



UNIVERSIDAD AUTONOMA DE COAHUILA
FACULTAD DE SISTEMAS

INTEGRANTES:

RICARDO GABRIEL RODRIGUEZ GONZALEZ

JUAN CARLOS MEDINA RODRIGUEZ

MARFIK EMILIANO SEGURA CALZADA

MODELOS COMPUTACIONALES
REPORTE PROYECTO FINAL

PROFESORA: VALERIA SOTO MENDOZA

28/05/2022

Descripción general del proyecto

En este proyecto se analizarán, como cadenas de Markov de tiempo discreto con espacio de estados finito, las acciones que tuvo una muestra de alumnos dentro de sus navegadores web durante la aplicación de una evaluación en línea y se tratará de encontrar la relación entre dichas acciones y la calificación que obtuvieron. En especial, se buscarán las diferencias entre las cadenas de Markov de los alumnos que sí pasaron y las de aquellos que no pasaron. Los datos serán extraídos de archivos separados por delimitadores.

Metas

Conseguir los resultados y diferencias de aquellos alumnos que pasaron o acreditaron respecto a los que no pasaron a través de cadenas de Markov, esto gracias a los estados que en este caso tenemos, como eventos. Son 9 eventos.

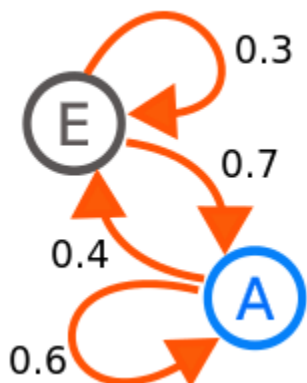
El modelo tiene como información una serie de datos, donde alumnos tienen una calificación específica, lo que haremos será primeramente llamarlos. El modelo realizará un análisis de todos los alumnos a través de cadenas de Markov que nos permitirá determinar las diferencias entre aquellos que aprobaron junto con los que no aprobaron, así como también podremos analizar entre esas mismas diferencias, que movimientos hicieron en comparación de otros en la plataforma.

Antecedentes

La cadena de Markov, también conocida como modelo de Markov o proceso de Markov, es un concepto desarrollado dentro de la teoría de la probabilidad y la estadística que establece una fuerte dependencia entre un evento y otro suceso anterior. Su principal utilidad es el análisis del comportamiento de procesos estocásticos.

La explicación de estas cadenas la desarrolló el matemático de origen ruso Andréi Márkov en 1907. Así, a lo largo del siglo XX, se ha podido emplear dicha metodología en numerosos casos prácticos de la vida cotidiana.

También se conoce como cadena simple biestable de Markov.



Según señaló Markov, en sistemas o procesos estocásticos (es decir, aleatorios) que presentan un estado presente es posible conocer sus antecedentes o desarrollo histórico. Por lo tanto, es factible establecer una descripción de la probabilidad futura de los mismos.

Más formalmente, la definición supone que en procesos estocásticos la probabilidad de que algo suceda solamente depende del pasado histórico de la realidad que estamos estudiando. Por este motivo, a menudo se dice que estas cadenas cuentan con memoria.

La base de las cadenas es la conocida como propiedad de Markov, la cual resume lo dicho anteriormente en la siguiente regla: lo que la cadena experimente en un momento $t + 1$ solamente depende de lo acontecido en el momento t (el inmediatamente anterior).

Dada esta sencilla explicación de la teoría, puede observarse que es posible a través de esta conocer la probabilidad de que un estado ocurra en el largo plazo. Esto ayuda indudablemente a la predicción y estimación en largos periodos de tiempo.

¿Dónde podemos encontrar las cadenas de Markov?

Las cadenas de Markov han experimentado una importante aplicación real en el ámbito de los negocios y las finanzas. Esto, al permitir, como se ha señalado, analizar y estimar futuros patrones de conducta de los individuos atendiendo a la experiencia y los resultados anteriores.

Lo anterior puede reflejarse en diferentes campos como la morosidad, el estudio de las conductas de consumidores, la demanda estacional de mano de obra, entre otros.

El sistema elaborado por Markov es bastante sencillo y cuenta, como hemos dicho, con una aplicación práctica bastante fácil. Sin embargo, muchas voces críticas señalan que un modelo tan simplificado no puede ser totalmente efectivo en procesos complejos.

Descripción del modelo

Parámetros

Para las cadenas de Markov se debe tener en cuenta que se necesitan de diversos parámetros

Como lo son la variable aleatoria, sus estados, así como su temporalidad.

Por lo que en este proyecto tenemos lo siguiente

Variable aleatoria: Movimientos que realiza un alumno aprobado y reprobado.

Estados $S_M = \{ \text{"Prueba LOGIN", "LOGIN", "Acceso Permitido", "Cambio de Pestaña", "Cambio de URL", "Creo una nueva pestaña", "Regreso al navegador", "Salió del navegador", "Usuario tenía sesión Activa"} \}$
\$9 estados/ Movimientos en el navegador.

Temporalidad: Por evento, ya que los intervalos no son constantes

Eventos

	A	B	C
1	Enrollment	Evento	FechaHora
2	9001	Regres— al navegador	10/04/2021 14:02
3	49455	[Login] Acceso concedido---sin Azure	10/04/2021 14:47
4	49455	Cambio de pesta-a	10/04/2021 14:48
5	49212	[Login] Acceso concedido---sin Azure	10/04/2021 14:48
6	49212	Salida del navegador	10/04/2021 14:48
7	49212	Regres— al navegador	10/04/2021 14:48
8	49570	[Login] Acceso concedido---sin Azure	10/04/2021 14:48
9	49455	Cambio de pesta-a	10/04/2021 14:48
10	49922	[Login] Acceso concedido---sin Azure	10/04/2021 14:49
11	49563	[Login] Acceso concedido---sin Azure	10/04/2021 14:49
12	49563	Salida del navegador	10/04/2021 14:49
13	49563	Regres— al navegador	10/04/2021 14:49
14	49690	[Login] Acceso concedido---sin Azure	10/04/2021 14:49
15	49585	[Login] Acceso concedido---sin Azure	10/04/2021 14:49
16	49455	Cambio de pesta-a	10/04/2021 14:49
17	49455	Cambio de pesta-a	10/04/2021 14:49
18	49269	[Login] Acceso concedido---sin Azure	10/04/2021 14:50
19	49269	Regres— al navegador	10/04/2021 14:50
20	49269	Salida del navegador	10/04/2021 14:50
21	49455	Cambio de pesta-a	10/04/2021 14:50
22	49455	Cambio de pesta-a	10/04/2021 14:51
23	49455	Cambio de pesta-a	10/04/2021 14:51
24	49455	Cambio de pesta-a	10/04/2021 14:51
25	49455	Cambio de pesta-a	10/04/2021 14:51
26	49455	Cambio de pesta-a	10/04/2021 14:52
27	49455	Cambio de pesta-a	10/04/2021 14:52
28	49647	[Login] Acceso concedido---sin Azure	10/04/2021 14:52
29	49671	[Login] Acceso concedido---sin Azure	10/04/2021 14:52
30	49551	[Login] Acceso concedido---sin Azure	10/04/2021 14:52
31	50615	[Login] Acceso concedido---sin Azure	10/04/2021 14:53
32	50615	Salida del navegador	10/04/2021 14:53
33	50615	Regres— al navegador	10/04/2021 14:53
34	50615	Salida del navegador	10/04/2021 14:53
35	50615	Regres— al navegador	10/04/2021 14:53
36	49186	[Login] Acceso concedido---sin Azure	10/04/2021 14:53

Calificaciones

	A	B	C
1	Enrollment	Final Grade	Explanación
2	49147 C	De vez en cuando, el examinado entiende frases y expresiones de uso frecuente relacionadas con áreas de su experiencia que le son relevantes tales como información básica sobre sí mismo, su familia, compras, lugares de interés, actividades de tiempo libre, etc.	
3	49148 NULL	NULL	
4	49152 D	El nivel de competencia lingüística del examinado es insuficiente para acreditar el Examen de Certificación de EXAVER 1 nivel A2. No acreditado el total de las cuatro habilidades evaluadas.	
5	49153 NULL	NULL	
6	49167 B	A menudo, el examinado entiende frases y expresiones de uso frecuente relacionadas con áreas de su experiencia que le son relevantes tales como información básica sobre sí mismo, su familia, compras, lugares de interés, actividades de tiempo libre, etc.	
7	49188 B	A menudo, el examinado entiende frases y expresiones de uso frecuente relacionadas con áreas de su experiencia que le son relevantes tales como información básica sobre sí mismo, su familia, compras, lugares de interés, actividades de tiempo libre, etc.	
8	49197 NULL	NULL	
9	49202 NULL	NULL	
10	49234 D	El nivel de competencia lingüística del examinado es insuficiente para acreditar el Examen de Certificación de EXAVER 1 nivel A2. No acreditado el total de las cuatro habilidades evaluadas.	
11	49247 E	El nivel de competencia lingüística del examinado es insuficiente para acreditar el Examen de Certificación de EXAVER 1 nivel A2.	
12	49256 C	De vez en cuando, el examinado entiende frases y expresiones de uso frecuente relacionadas con áreas de su experiencia que le son relevantes tales como información básica sobre sí mismo, su familia, compras, lugares de interés, actividades de tiempo libre, etc.	
13	49293 C	De vez en cuando, el examinado entiende frases y expresiones de uso frecuente relacionadas con áreas de su experiencia que le son relevantes tales como información básica sobre sí mismo, su familia, compras, lugares de interés, actividades de tiempo libre, etc.	
14	49314 C	De vez en cuando, el examinado entiende frases y expresiones de uso frecuente relacionadas con áreas de su experiencia que le son relevantes tales como información básica sobre sí mismo, su familia, compras, lugares de interés, actividades de tiempo libre, etc.	
15	49323 NULL	NULL	
16	49333 NULL	NULL	
17	49334 D	El nivel de competencia lingüística del examinado es insuficiente para acreditar el Examen de Certificación de EXAVER 1 nivel A2. No acreditado el total de las cuatro habilidades evaluadas.	
18	49344 E	El nivel de competencia lingüística del examinado es insuficiente para acreditar el Examen de Certificación de EXAVER 1 nivel A2.	
19	49345 C	De vez en cuando, el examinado entiende frases y expresiones de uso frecuente relacionadas con áreas de su experiencia que le son relevantes tales como información básica sobre sí mismo, su familia, compras, lugares de interés, actividades de tiempo libre, etc.	
20	49346 D	El nivel de competencia lingüística del examinado es insuficiente para acreditar el Examen de Certificación de EXAVER 1 nivel A2. No acreditado el total de las cuatro habilidades evaluadas.	
21	49350 B	A menudo, el examinado entiende frases y expresiones de uso frecuente relacionadas con áreas de su experiencia que le son relevantes tales como información básica sobre sí mismo, su familia, compras, lugares de interés, actividades de tiempo libre, etc.	
22	49375 C	De vez en cuando, el examinado entiende frases y expresiones de uso frecuente relacionadas con áreas de su experiencia que le son relevantes tales como información básica sobre sí mismo, su familia, compras, lugares de interés, actividades de tiempo libre, etc.	
23	49401 NULL	NULL	
24	49412 NULL	NULL	
25	49473 C	De vez en cuando, el examinado entiende frases y expresiones de uso frecuente relacionadas con áreas de su experiencia que le son relevantes tales como información básica sobre sí mismo, su familia, compras, lugares de interés, actividades de tiempo libre, etc.	
26	49484 NULL	NULL	
27	49494 NULL	NULL	
28	49534 NULL	NULL	
29	49559 NULL	NULL	
30	49568 B	A menudo, el examinado entiende frases y expresiones de uso frecuente relacionadas con áreas de su experiencia que le son relevantes tales como información básica sobre sí mismo, su familia, compras, lugares de interés, actividades de tiempo libre, etc.	
31	49576 NULL	NULL	
32	49582 -	Cancelado	
33	49591 NULL	NULL	
34	49598 NULL	NULL	
35	49609 D	El nivel de competencia lingüística del examinado es insuficiente para acreditar el Examen de Certificación de EXAVER 1 nivel A2. No acreditado el total de las cuatro habilidades evaluadas.	
36	49635 A	Generalmente, el examinado entiende frases y expresiones de uso frecuente relacionadas con áreas de su experiencia que le son relevantes tales como información básica sobre sí mismo, su familia, compras, lugares de interés, actividades de tiempo libre, etc.	

Desarrollo del programa

Búsqueda de las herramientas necesarias

Debido a la gran cantidad de registros presentes en los dos archivos proporcionados, era indispensable indagar respecto a herramientas que sirvieran para manejar semejante volumen de datos de una manera sencilla, rápida y legible. De lo contrario, se desperdiciaría mucho tiempo fabricando funciones que probablemente ya habrían sido creadas y probadas de manera más rigurosa por alguien más, además de que nos arriesgábamos a tener fallas cuyos efectos pasarían desapercibidos.

Una de las primeras herramientas que cumplían con la cualidades mencionadas anteriormente era la biblioteca de R purr, un paquete incluido en la colección Tydiverse, para la ciencia de datos. Pese a que incluía algunas funciones que posiblemente nos ayudarían a transformar los datos para su análisis, carecía de otras también indispensables para nosotros, por lo que consideramos a esta opción como incompleta. Además de que los objetos principales sobre los que operaba eran matrices y tablas, estructuras de datos que no podían contener los datos que queríamos, como matrices internas.

La segunda opción potencial, y con la cual nos quedamos, fue la biblioteca de R rlist (renkun-ken, s.f.). Poseía todas las formas de consulta y manipulación que pudiéramos necesitar, funcionaba a base de listas, que son un elemento más flexible que vectores, matrices o tablas, y además trabajaba de manera conjunta con otra biblioteca llamada pipeR, la cual ayuda a encadenar operaciones de una manera más breve y legible.

Por último, se requirió de una herramienta para graficar, de alguna forma, los resultados obtenidos a través de las operaciones previas y, así, tener un mejor panorama para el análisis y la obtención de conclusiones. Por su simplicidad, se escogió la biblioteca plot.matrix (Klinke, 2022), la cual sirve para obtener una representación visual de una matriz mediante un mapa de calor.

Código fuente

El código fuente, el cual se encuentra en el repositorio [marfikuac/6ModMarkov: Código en R del proyecto final de la clase de Modelos Computacionales sobre cadenas de Markov. \(github.com\)](https://github.com/marfikuac/6ModMarkov), está

escrito en el lenguaje de programación R en un único archivo. Se encuentra documentado para que personas inexpertas lo puedan comprender, así como para plasmar el proceso mental al desarrollarlo, para futuras referencias. Las únicas entradas son los valores de las rutas de los dos archivos que contienen los eventos y las calificaciones registradas. Los resultados son dos mapas de calor, obtenidos con la biblioteca descrita en la sección anterior, que representan, de manera visual, el promedio de las probabilidades de transición de estado (de un paso) de todos los alumnos que aprobaron y de los que reprobaron. Ya que los datos son proporcionados en archivos separados por delimitadores, el programa no requiere interacción por parte del usuario y simplemente se tiene que ejecutar para obtener los resultados.

El código fuente es el siguiente:

```
rutaArchivoEventos<-"6_bitacoraRegistrosNavegador_MC.csv"
rutaArchivoCalificaciones="1_2_Calificaciones_MC.csv"

#Para obtener la matriz de transición:
library(markovchain)
#Para las consultas y el procesamiento de los datos:
library(rlist)
#Para usar la sintaxis de tubería:
library(pipeR)
#Para el mapa de calor:
library(plot.matrix)

#Obtener los datos como un objeto de tipo data.frame.
eventos<-read.csv(rutaArchivoEventos,sep="\t",fileEncoding="x-mac-hebrew")
calificaciones<-read.csv(rutaArchivoCalificaciones,sep="\t",fileEncoding="x-mac-hebrew")

#Obtener los estados posibles.
states<-list.parse(eventos)%>>%
  list.group(Evento)%>>%
  list.map(.i-1L)

#Obtener solo el valor numérico de los estados (excluyendo el nombre).
M<-as.integer(states)

lst<-
  #Convertir a lista.
  list.parse(eventos)%>>%
  #Agregar campos previos de consulta y ordenamiento.

list.select(Enrollment,Evento,Evento2=states[[Evento]],Momento=as.POSIXct(FechaHora,format="%d/%m/%y %H:%M"))%>>%
  list.select(Enrollment,Evento,Evento2,Momento,Momento2=as.numeric(Momento))%>>%
  #Ordenar lista por fecha de manera ascendente.
  list.sort(Momento2)%>>%
  #Agrupar por matrícula.
  list.group(Enrollment)%>>%
  #Obtener solo la cadena de markov como vector de enteros.
  list.map(.%>>%list.map(Evento2))%>>%
  list.map(as.integer(.))%>>%
  list.select(Enrollment=.name,CadenaDeMarkov=.)%>>%
```

```

#Calcular y agregar la matriz de transición de estados
#(de un paso) con ayuda de la biblioteca markovchain.

list.select(Enrollment,CadenaDeMarkov,MatrizDeTransicion=markovchainFit(CadenaDeMarkov,possibleStates=M)$estimate@transitionMatrix)%>>%
#Adjuntar la matrícula obteniéndola del archivo de calificaciones.
list.join(list.parse(calificaciones),Enrollment)%>>%
#Formalizar las propiedades de cada elemento de la lista.
list.select(Enrollment,CadenaDeMarkov,MatrizDeTransicion,Calificacion=Final_Grade)

#Identificar las calificaciones aprobatorias y reprobatorias.
califsAprob<-c("A","B","C")
califsReprob<-c("D","E")

#Obtener solo los alumnos que aprobaron.
alumsAprob<-lst%>>%
  list.filter(Calificacion %in% califsAprob)

#Obtener solo los alumnos que aprobaron.
alumsReprob<-lst%>>%
  list.filter(Calificacion %in% califsReprob)

#(Opcional) Obtener los alumnos restantes (valores erróneos).
# alumsOtros<-lst%>>%
#   list.filter(!(Calificacion %in% c(califsAprob,califsReprob)))

#(Opcional) Obtener los alumnos agrupados por calificación.
# alumsPorCalif<-lst%>>%
#   list.group(Calificacion)

#Para una lista de matrices con las mismas dimensiones,
#obtener una matriz que sea el promedio de ellas,
#(Promedio de cada elemento en la misma posición de
#las matrices.)
calcularPromedioDeMatrices <- function(matrices, rows, cols) {
  result <- matrix(0, rows, cols)
  for (i in 1:rows) {
    for (j in 1:cols) {
      zSum <- 0
      for (k in 1:length(matrices)) {
        zSum <- zSum + matrices[[k]][i, j]
      }
      zMean <- zSum / length(matrices)
      result[i, j] <- zMean
    }
  }
  return (result)
}

#Calcular el promedio de transiciones de un estado a otro con las matrices de
transición de los alumnos aprobados y mostrar los resultados con un mapa de calor.
promedioAprobados<-
calcularPromedioDeMatrices(alumsAprob%>>%list.map(MatrizDeTransicion),length(M),length(M))
plot(promedioAprobados,breaks=range(promedioAprobados))

```

```
#Calcular el promedio de transiciones de un estado a otro con las matrices de
transición de los alumnos reprobados y mostrar los resultados con un mapa de calor.
promedioReprobados<-
calcularPromedioDeMatrices(alumsReprob%>%list.map(MatrizDeTransicion),length(M),length(M))
plot(promedioReprobados,breaks=range(promedioReprobados))
```

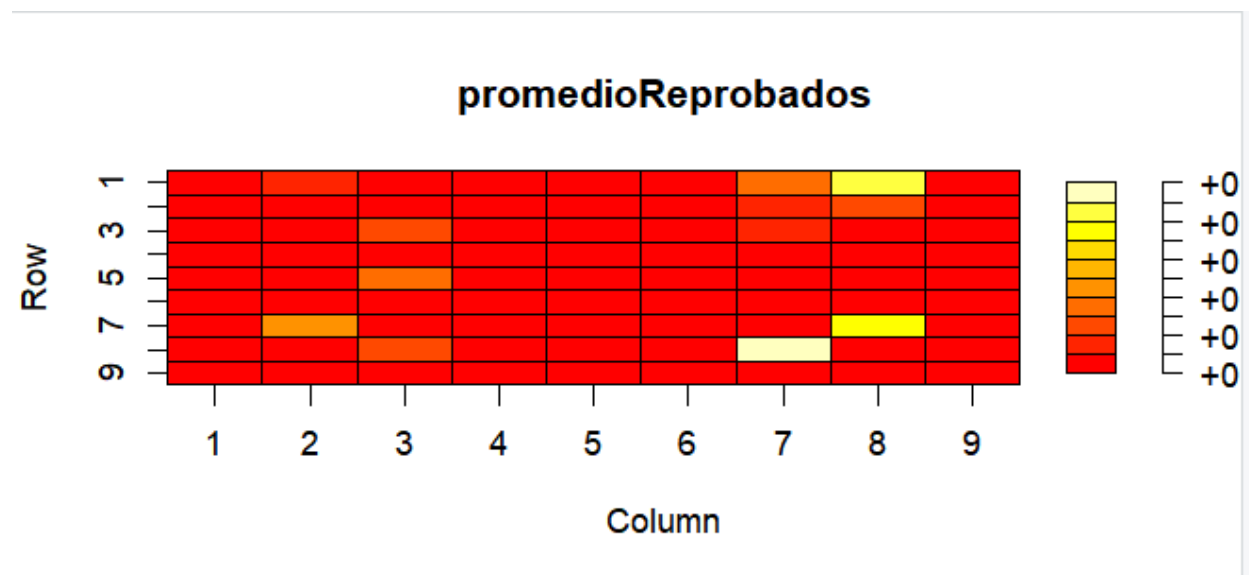
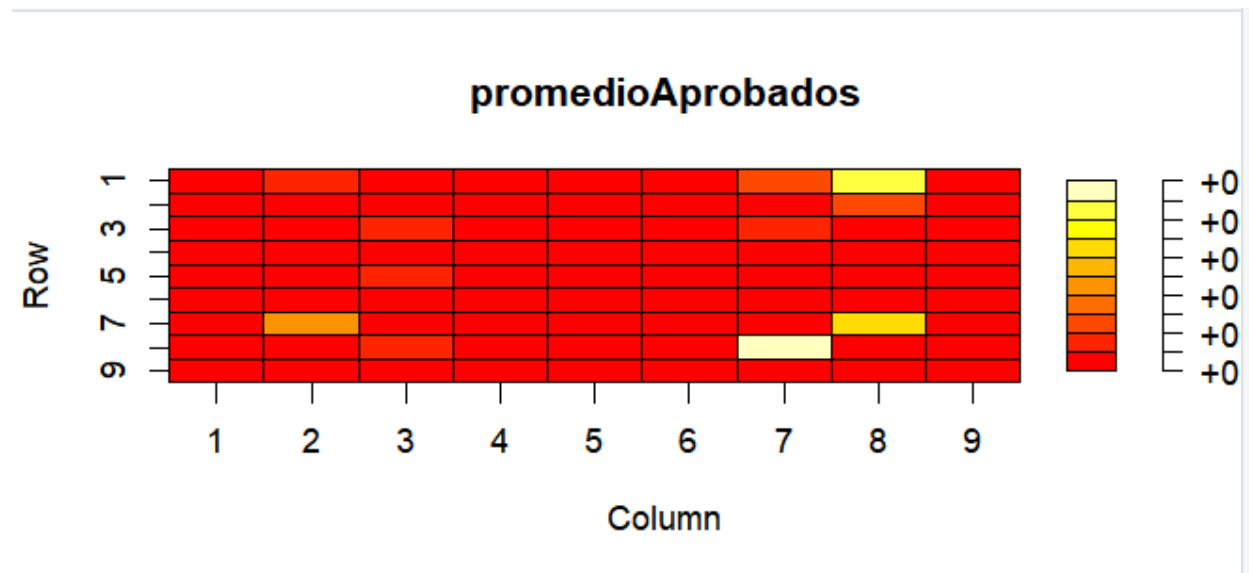
Para comprender mejor la estructura de los datos al irse procesando (o el valor de las variables), a continuación, se proporciona una visualización de las variables indispensables:

lst	Large list (340 elements, 1.1 MB)
\$ 49147:List of 4	
..\$ Enrollment	: int 49147
..\$ CadenaDeMarkov	: int [1:4] 0 7 6 1
..\$ MatrizDeTransicion:	num [1:9, 1:9] 0 0 0 0 0 0 0 0 0 0 ...
.. ..- attr(*, "dimnames")=	List of 2
..\$:	chr [1:9] "0" "1" "2" "3" ...
..\$:	chr [1:9] "0" "1" "2" "3" ...
..\$ Calificacion	: chr "C"
\$ 49152:List of 4	
..\$ Enrollment	: int 49152
..\$ CadenaDeMarkov	: int [1:23] 0 7 6 7 6 6 7 6 7 6 ...
..\$ MatrizDeTransicion:	num [1:9, 1:9] 0 0 0 0 0 0 0 0 0 0 ...
.. ..- attr(*, "dimnames")=	List of 2
..\$:	chr [1:9] "0" "1" "2" "3" ...
..\$:	chr [1:9] "0" "1" "2" "3" ...
..\$ Calificacion	: chr "D"
\$ 49160:List of 4	
estados	List of 9
\$ [Login] Acceso concedido---sin Azure	: int 0
\$ [PaperTwo] Acceso permitido	: int 1
\$ Cambio de pestaña	: int 2
\$ Cambio de URL	: int 3
\$ Creo una nueva pestana	: int 4
\$ PRUEBA [Login] Acceso concedido---sin Azure:	int 5
\$ Regresó al navegador	: int 6
\$ Salida del navegador	: int 7
\$ Usuario tenía sesion activa	: int 8

Resultados

Al ejecutar el programa se obtienen los siguientes mapas de calor como única salida. Al representar promedios de probabilidades de transición de estados (valores de 0 a 1), la escala que se muestra del lado derecho de las imágenes no pasa de uno y, por cuestiones de redondeo, no se alcanzan a ver decimales. Sin embargo, de acuerdo con la leyenda, un color cercano a rojo en una celda indica un valor cercano al valor mínimo de la matriz y, en contraste, mientras más blanco sea el color de una celda, más cercano estará al valor máximo de la matriz. La matriz es de 9 filas por 9 columnas, que corresponden a los 9 estados posibles de las cadenas de Markov con las que se modela este problema, los cuales van de 0 a 8. Obsérvese que la principal diferencia entre estos dos resultados (los promedios de las matrices de transición de estado de los alumnos que aprobaron y de los que reprobaron) se encuentra en la tercera

columna, indicando con un color más cercano al blanco, un valor mayor de las probabilidades de transición promedio al estado 2 para los alumnos que reprobaron.



Conclusiones

Una forma de analizar una cadena de Markov es mediante su matriz de transición de estados de un paso. Pero cuando se requiere analizar una serie de cadenas de Markov, surge la necesidad de encontrar algún valor representativo que no nos obligue a estar analizando un grupo grande de matrices de transición. Este valor representativo puede ser su promedio, el cual sigue siendo una matriz con las mismas dimensiones, pero deja de ser una matriz de transición porque pierde las propiedades de una.

Esta operación fue la que se realizó para encontrar alguna diferencia en el comportamiento de los alumnos que aprobaron y que reprobaron un examen en línea, mediante el modelamiento de sus acciones en el navegador usando cadenas de Markov (y sus respectivas matrices de transición).

De acuerdo con los resultados obtenidos en la sección previa, la diferencia principal entre estas dos matrices (que representan las matrices de transición individuales) es que la probabilidad promedio de pasar al estado 2 es mayor para el grupo que reprobó el examen en comparación con el grupo que lo aprobó. Dado que el estado 2 es el estado “Cambió de pestaña”, se puede concluir que los alumnos que reprobaron tienen, en promedio, la característica de que cambian más veces de pestaña en el navegador.

La resolución de este problema comenzó con una ardua búsqueda de alguna manera de analizar un grupo grande de cadenas de Markov (o sus matrices de transición de estados). La forma fue mediante su promedio. El siguiente reto fue encontrar alguna biblioteca en R que sirviera para transformar los datos de entrada para su posterior análisis. Esta fue la biblioteca “rlist”. Al final se encontró una forma de representar de manera visual los resultados.

Durante este proyecto, se ahondó en la comprensión y la utilidad de las cadenas de Markov en la vida diaria, así como en el uso del lenguaje R, lo cual sin duda servirá para realizar análisis futuros en esta área.

Bibliografía

Hiller F., Lieberman G. (2010). *Introducción a la investigación de operaciones*. McGraw Hill.

Klinke, S. (2022). *plot.matrix v1.6.2*. Retrieved from <https://cran.r-project.org/web/packages/plot.matrix/vignettes/plot.matrix.html>

renkun-ken. (s.f.). *rlist*. Obtenido de <http://renkun-ken.github.io/rlist/>

RStudio. (n.d.). *Tidyverse*. Retrieved from <https://www.tidyverse.org/>