

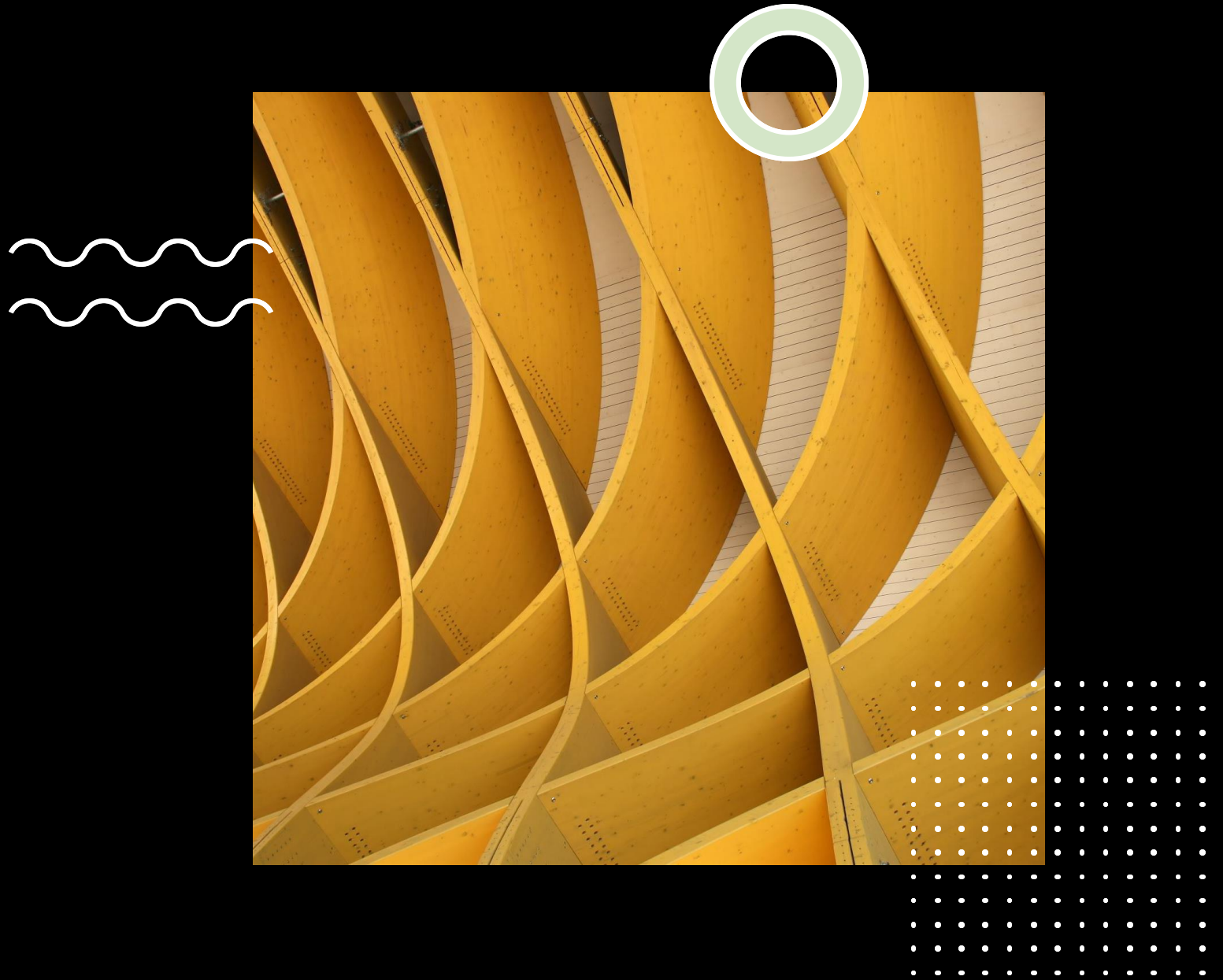
Patrones Secuenciales

Gerardo Mora Martinez 1849522

Orlando García Salazar 1811901

Cristina Vidal Mosqueda 1941502

Adriel Isaí Macías Cortés 1644680



Definición

- La minería de patrones secuenciales es la minería de patrones que ocurren frecuentemente relacionados al tiempo u a otras secuencias.
- Algunos ejemplos son, patrones de llamadas telefónicas, flujos de navegación en la web, secuencias de ADN.



- *El objetivo es buscar asociaciones de forma en que, si el evento X sucede en el tiempo t , sucederá en el tiempo $t + n$ el evento Y .*
- Los patrones secuenciales tienen ciertas características:
- El orden importa.
- Es una lista de datos, donde cada dato es un elemento de la secuencia.
- El soporte de la secuencia es el porcentaje de las secuencias que la contienen dentro de un grupo de secuencias S .
- Los patrones son las subsecuencias que se forman dentro de la secuencia de datos con un porcentaje mínimo de soporte.





- *Por ejemplo, esta técnica puede revelar qué artículos de ropa es más probable que compren los clientes después de una compra inicial de, por ejemplo, un par de zapatos. Entender los patrones secuenciales puede ayudar a las organizaciones a recomendar a los clientes artículos adicionales para estimular las ventas.*

Retos de ésta Técnica



- En las bases de datos se oculta un enorme número de posibles patrones secuenciales
- Un algoritmo de minería debería encontrar el conjunto completo de patrones, cuando posible, que satisfaga el umbral de apoyo mínimo (frecuencia) mínimo
- Ser altamente eficiente, escalable, y requerir sólo un número de exploraciones de la base de datos
- Ser capaz de incorporar varios tipos de restricciones específicas del usuario

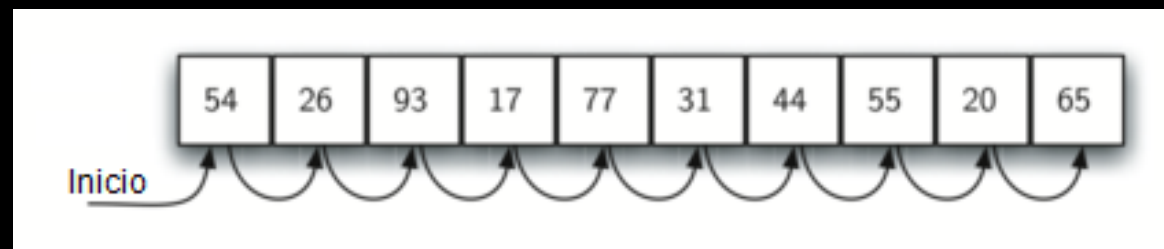




- **Busqueda de secuencias.** Cuando los ítems de datos se almacenan en una colección. Cada ítem de datos se almacena en una posición relativa a los demás. Dado que estos valores de los índices están ordenados, es posible para nosotros visitarlos en secuencia. Este proceso da lugar a nuestra primera técnica de búsqueda, la **búsqueda secuencial**.

Ejemplo

Comenzando en el primer ítem de la lista, simplemente nos trasladamos de un ítem a otro, siguiendo el orden secuencial subyacente hasta que encontremos lo que buscamos o nos quedemos sin ítems. Si nos quedamos sin ítems, hemos descubierto que el ítem que estábamos buscando no estaba presente.





Ejemplo

```
def busquedaSecuencial(unaLista, item):  
    pos = 0  
    encontrado = False  
    while pos < len(unaLista) and not encontrado:  
        if unaLista[pos] == item:  
            encontrado = True  
        else:  
            pos = pos+1  
    return encontrado
```

```
listaPrueba = [1, 2, 32, 8, 17, 19, 42, 13, 0]  
print(busquedaSecuencial(listaPrueba, 3))  
print(busquedaSecuencial(listaPrueba, 13))
```

```
False  
True
```

La función necesita la lista y el ítem que estamos buscando y devuelve un valor booleano que indica si el ítem está o no presente. La variable booleana encontrado se inicializa en False y se le asigna el valor True si descubrimos el ítem en la lista.





Algoritmos de extracción de patrones secuenciales

Algoritmos de extracción de patrones secuenciales

- **Método basado en Apriori:** GSP (Generalized Sequential Patterns)
- **Métodos de crecimiento de patrones:** FreeSpan & PrefixSpan
- **Minería basada en formatos verticales:** SPADE
- **Minería de patrones secuenciales basada en restricciones:** SPIRIT
- **Minería de patrones secuenciales cerrados:** CloSpan



Algoritmos de extracción de patrones secuenciales

Algoritmos de extracción de patrones secuenciales

- **Método basado en Apriori:** GSP (Generalized Sequential Patterns)
- **Métodos de crecimiento de patrones:** FreeSpan & PrefixSpan
- **Minería basada en formatos verticales:** SPADE
- **Minería de patrones secuenciales basada en restricciones:** SPIRIT
- **Minería de patrones secuenciales cerrados:** CloSpan

Ejemplo 2



CustomerID	CleanMonth	ServiceLevel	Revenue
1	5/1/16	C	\$3,000
1	5/1/16	D	\$4,000
1	10/1/16	A	\$1,000
1	10/1/16	B	\$2,000
1	10/1/16	C	\$3,000
1	3/1/17	A	\$1,000
1	3/1/17	B	\$2,000
1	3/1/17	F	\$6,000
1	8/1/17	A	\$1,000
1	8/1/17	C	\$3,000
1	8/1/17	D	\$4,000
1	8/1/17	F	\$6,000
2	10/1/16	A	\$1,000
2	10/1/16	B	\$2,000
2	10/1/16	F	\$6,000
2	3/1/17	E	\$5,000
3	5/1/16	A	\$1,000
3	5/1/16	B	\$2,000
3	5/1/16	F	\$6,000
4	5/1/16	D	\$4,000
4	5/1/16	G	\$7,000
4	5/1/16	H	\$8,000
4	3/1/17	B	\$2,000
4	3/1/17	F	\$6,000
4	8/1/17	A	\$1,000
4	8/1/17	G	\$7,000
4	8/1/17	H	\$8,000



sequenceID	eventID	SIZE	items
1	10	2	C, D
1	15	3	A, B, C
1	20	3	A, B, F
1	25	4	A, C, D, F
2	15	3	A, B, F
2	20	1	E
3	10	3	A, B, F
4	10	3	D, G, H
4	20	2	B, F
4	25	3	A, G, H

Click to add text

El ejemplo buscamos encontrar patrones entre si un cliente hizo una compra con anterioridad, cual sera su siguiente compra.

Lo siguiente es correr el algoritmo SPADE.

DATABASE			FREQUENT SEQUENCES	
SID	Time (EID)	Items	Frequent 1-Sequences	
Azure Customer 1	1	10	C	4
	1	15	A B C	4
	1	20	A B F	2
	1	25	A C D F	4
Azure Customer 2	2	15	A B F	3
	2	20	E	3
Azure Customer 3	3	10	A B F	2
	3	10	A B F	4
Azure Customer 4	4	10	D G H	2
	4	20	B F	2
	4	25	A G H	2

Frequent 2-Sequences	
AB	3
AF	3
B→A	2
BF	4
D→A	2
D→B	2
D→F	2
F→A	2

Frequent 3-Sequences	
ABF	3
BF→A	2
D→BF	2
D→B→A	2
D→F→A	2

Frequent 4-Sequences	
D→BF→A	2

Single Azure services, e.g. A = "Compute", B = "Networking"

Interpretation: It is likely for service D to be purchased, and then for services B and F to be simultaneously purchased, and finally for A to be purchased

Azure Customer ID Time in Months Azure Services Purchased

Ejemplo 2

Ahora, con esa funcion spade, se separan por clientes y los items se separan por subsecuencias.

```

set of 18 sequences with
most frequent items:
  A      B      F      D (other)
  11     10     8      8      28

most frequent elements:
  {A}    {D}    {B}    {F}    {B,F} (other)
   8     8     4     4     4     3

element (sequence) size distribution:
sizes
1 2 3
8 7 3

sequence length distribution:
lengths
1 2 3 4
4 8 5 1

summary of quality measures:
support
Min. :0.5000
1st Qu.:0.5000
Median :0.5000
Mean :0.6528
3rd Qu.:0.7500
Max. :1.0000


includes transaction ID lists: FALSE

mining info:
  data ntransactions nsequences support
trans_matrix      10         4      0.3

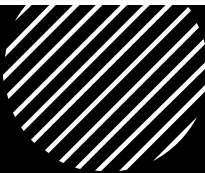
```

Despues con la funcion summary no deja ver, la frecuencia, los items mas repetids, etc.

Ejemplo 2



	rule	support	confidence	lift
1	$\langle\{D\}\rangle \Rightarrow \langle\{F\}\rangle$	0.5	1.0	1.0
2	$\langle\{D\}\rangle \Rightarrow \langle\{B,F\}\rangle$	0.5	1.0	1.0
3	$\langle\{D\}\rangle \Rightarrow \langle\{B\}\rangle$	0.5	1.0	1.0
4	$\langle\{B\}\rangle \Rightarrow \langle\{A\}\rangle$	0.5	0.5	0.5
5	$\langle\{D\}\rangle \Rightarrow \langle\{A\}\rangle$	0.5	1.0	1.0
6	$\langle\{F\}\rangle \Rightarrow \langle\{A\}\rangle$	0.5	0.5	0.5
7	$\langle\{D\},\{F\}\rangle \Rightarrow \langle\{A\}\rangle$	0.5	1.0	1.0
8	$\langle\{B,F\}\rangle \Rightarrow \langle\{A\}\rangle$	0.5	0.5	0.5
9	$\langle\{D\},\{B,F\}\rangle \Rightarrow \langle\{A\}\rangle$	0.5	1.0	1.0
10	$\langle\{D\},\{B\}\rangle \Rightarrow \langle\{A\}\rangle$	0.5	1.0	1.0



Ahora, en las siguientes columnas de las tablas de reglas, se observan que objetos son los que hacen que despues de comprado uno son los siguientes de comprar.

Esto se interpreta como, por ejemplo, despues de comprar un item D, se comprar un item F en un tiempo futuro por la probabilidad y con una certeza del 100%

Bibliografía

<https://apiumhub.com/es/tech-blog-barcelona/mineria-de-datos-casos-de-uso-beneficios/>

http://ri.uaemex.mx/bitstream/handle/20.500.11799/103630/secme-25335_1.pdf?sequence=1&isAllowed=y

http://web.cs.ucla.edu/~yzsun/classes/2018Fall_CS145/Slides/13Sequence_Data.pdf

<https://es.slideshare.net/marilynsilvana/mineria-de-datos-secuenciales-5571523>

<https://www.vldb.org/conf/1999/P22.pdf>

<https://runestone.academy/runestone/static/pythoned/SortSearch/LaBusquedaSecuencial.html>

Bibliografía

<https://philippe-fournierviger.com/spmf/SPADE.pdf>

<https://www.cs.sfu.ca/~jpei/publications/span.pdf>

<https://programmerclic.com/article/50051245985/>

<https://riull.ull.es/xmlui/bitstream/handle/915/1136/Inteligencia%20Empresarial%20basada%20en%20tecnicas%20de%20minería%20de%20datos%20y%20de%20procesos.%20Fusion%20Minería%20de%20procesos%20evolutiva%20y%20patrones%20secuenciales%20frecuentes.pdf?sequence=1&isAllowed=y>

<https://rdr.io/cran/arulesSequences/man/zaki.html>

<https://blog.revolutionanalytics.com/2019/02/sequential-pattern-mining-in-r.html>

Preguntas

- 1) Define la Técnica de Patrones Secuenciales
- 2) ¿Cuál es el objetivo?
- 3) Menciona una de su características
- 4) Menciona una de las dificultades de ésta Técnica
- 5) ¿Cuál es el método basado en Apriori?