamiento probabilístico en el tiempo, que es el modelo de transición, que de escribe matemáticamente de la siguiente manera:

$$P(X_k|X_{0:k-1}) \tag{1}$$

Lo que esta expresión describe es la distribución de probabilidad -es decir, la probabilidad de encontrarse en cada uno de los estados posibles- del estado actual dados los estados previos, donde k es un momento en el tiempo -que representamos de manera discreta, creando una sucesión de momentos individuales-, por lo que $X_{0:k-1}$ representa la información de los estados desde el momento 0 hasta el momento inmediatamente previo al actual.

En el aprendizaje automático se suelen hacer ciertos supuestos que sacrifican exactitud, en su mayoría de manera intrascendente, en aras de facilitar -como en el clasificador ingenuo de Bayes- o directamente posibilitar el desarrollo de modelos. En este caso, se realizan supuestos denominados "de Markov". Uno de estos supuestos se hace respecto al ya explicado modelo de transición, con lo cual se reescribe más simple como:

$$P(X_k|X_{k-1}) \tag{2}$$

Con esto se reduce la cantidad de información relevante para determinar el estado actual a exclusivamente el estado inmediatamente previo, en lugar de todos los estados anteriores. Otro de ellos se conoce como el supuesto de Markov del sensor, y se escribe matemáticamente de esta forma:

$$P(E_k|X_{0:k}, E_{1:k-1}) = P(E_k|X_k) \tag{3}$$

Es decir, para calcular la probabilidad de recuperar cierta evidencia en un momento K, se considera solamente el estado en ese mismo momento, despreciando las evidencias y los estados previos. Se podría decir que se asume que la información necesaria está contenida en el estado actual, lo cual implica que éste absorbe la información relevante de las evidencias y estados previos. Esto se aplica directamente en el modelo. Como se puede observar, estos supuestos son similares entre sí, basándose en los mismos razonamientos y proponiéndose con el los mismos propósitos: reducir la complejidad de los cálculos, lo cual, en el campo del aprendizaje automático, conlleva también disminuir el costo computacional.

Adicional a estas, también se consideran otras expresiones que se utilizan para determinar información relevante -lo cual es el objetivo del modelo-, de entre las cuales destaca pero no se limita al modelo del estado inicial $P(X_0)$.

Al generar un modelo temporal considerando entre muchas otras cosas, estas relaciones, se pueden llevar a cabo inferencias -lo que constituye el objetivo primordial-, de entre las cuales tenemos 4 básicas:

- Filtrado:
  Calcular la distribución de probabilidad de los estados posibles en el momento k dada la evidencia recolectada desde el primer hasta el momento que se quiere estimar. Matemáticamente se escribe como:

$$P(X_k|E_{1:k} = e_{1:k}) \tag{4}$$

- Predicción:
  Exactamente como su nombre lo sugiere, consiste en predecir un estado futuro. Esto se hace calculando la distribución de probabilidad del momento deseado -que se escribe como el momento $k + j$, donde j es la cantidad de momentos que faltan para llegar al punto que se desea predecir- dada toda la evidencia hasta la actualidad.

$$P(X_{k+j}|E_{1:k} = e_{1:k}) \tag{5}$$

- "Smoothing":
  Radica en calcular la distribución de probabilidad de los estados posibles en un momento pasado j, dada toda la evidencia hasta la actualidad.

$$P(X_j|E_{1:k} = e_{1:k}) \tag{6}$$

- Explicación más probable:
  Se estima la secuencia de estados más probable dada una serie de observaciones, por lo que en teoría se calculan las probabilidades de todas las secuencias de estados -de la misma longitud que las secuencias de observaciones- dada la secuencias de observaciones en cuestión para posteriormente compararlas, encontrar la más alta y extraer la secuencia de observaciones correspondiente a dicha probabilidad.

$$maxP(x_{1:k}|E_{1:k} = e_{1:k}) \tag{7}$$

A toda esta teoría estadística se le suma la programación necesaria para traducir este conocimiento a un código funcional. Para este trabajo se utilizó el lenguaje de programación Python, del fue particularmente importante el manejo de la librería numpy, la manipulación de sus matrices, listas anidadas, diccionarios y conversión de uno de estos formatos a los demás.

5. Razonamiento

El primer paso que se tomó para traducir la teoría a un código funcional fue declarar una matriz "de transición"que contiene las probabilidades de haber pasado de un estado a cada uno de los demás, o bien, la probabilidad de un estado dado el estado previo.

Posteriormente, se almacenó la información que describe a los estados en un diccionario, donde cada valor del mismo es una set de coordenadas correspondiente a un estado -un cuadro o coordenada disponible-. Las keys del diccionario son índices del 0 al 41 -la cantidad de cuadros disponibles-, y el orden en el que se almacenaron las coordenadas es de izquierda a derecha de abajo a arriba referente a la figura 1.

El paso siguiente consistió en elaborar una función que recibe el diccionario que contiene las coordenadas y con base en él, genere un arreglo bidimensional en el que se almacenan las direcciones relativas a cada cuadro/coordenada en las que hay espacios disponibles. Este arreglo se puede describir como una matriz de 42 filas y 4 columnas, donde en cada fila está la información del cuadro/coordenada correspondiente. Cada columna corresponde a una dirección, siguiendo la convención ESWN -este, sur, oeste, norte, en ese orden-. Por ejemplo, cuando la primera columna de un registro almacena un 1, implica que el cuadro al este del cuadro al que corresponde la fila está disponible. En el caso contrario, se almacena un 0.

Lo que siguió en el proceso fue declarar las 16 observaciones posibles -todos los números binarios posibles de 4 bits, ya que son 4 posiciones y 2 posibilidades para cada una, disponible o no- en un arreglo bidimensional de 16 filas y 4 columnas, con la misma mecánica utilizada en el arreglo explicado en el anterior párrafo y diseñar una función que calcula la discrepancia, para llamarla dentro de otra función que generará las 16 matrices -una por cada estado posible- diagonales de 42x42 que contienen la probabilidad de los estados dada la observación a la que corresponde la matriz. Es decir, cada elemento de la diagonal de la matriz que corresponde a la primera observación posible -digamos que es 1111, o bien, espacios disponibles en todas direcciones- representa la probabilidad de que se dé un estado específico dado que se registró dicha observación -el orden en el que están almacenados en el diccionario es el mismo en el que aparecen en la matriz-. Todo esto considerando un porcentaje de error que se puede ajustar alimentándole el valor deseado a la función, en conjunto con las observaciones posibles y la información de las adyacencias a cada estado/coordenada/cuadro.

Como último paso previo a la ejecución de las funciones que realizan las inferencias, se manipulan los arreglos y diccionarios de modo que se encuentren en el formato adecuado para alimentarse a dichas funciones.

Finalmente, se ejecutan las funciones de filtrado, smoothing y explicación más probable, además de visualizarse los resultados para facilitar la interpretación de los mismos y por lo tanto extraer información relevante.

6. Conclusiones

La primera inferencia que se realizó y visualizó fue filtrado. Ésta se probó de gran utilidad para determinar la localización del robot, ya que obtenemos la probabilidad de cada estado dada cierta evidencia. Experimentando con distintas cantidades de observaciones y rutas específicas trazadas, se concluyó que en el entorno particular en el que se llevó a cabo el ejercicio, tres observaciones fueron suficientes para tener una respuesta clara -en la que un estado tiene una probabilidad inmensamente mayor a los demás-. Esto se debe probablemente a que las secuencias de adyacencias casi nunca se repiten una vez que son tres observaciones, es decir, cada "tramo"de tres cuadros es usualmente único en el entorno, con sus limitadas excepciones.

Para la inferencia de *"smoothing"* no se realizó visualización, ya que la información que se genera a través de la misma, es regresada en forma de matriz, ya que se podría decir que el resultado es una matriz de transición actualizada con las observaciones que se le alimentó.

Finalmente, explicación más probable o secuencia de estados más probables se mostró

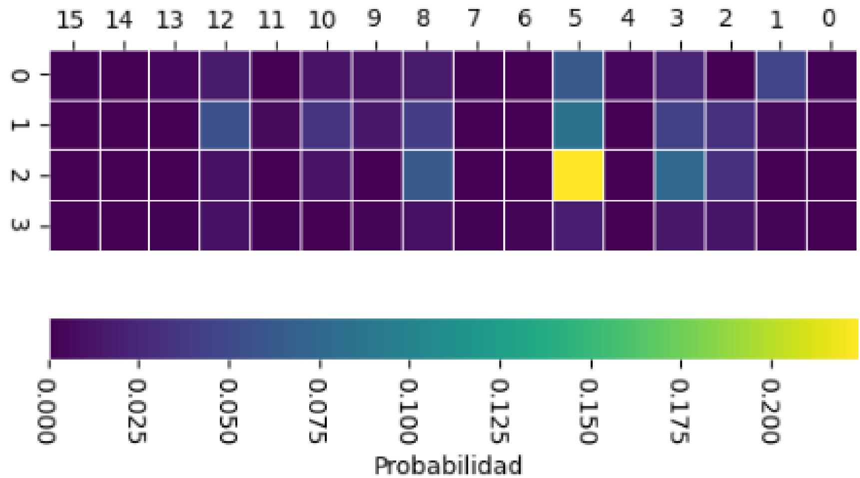

Figura 2: Filtrado realizado con 3 observaciones con 0.35 de error

igual de efectiva que inferencia, ya que en todos los casos en los que se experimentó alimentándole evidencia posible, llegó a la secuencia -que en este caso es una ruta- esperada.

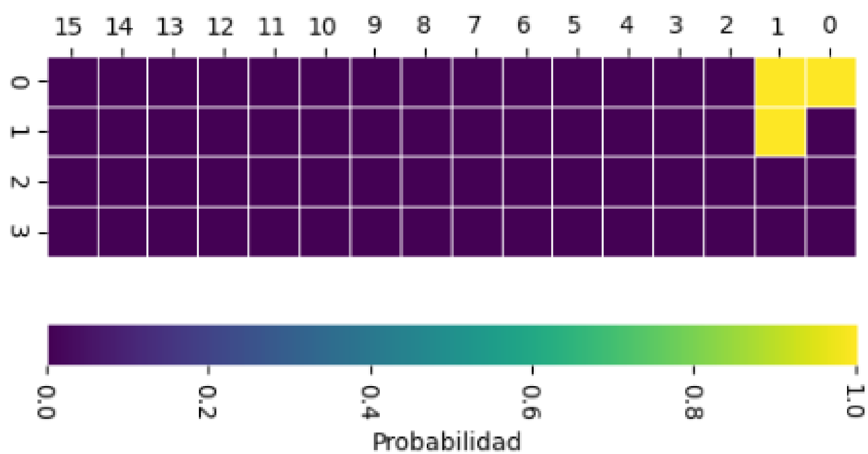

Figura 3: Secuencia más probable con 0.35 de error

Es necesario añadir que también se experimentó con distintas probabilidades de error, con las cuales se observó que la función de filtrado por lo general mantuvo efectividad con sólo 3 observaciones con debajo de 0.4 de tasa de error -en los casos en los que se perdió la efectividad antes de subir a ese punto la tasa de error fue en los que se emplearon los puntos previamente mencionados donde la ruta es similar a otra o incluso igual-. La función de secuencia de estados más probables se mostró más resistente al aumento en la tasa de error.

La aplicación de la teoría de Modelos Ocultos de Markov para la resolución de la problemática planteada como un tipo de razonamiento probabilístico en el tiempo, demuestra ser de gran utilidad con una eficiencia notable. Dada la comprensión de estas herramientas, la ejecución de una variedad de tipos de inferencias se realizó exitosamente, obteniendo resultados satisfactorios de forma general.

Referencias

[1] KeepCoding, R. (2023, 7 julio). ¿Qué es el modelo oculto de Markov?, KeepCoding Bootcamps. KeepCoding Bootcamps. https://keepcoding.io/blog/que-es-el-modelo-oculto-de-markov/#Funcionamiento_del_modelo_oculto_de_Markov

[2] Rafael Belloso Chacín. (2012, septiembre). Los modelos ocultos de Markov, MOM. TeloS. https://www.redalyc.org/pdf/993/99324907003.pdf

[3] Modelo oculto de Markov Numerentur.org. (s.f.). https://numerentur.org/markov-hmm/

[4] ámbito.com. (2023, 28 junio). Usan drones con inteligencia artificial para recolectar frutas. Ámbito Financiero. https://www.ambito.com/tecnologia/usan-drones-inteligencia-artificial-recolectar-frutas-n5756812

[5] CNN. (2023, 27 abril). Así usan robots y drones en Nueva York para tareas de rescate. https://cnnespanol.cnn.com/video/drones-robot-rescates-bomberos-nueva-york-toro-ush-pkg/

[6] Ferguson, A. (2021). Los robots eléctricos están cartografiando el fondo marino, la última frontera de la Tierra. CNN.https://cnnespanol.cnn.com/2021/10/27/robots-electricos-cartografiando-fondo-marino-trax/

[7] Johnson, A., Johnson, A. (2022). Así son los robots de la NASA en Marte y las investigadoras que los crean. Forbes España. https://forbes.es/actualidad/200600/asi-son-los-robots-de-la-nasa-en-marte-y-las-investigadoras-que-los-crean/#:~:text=Desde%20el%20Opportunity%20ha%20habido,sus%20equipos%20en%20la%20Tierra.

[8] Otero Fadul, D. (2023). Razonamiento probabilístico en el tiempo. Análisis de Métodos de Razonamiento e Incertidumbre. Tecnológico de Monterrey.