UANL
UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN

FCFM
FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS

PATRONES SECUENCIALES

Minería de datos/003 *Mayra Cristina Berrones Reyes*

INTEGRANTES

Marlene Calderon Rangel 1811330
Perla Millet Díaz Talamantes 1809285
Leslye Marisol Hernandez Bolaños 1819111
Valeria Esthepania Urbina Gallegos 1799959
Ulises Solis Moises 1887850

Conceptos

Minería de Datos Secuenciales: Es la extracción de patrones frecuentes relacionados con el tiempo u otro tipo de secuencia. Son eventos que se enlazan con el paso del tiempo
El orden de acontecimientos es considerado.

Se busca asociaciones de la forma "si sucede de la forma X en el instante de tiempo t entonces sucederá en el evento Y en el instante *t+n*". El objetivo es poder describir de forma concisa relaciones temporales que existen entre los valores de los atributos del conjunto de ejemplos.

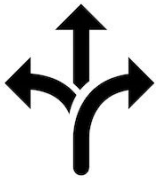
Reglas de asociación secuencial: Expresan patrones secuenciales, esto quiere decir que se dan en instantes distintos en el tiempo.

Características

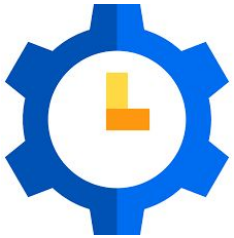
- El orden importa.
- Objetivo: encontrar patrones secuenciales.
- El tamaño de una secuencia es su cantidad de elementos.
- La longitud de la secuencia es la cantidad de ítems.
- El soporte de una secuencia es el porcentaje de secuencias que la contienen en un conjunto de secuencias S.
- Las secuencias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

VENTAJAS

Flexibilidad



Eficiencia



DESVENTAJAS

Utilización

Sesgado por los primeros patrones

Tipos de datos

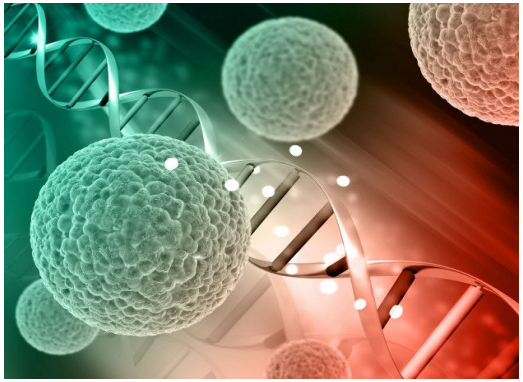
ADN y Proteínas

Recorrido de clientes en un supermercado

Registros de accesos a una página web

Aplicaciones

Medicina



Predecir si un compuesto químico causa cáncer

Análisis de Mercado

Comportamiento de compras

Web

Reconocimiento de spam de un correo electrónico

Agrupamiento de patrones secuenciales

Clasificación con datos secuenciales

El Proceso de los patrones secuenciales

Secuencias

$s = < e_1\ e_2\ e_3\ e_4 \dots >$

$|s|$ es el número de elementos en una secuencia

$e_i = \{ i_1, i_2, i_3, \dots, i_k \}$

Una k-secuencia es una secuencia con k eventos

Secuencia de visitas en una página web

< {Homepage} {Electronics} {Tablets} {Kindle Fire HD} {Shopping Cart} {Order Confirmation} {Return to Shopping} >

Subsecuencias

Una subsecuencia es una secuencia que está dentro de otra. Pero cumpliendo ciertas normas.

Secuencia	Subsecuencia	¿incluida?
< {2,4} {3,5,6} {8} >	< {2} {3,5} >	Sí
< {1,2} {3,4} >	< {1} {2} >	No
< {2,4} {2,4} {2,5} >	< {2} {4} >	Sí

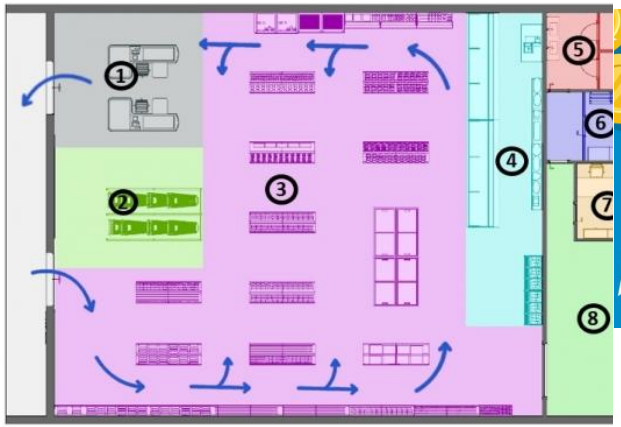
El ítem del evento i de la subsecuencia, tiene que estar dentro del evento i de la secuencia

Análisis de secuencias

Base de datos

Secuencia

Elemento (Transacción)



Elemento (Ítem)

Fases del método GSP (Generalized Sequential Pattern)

Fase 1:

Recorrer la base de datos para obtener todas las secuencias frecuentes de 1 elemento.

Fase 2:

Generación:

Generar k-secuencias candidatas a partir de las (k-1)-secuencias frecuentes.

Poda:

Podar k-secuencias candidatas que contengan alguna (k-1)-secuencia no frecuente.

Conteo:

Obtener el soporte de las candidatas

Eliminación:

Eliminar las k-secuencias candidatas cuyo soporte real esté por debajo del Umbral de soporte mínimo de frecuencia

Generación de candidatos

Para k = 2

La combinación de dos 1-secuencias $\langle\{i_1\}\rangle$ y $\langle\{i_2\}\rangle$ produce 2-secuencias candidatas

$$\langle\{i_1\}\{i_2\}\rangle \;\&\; \langle\{i_1\ i_2\}\rangle$$

Para k > 2

Dos (k-1)-secuencias frecuentes, w_1 y w_2, se combinan para producir una k-secuencia candidata si la subsecuencia obtenida al eliminar el primer evento de w_1 es la misma que la subsecuencia obtenida al eliminar el último evento de w_2 .

La k-secuencia candidata vendrá dada por la extensión de w_1 .

Si los dos últimos eventos de w_2 corresponden al mismo elemento, el último evento de w_2 pasa a formar parte del último elemento de w_1 .

Ejemplos de candidatos

$w_1 = \langle \{1\}\ \{2\ 3\}\ \{4\} \rangle$ $w_2 = \langle \{2\ 3\}\ \{4\ 5\} \rangle$

Producen $\langle \{1\}\ \{2\ 3\}\ \{4\ 5\} \rangle$

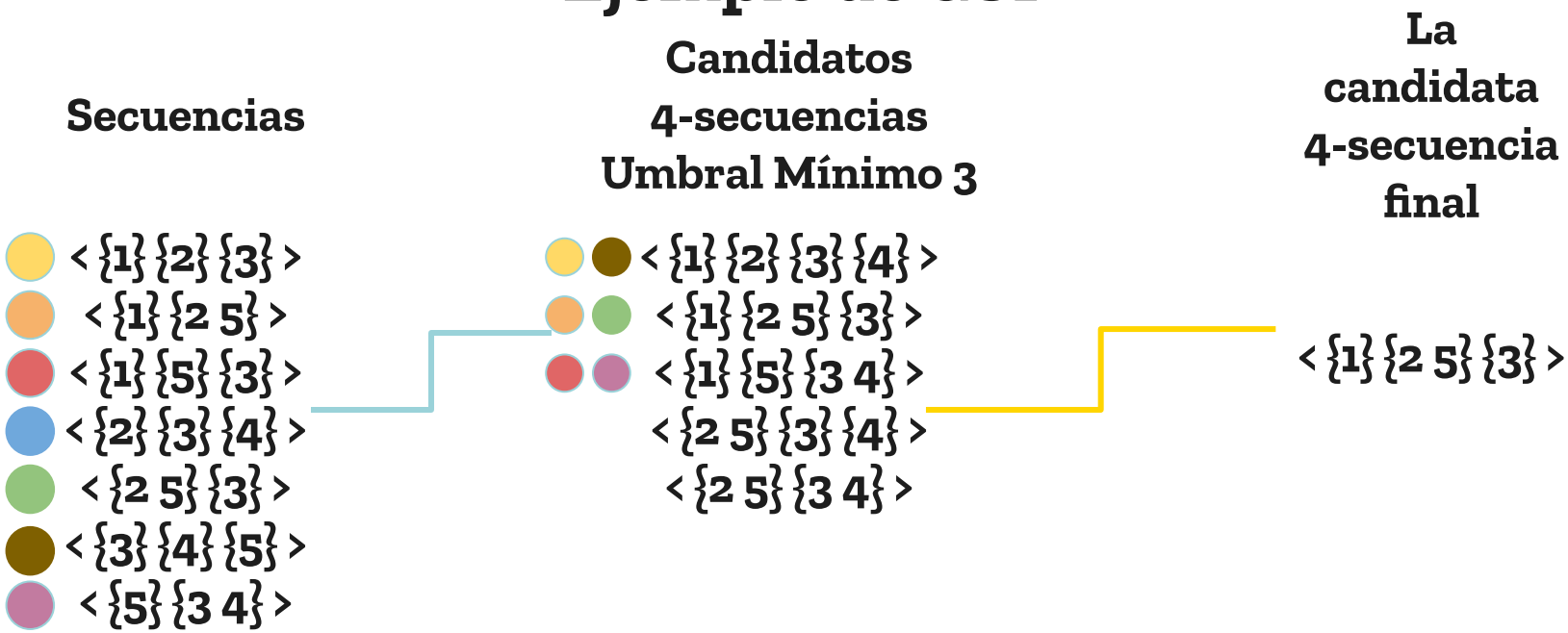
$w_1 = \langle \{1\}\ \{2\ 3\}\ \{4\} \rangle$ $w_2 = \langle \{2\ 3\}\ \{4\}\ \{5\} \rangle$

Producen $\langle \{1\}\ \{2\ 3\}\ \{4\}\ \{5\} \rangle$

$w_1 = \langle \{1\}\ \{2\ 6\}\ \{4\} \rangle$ $w_2 = \langle \{1\}\ \{2\}\ \{4\ 5\} \rangle$

Producen $\langle \{1\}\ \{2\ 6\}\ \{4\ 5\} \rangle$

Ejemplo de GSP



Bibliografía

Jaramillo, Marilyn. (2010, Octubre).

Minería de datos secuenciales. Slideshare.com .

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

Berzal, Fernando. (2018, Febrero).

Patrones secuenciales. Decsai.ugr.es;

Universidad de Granada.

https://elvex.ugr.es/idbis/dm/slides/22%20Pattern%20Mining%20-%20Sequences.pdf

Bibliografía

Agrawal, R., Imielinski, T., and Swami, A.

Mining association rules between sets of items il large Databases. In Proceedings of the 1993 ACM SIGMOD. International Conference on Management of Data (New York, NY, USA, 1993), SIGMOD '93, ACM, pp. 207{216.

Tan, P.-N., Steinbach, M., and Kumar, V.

Introduction to Data Mining. Addison-Wesley, 2006.