



UANL

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM

FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS



...

PATRONES SECUENCIALES

MINERÍA DE DATOS

GRUPO: 003

...

INTEGRANTES:

Arlette Martinez Martinez

Leslie Paola Navarro Sosa

Nadia Valeria Prado Porras

Areli Fernanda Tolentino Vargas

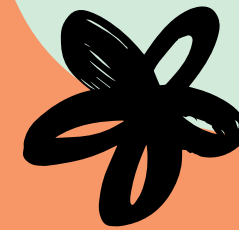
Dayana Guadalupe Reyes Joaquin

La minería de patrones secuenciales, es la búsqueda de patrones que ocurren frecuentemente relacionados con el tiempo o con otras secuencias.

...



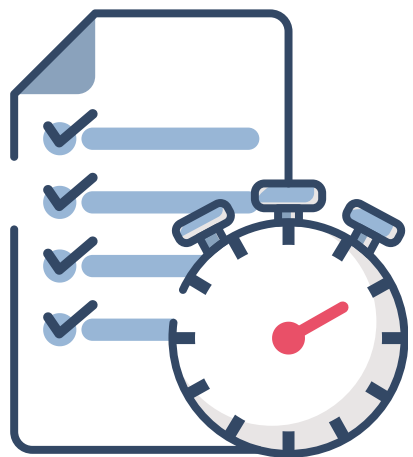
MINERÍA DE PATRONES SECUENCIALES





Patrones secuenciales

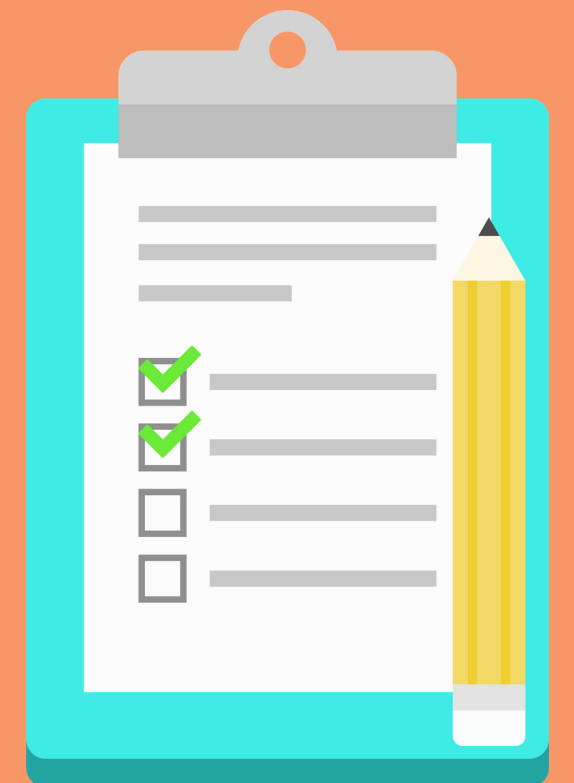
Es muy utilizado en la industria de ventas al por menor, y también en el dominio de la medicina. El resultado de ésta técnica se presenta como una lista de transacciones.



Los algoritmos de patrones secuenciales son muy difíciles a la hora de descubrir la tendencia de los datos como: El número de revistas deportivas vendidas a clientes con N° de crédito

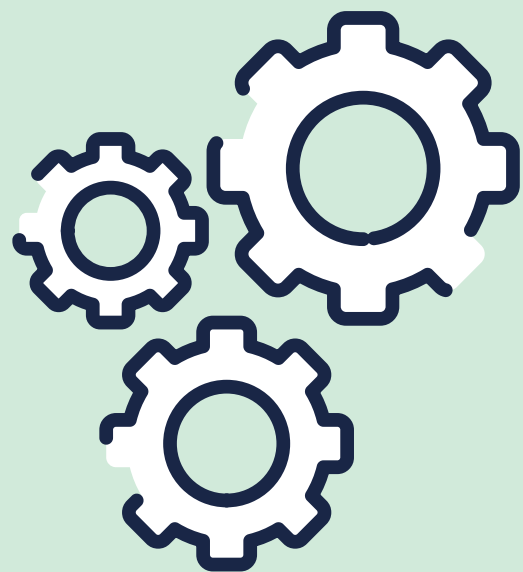
... CARACTERÍSTICAS

- El orden importa.
- Objetivo: encontrar patrones secuenciales.
- El tamaño de una secuencia es su cantidad de elementos.
- La longitud de la secuencia es la cantidad de items.
- El soporte de una secuencia es el porcentaje de secuencias que la contienen en un conjunto de secuencias S.
- Las secuencias frecuentes son las subsecuencias de una secuencia que tiene soporte mínimo.



VENTAJAS

FLEXIBILIDAD



EFICIENCIA

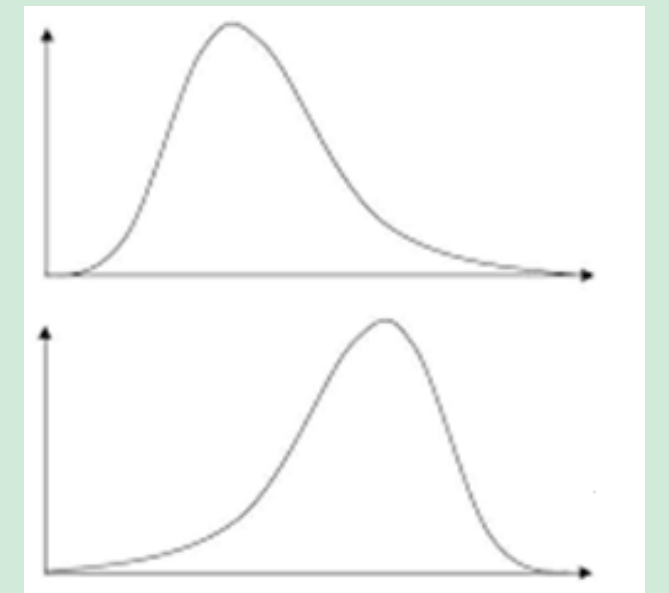


DESVENTAJAS

UTILIZACIÓN



**SESGADO POR
LOS PRIMEROS
PATRONES**



Tipos de datos

✿ ADN y Proteínas.



...

✿ Recorrido de clientes en un supermercado.



✿ Registros de accesos a una página web



...

APLICACIONES

MÚSICA



Una melodía puede verse como una secuencia de elementos musicales.

ANÁLISIS DE MERCADO




Comportamiento de compras.

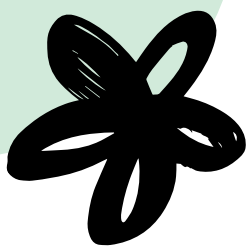
TELECOMUNICACIONES



Búsqueda de patrones de llamadas telefónicas y para servicios basados en la localización.



Existen varios algoritmos para resolver problemas de patrones secuenciales. Algunos son:



- GSP [Patrones secuenciales generalizados].
- SPADE [Descubrimiento de patrón secuencial usando clase equivalente].
- PrefixSpan

Secuencia / Subsecuencia



Subsecuencia

Una subsecuencia es una secuencia dentro de otra secuencia.

Secuencia $\langle a_1 a_2 \dots a_n \rangle$ esta contenida en otra secuencia $\langle b_1 b_2 \dots b_m \rangle$

- Entonces m tiene que ser mayor o igual a n.
- El item del evento i de la subsecuencia tiene que estar dentro del evento i de la secuencia

Secuencia



Una secuencia es una lista ordenada de elementos:

$$S = \langle e_1 e_2 e_3 \dots \rangle$$

La longitud de la secuencia ($|S|$) es el número de elementos de la secuencia.

Una k-secuencia es una secuencia de k eventos.

- Cada elemento contiene una colección de eventos:

$$e_i = \{i_1, i_2, \dots, i_k\}$$

Secuencia	Subsecuencia	¿incluida?
$\langle \{2,4\} \{3,5,6\} \{8\} \rangle$	$\langle \{2\} \{3,5\} \rangle$	Sí
$\langle \{1,2\} \{3,4\} \rangle$	$\langle \{1\} \{2\} \rangle$	No
$\langle \{2,4\} \{2,4\} \{2,5\} \rangle$	$\langle \{2\} \{4\} \rangle$	Sí

ANÁLISIS DE SECUENCIAS

BASE DE DATOS



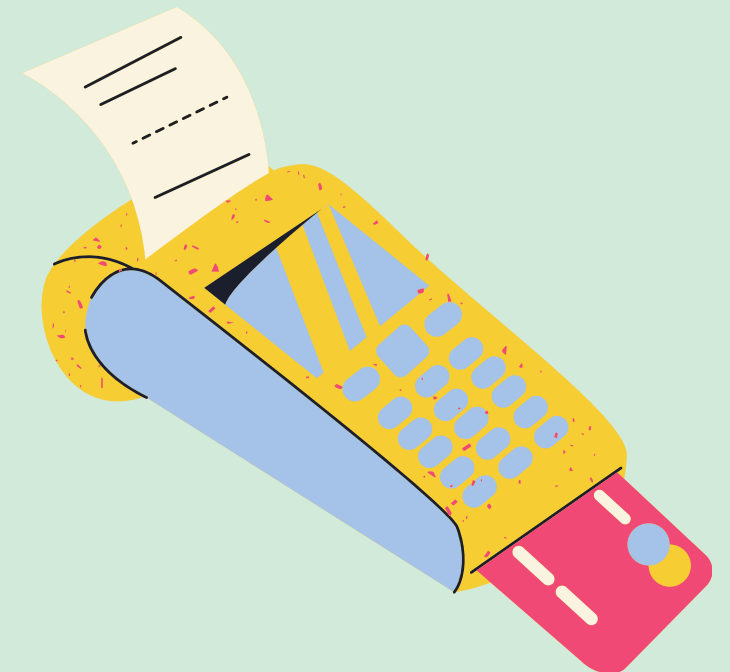
SECUENCIA



**ELEMENTO
(TRANSACCIÓN)**



EVENTO (ITEM)



MÉTODO GSP

...

1

En la primera pasada producimos todas las secuencias frecuentes de 1 elemento

2

Repetir hasta encontrar nuevas secuencias frecuentes

2.1

Generación de candidatos: Fusionar los pares encontrados en el paso $k-1$. $W1$ y $W2$ se pueden combinar si las subsecuencias obtenidas por la eliminación del primer elemento de $W1$ y el último elemento de $W2$ son iguales

2.2

Eliminación de candidatos: Elimine los candidatos que contienen una subsecuencia que es infrecuente en $k-1$ subsecuencias

2.3

Soporte de conteo: necesidad de nuevo pase a la base de datos

2.4

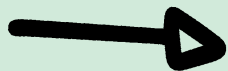
Eliminación de candidatos: implica umbrales basados en minsup

Ejemplo

SECUENCIAS

	A	B	C
1	num	elementos	
2		1 a,d,c,b,a,c	
3		2 b,f,c,e,b,f,g	
4		3 a,h,b,f,a,b,f	
5		4 b,e,c,e,d	
6		5 a,b,d,b,c,b,a,d,e	
7			

RESPUESTA



GENERACIÓN DE CANDIDÁTOS

```
L1:
defaultdict(<class 'int'>, {'b': 5, 'c': 4, 'd': 3, 'a': 3,
'f': 2, 'e': 3})
L2:
defaultdict(<class 'int'>, {'bb': 3, 'bc': 4, 'bd': 2, 'ba':
3, 'bf': 2, 'be': 3, 'cb': 3, 'cd': 2, 'ca': 2, 'ce': 3, 'db'
: 2, 'dc': 2, 'da': 2, 'ab': 3, 'ac': 2, 'ad': 2, 'aa': 3, 'f
b': 2, 'ff': 2})
L3:
defaultdict(<class 'int'>, {'bbf': 2, 'bcb': 2, 'bcd': 2, 'bc
e': 3, 'bfb': 2, 'bff': 2, 'cba': 2, 'dbc': 2, 'dba': 2, 'dcb'
': 2, 'dca': 2, 'abb': 2, 'abc': 2, 'aba': 3, 'acb': 2, 'aca'
: 2, 'adb': 2, 'adc': 2, 'ada': 2, 'fbf': 2})
L4:
defaultdict(<class 'int'>, {'bfbf': 2, 'dcba': 2, 'acba': 2,
'adbc': 2, 'adba': 2, 'adcb': 2, 'adca': 2})
L5:
defaultdict(<class 'int'>, {'adcba': 2})
```

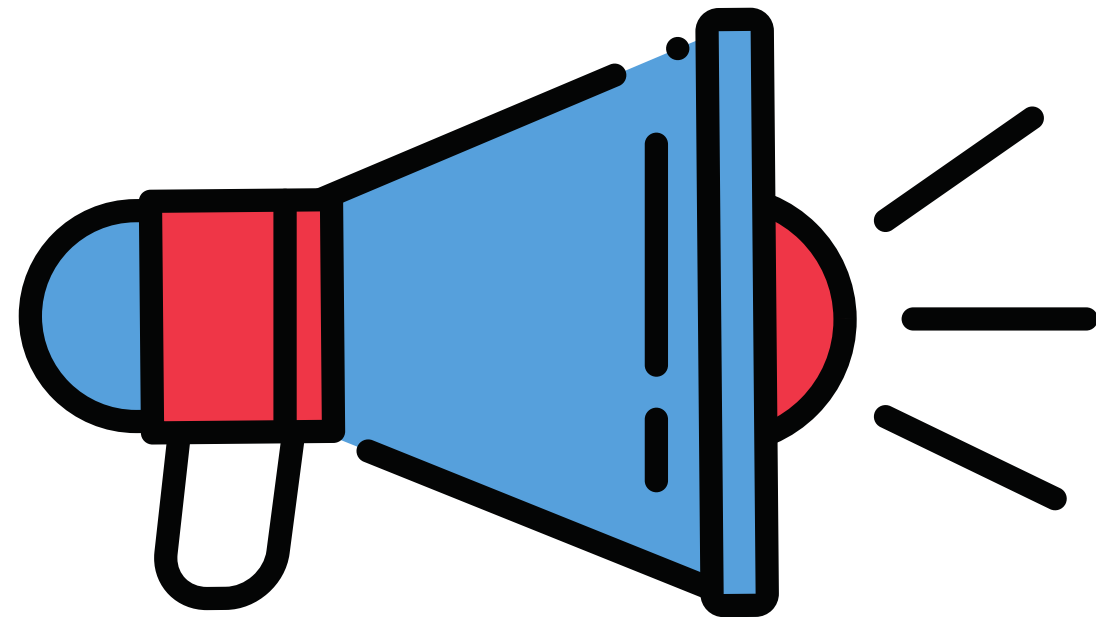
Bibliografía

...

Jaramillo, Marilyn. "Minería de datos secuenciales". Slideshare.com. (Octubre, 2010).
<https://es.slideshare.net/marilynsilvana/mineria-de-datos-secuenciales-5571523>

"Python: algoritmo GSP en minería de datos". Programador Clic.
<https://programmerclick.com/article/2629897547/>

Berzal, Fernando. "Patrones Secuenciales". Universidad de Granada. (Febrero, 2018).
<https://elvex.ugr.es/idbis/dm/slides/22%20Pattern%20Mining%20-%20Sequences.pdf>



**¡Gracias por
su Atención!**

