

FP-tree algorithm

FP tree:

- تستخدم لتميز انماط متكررة في منطقه تنقيب البيانات وهي إحدى خوارزميات association .
- خوارزميه FP tree خوارزميه بديله تستخدم لإيجاد العناصر المتكررة .
- تعتبر بشكل واسع مختلفة عن خوارزميه Apriori .

وتنقسم الى جزئين رئيسيين :

- اولاً : التعامل مع تمثيل FP tree .
- ثانياً : التفاصيل عن العناصر المتكررة بإستخدام هذه الشجرة وخوارزميتها .

خطوه ١ : نحسب الدعم الادنى :

أولا نحسب إحصاء الدعم الادنى . في السؤال الدعم الادنى يجب أن يكون ٣٠% نحسب كالتالي :

$$\text{إحصاء الدعم الادنى} = (30/100 * 8) = 2.4$$

تظهر لتمنح حساب اسهل يمكن أن يدور الحد الاعلى ، الآن

$$\text{الحد الاعلى هو عدد الدعم الادنى} = (30/100 * 8) = 3$$

خطوه ٢ : إيجاد تكرار الظهور أو الحدث :

إيجاد تكرار ظهور كل عنصر في جدول DB . كمثال عنصر A ظهر في الصف ١ او ٢ و ٣ و ٤ و ٧ ظهرت كليا ٥ مرات في جدول DB يمكن أن ترى عدد تكرار الظهور لكل عنصر في جدول ٢ .

خطوه ٣ : أولويه العناصر :

في الجدول ٢ يمكن أن ترى الاعداد المكتوبة بالقلم الاحمر . تلك أولويه كل عنصر طبقاً لتكرار ظهورها .
العنصر B حصل على الأولوية الاعلى بسبب أنه أعلى عدد في الظهور .
في نفس الوقت لديك فرصه لأسقاط العناصر التي لا تنجز متطلبات الدعم الأدنى . على سبيل المثال قاعده البيانات تحتوي على F التي تكرارها ١ حيث يتم اسقاطه .

Item	Frequency
A	5
B	6
C	3
D	6
E	4

Table2 -Frequency of Occurrence

Construct FP-tree from a Transaction Database

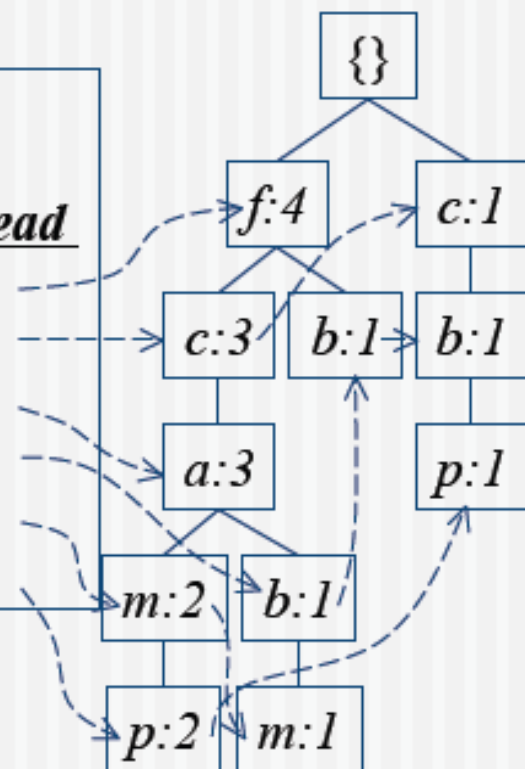
<i>TID</i>	<i>Items bought</i>	<i>(ordered) frequent items</i>
100	{f, a, c, d, g, i, m, p}	{f, c, a, m, p}
200	{a, b, c, f, l, m, o}	{f, c, a, b, m}
300	{b, f, h, j, o, w}	{f, b}
400	{b, c, k, s, p}	{c, b, p}
500	{a, f, c, e, l, p, m, n}	{f, c, a, m, p}

min_support = 3

1. Scan DB once, find frequent 1-itemset (single item pattern)
2. Sort frequent items in frequency descending order, f-list
3. Scan DB again, construct FP-tree

Header Table	
<i>Item frequency head</i>	
<i>f</i>	4
<i>c</i>	4
<i>a</i>	3
<i>b</i>	3
<i>m</i>	3
<i>p</i>	3

F-list=f-c-a-b-m-p



تقسيم الانماط و DB :

الانماط المتكررة يمكن أن تقسم الى مجموعات فرعية طبقاً الى f-list

$$\mathbf{F-list = f - c - b - m - p}$$

- الانماط التي تحتوي p .
- الانماط التي لديها m ولكن ليس p .
- .
- .
- الانماط التي لديها c ولكن ليس a ولا p,m,b .
- نمط f .

بحث الانماط التي لديها p من قاعده البيانات الشرطية :

- نبدئ بالعناصر المتكررة من راس الجدول للشجرة FP .
- عارض شجره FP بواسطة اتباع مسار كل العنصر مكرر p .
- جمع كل تحول للمسارات البادئة للعنصر P لتشكيل $P's$ حسب قاعده النمط الشرطية .

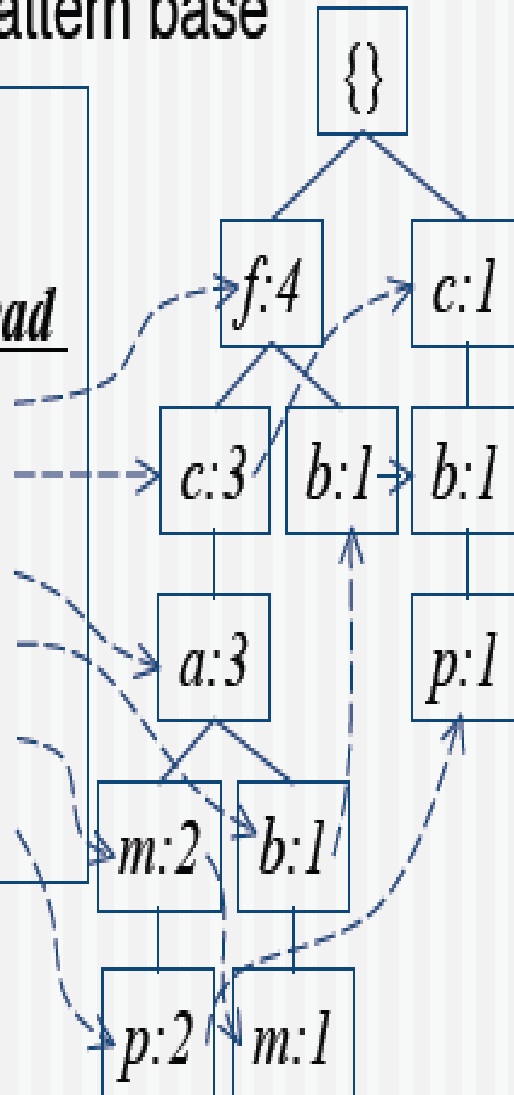
Finding all patterns with 'p'

conditional pattern base

Header Table

Item frequency head

<i>f</i>	4
<i>c</i>	4
<i>a</i>	3
<i>b</i>	3
<i>m</i>	3
<i>p</i>	3



Conditional pattern bases

item cond. pattern base

c *f:3*

a *fc:3*

b *fca:1, f:1, c:1*

m *fca:2, fcab:1*

p *fcam:2, cb:1*

العقدة p نمطه المتكرر (p:3) وله مساران في الشجرة p:1 , b:1 , c:3 و
f:4 , c:3 , a:3 , m:2

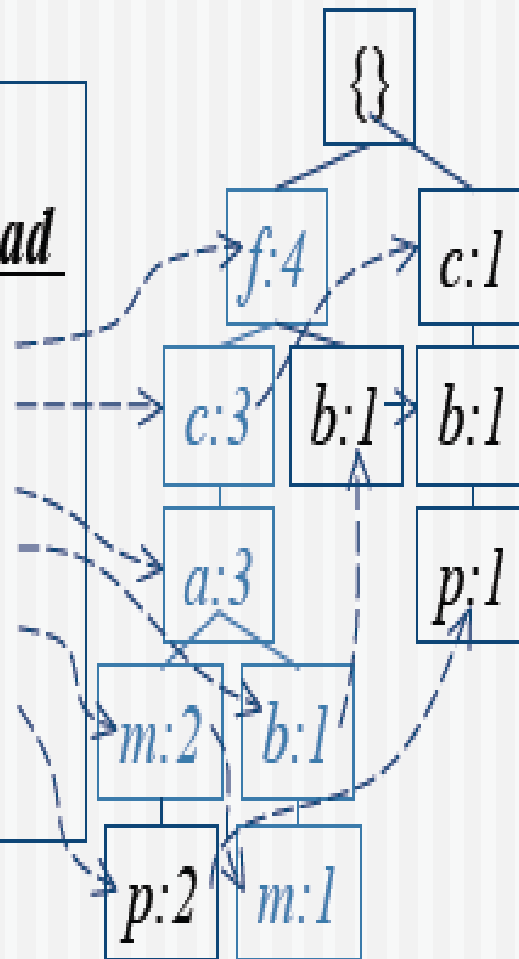
يشير المسار الاول الى الخط (f , c , a , m , p) يظهر مرتين في قاعده
البيانات .

لاحظ المسار يشير الى الخط (f , c , a) تظهر ثلاث مرات و f نفسه يظهر
اربع مرات ، على أي حال تظهر مرتين سوية مع p فيصبح f:2 , c:2 , a:2
m:2 , ، المسار الثاني يشير الى الخط (c , b , p) يظهر مره في مجموعه
المعاملات او p مسار بادئه cb:1 .

شكل قاعده النمط الشرطية بناء شجره FP على هذه القاعدة تؤدي الى نمط
متكرر واحد cp:3

Finding all patterns with 'm'

Header Table	
Item	frequency head
<i>f</i>	4
<i>c</i>	4
<i>a</i>	3
<i>b</i>	3
<i>m</i>	3
<i>p</i>	3



m-conditional pattern base:

fca:2, fcab:1



{

f:3

c:3

a:3

All frequent patterns relate to *m*

m,

fm, cm, am,

fcm, fam, cam,

fcam

-> associations

m-conditional FP-tree

العقدة m نمطه المتكرر $(m:3)$ وله مساران $a:3, c:3, f:4$ و $a:3, c:3, f:4$
 $a:3, m:2$

ليس هناك حاجة لتضمين p هنا في التحليل لأنها انماط متكررة تتضمن p
قاعده النمط الشرطية $(fcab:1)$ و $(fca:2)$ بناء شجره FP الشرطية عليه ،
تشتق m الشجرة الشرطية FP $a:3, c:3, f:3$ مسار متكرر وحيد .
هذه الشجرة الشرطية FP نقبت بشكل تكراري $(f:3, c:3, a:3, |m)$

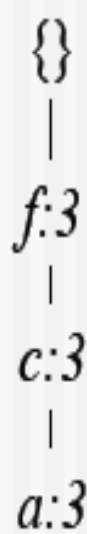
يتضمن تنقيب ثلاث عناصر (a),(c),(f)

يشتق أولاً من النمط المتكرر (am:3) قاعده النمط الشرطية (fc:3) وبعد ذلك تسمى (f:3,c:3 | am)

الثانية تشتق من تكرار النمط (cm:3) قاعده النمط الشرطية (f:3) وبعد ذلك تسمى (f:3 | cm)

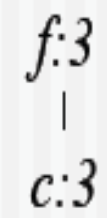
والثالث يشتق من تكرار النمط (fm:3) بشكل تكراري يسمى (, f:3)
(c:3 | am) يشتق نمطان (cam:3) و (fam:3) وقاعده نمط شرطيه
(f:3) وتسمى (f:3 | cam) الذي يشتق النمط الاطول (fcam:3) بنفس
الطريقة (f:3 | cm) يشتق نمط واحد (fcm:3) مجموعته الانماط
المتكررة { (m:3), (am:3), (cm:3), (fm:3), (cam:3), (fam:3),
(fcam:3), (fcm:3) }.

يشير بأن مسار وحيد شجرة FP يمكن أن ينقب بإنتاج كل مجموعات العناصر في المسار .



m-conditional FP-tree

Cond. pattern base of "am": (fc:3)



am-conditional FP-tree

Cond. pattern base of "cm": (f:3)



cm-conditional FP-tree

Cond. pattern base of "cam": (f:3)



cam-conditional FP-tree

Finding all patterns with 'b'

بنفس الطريقة عقده b يشتق (b:3) وله ثلاث مسارات :

f: 4, c: 3 , a: 3 , b: 1 , f: 4 , b: 1 , f: 1 , c: 1

من قاعده النمط الشرطية لـ b (fca : 1) ، (c : 1) ، (f : 1) لن يولد أي متكرر..

جدول Conditional FP-tree و Conditional pattern base

item	conditional pattern base	conditional FP-tree
p	{(fcam:2), (cb:1)}	{(c:3)} p
m	{(fca:2), (fcab:1)}	{(f:3, c:3, a:3)} m
b	{(fca:1), (f:1), (c:1)}	leer
a	{(fc:3)}	{(f:3, c:3)} a
c	{(f:3)}	{(f:3)} c
f	leer	leer

إعداد : صفاء قحمان

إشراف : أ / صلاح الحجري