第三章 出現摘要資訊儲存方法

如同前一章所述，目前交易視窗會隨時間移動，當目前時間點由 t 進到 t+1 時，若要精確地移除過時的資料項集，必須有交易 $T_{t\text{-}w+1}$ 的資料，才能精確將交易 $T_{t\text{-}w+1}$ 所包含的資料項集之支持度計數值扣除 1 次。如此一來，則需要隨時暫存目前交易視窗內各時間點對應的交易記錄。

為了不需將目前交易視窗內的交易原始資料儲存下來，本論文提出兩個記錄資料項集出現摘要資訊的方法，分別為平均時戳法與出現頻率改變點法。兩個方法各有自己的維護方法以及刪除過時資訊的規則，運用這兩種摘要資訊儲存法，可以較小儲存空間需求，近似找出目前交易視窗內的常見資料項集。

3-1 平均時戳法

資料流裡每一筆交易都有一個對應的時間點，此時間點為該交易所包含資料項集的一個出現時間點。平均時戳法是記錄資料項集的平均出現時間點，資料項集 p 的平均出現時間點是指：從開始記錄 p 到目前時間點為止，將所有出現時間點之總合，除以此期間 p 的支持度計數值所得之值。因此在出現摘要資料結構中對 p 記錄以下資訊：

(1) e：資料項集內容；

(2) t_s：開始時間點，即 p 開始加入此儲存結構的時間點；

(3) f：支持度計數值，即從 t_s 到目前時間點之間 p 的出現次數；

(4) sum：時戳總合，即從 t_s 到目前時間點之間 p 的出現時間之總合。

每當目前時間點 t 新輸入一筆交易 T_t，維護的步驟會先將新交易 T_t 中出現的資料項集資訊加入出現摘要結構中。由於探勘最近常見資料項集時，需估算資料項集在目前視窗中的支持度計數值，即必須去除資料項集在視窗外過時的計數值，因此第二個步驟便是從摘要資料結構中刪除過時資料項集。

我們以資料項集的平均時戳來代表其出現的分佈位置中心，資料項集 p 之平均時戳 $avg_t(p)$由其時戳總合 $p.sum$ 除以其支持度計數值 $p.f$ 計算得之。當一個資料項集 p 的平均出現時戳小於 $p.t_s + (CTL_t^{first} - p.t_s)/2$，表示其出現位置極可能落在目前交易視窗 CTL_t 之外，因此將 p 從摘要資料結構內移除。在目前時間 t 時，維護出現摘要資料結構的步驟如下：

步驟 1) 加入新輸入交易 T_t。

將 T_t 的所有子集找到對應於摘要資料結構中的資料項集位置，將其支持度計數值 f 加上 1、出現時戳總合 sum 加上 t。若該子集在摘要資料結構中沒有儲存對應的資料項集，則將此資料項集加入摘要資料結構，其支持度計數值設為 1，出現時戳總合及開始時間點皆設為 t。

當 $t \geq w$，表示目前交易視窗已有 w 筆交易後，便可繼續進行以下步驟。

步驟 2) 刪除過時的資料項集。

檢查整個摘要資料結構是否有資料項集 p，其開始時間點 $P.t_s$ 小於 CTL_t^{first}，且平均時戳小於 $p.t_s + (CTL_t^{first} - p.t_s)/2$，將這些資料項集從摘

要資料結構中刪除。

當有過時的資料項集被刪除時，依據 downward closure 之特性可知，其包含此資料項集的超集合（super Set）亦會過時，需一併刪除。

為避免摘要資料結構過大，我們只將常見及準常見資料項集保留下來。因此會定期或視需要情況，刪除摘要資料結構中最近估算支持度小於 ε 之資料項集。每個資料項集 p 的最近支持度計數值估算方式如下：當 $p.t_s$ 大於 CTL_t^{first} ，表示其有可能因非常見而被刪除過，因此其最近支持度計數值的最大值 $RC_t^{\max}(p)$ 由 $p.f + \lfloor (p.t_s - CTL_t^{first}) \times \varepsilon \rfloor$ 計算得之，而其最近支持度 $R\sup_t^{DS}(p)$ 便由 $RC_t^{\max}(p)$ 除以 $|CTL_t|$ 計算得之；當 $p.t_s$ 小於等於 CTL_t^{first} ，則表示 $p.f$ 的計數區間大於目前區間，但資料項集 p 還不是過時資料項集（在目前視窗中仍有出現），因此將 $p.f$ 除以其總計數區間 $(t - p.t_s + 1)$ 來估算其最近支持度 $R\sup_t^{DS}(p)$ 。

當需要探勘最近常見資料項集時，只要從找出摘要資料結構中找出最近支持度 $R\sup_t^{DS}(p)$ 大於等於 S_{min} 之資料項集 p 即可。

[範例 3.1]

以表 3.1 所示之資料流範例（一），每個時間點輸入一筆新交易處理，每 5 個時間點執行一次刪除非常見資料項集的步驟。若最小支持度門檻值 S_{min} 設為 0.5、最大支持度誤差值 ε 設為 0.25，且視窗大小 W 設為 10，則建立平均時戳摘要資訊的處理步驟說明如下。

表 3.1 資料流範例（一）

交易時間	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
交易所包含資料項	AB	AB	D	A	AB	AC	AE	AC	E	AE	AD	B	AE	B	AE

時間點：t_1

e	t_s	*f*	*sum*
A	1	1	1
B	1	1	1
AB	1	1	1

(a) 加入交易T_1
(CTL_1：T_1)

時間點：t_2

e	t_s	*f*	*sum*
A	1	2	3
B	1	2	3
AB	1	2	3

(b) 加入交易T_2
(CTL_2：T_1~T_2)

圖 3.1.1 平均時戳法之摘要資料結構（時間點 t_1~t_2）

在時間點 1 時，資料流輸入第一筆交易 T_1（AB），將 T_1 的所有子集{A, B, AB}之(e, t_s, f, sum)資訊加入摘要資料結構中，新增(A, 1, 1, 1)、(B, 1, 1, 1)及(AB, 1, 1, 1)三個資料項集摘要資訊，加入結果如圖 3.1.1(a)所示。接著資料流輸入第二筆交易 T_2（AB），將 T_2 的所有子集{A, B, AB}找出其在摘要資料結構中對應的資料

項集，將資料項集摘要資訊(A, 1, 1, 1)更新為(A, 1, 2, 3)，(B, 1, 1, 1)更新為(B, 1, 2, 3)，(AB, 1, 1, 1)更新為(AB, 1, 2, 3)，加入結果如圖 3.1.1(b)所示。

時間點：t_5

e	t_s	*f*	*sum*
A	1	4	12
B	1	3	8
AB	1	3	8
D	3	1	3

(a) 加入交易T_5
(CTL_5：T_1~T_5)

時間點：t_5

e	t_s	*f*	*sum*
A	1	4	12
B	1	3	8
AB	1	3	8

(b) 刪除非常見資料項集
(CTL_5：T_1~T_5)

圖 3.1.2 平均時戳法之摘要資料結構（時間點 t_5）

以此方式繼續處理，在時間點 5 時，加入交易 T_5 後之摘要資料結構如圖 3.1.2(a)所示。由於設定每五個時間點刪除非常見資料項集，因此從圖 3.1.2(a)所示之摘要資料結構中，計算並找到資料項集 D 的最近支持度為 0.2（即 1/5=0.2）小於 0.25，將 D 從摘要資料項集內刪除，刪除後結果如圖 3.1.2(b)所示。

時間點：t_{10}

e	t_s	*f*	*sum*
A	1	8	43
B	1	3	8
C	6	2	14
E	7	3	26
AB	1	3	8
AC	6	3	14
AE	7	3	17

(a) 加入交易T_{10}
(CTL_{10}：T_1~T_{10})

時間點：t_{11}

e	t_s	*f*	*sum*
A	1	9	54
B	1	3	8
C	6	2	14
D	11	1	11
E	7	3	26
AB	1	3	8
AC	6	3	14
AD	11	1	11
AE	7	3	17

(b) 加入交易T_{11}
(CTL_{11}：T_2~T_{11})

圖 3.1.3 平均時戳法之摘要資料結構（時間點 t_{10}~ t_{11}）

繼續處理到時間點 10 時，加入交易 T_{10}，此時為所設定需檢查非常見資料項集之時間點，但沒有找到非常見資料項集需被刪除，處理後結果如圖 3.1.3(a)所示。

在時間點 10（視窗大小 w）之後的每個時間點，皆要檢查摘要資料結構中是否有過時的資料項集。在時間點 11 時，加入交易 T_{11} 到摘要資料結構中，沒有過時的資料項集，處理結果如圖 3.1.3(b)所示。

時間點：t_{12}

e	t_s	*f*	*sum*
A	1	9	54
B	1	4	20
C	6	2	14
D	11	1	11
E	7	3	26
AB	1	3	8
AC	6	2	14
AD	11	1	11
AE	7	2	17

(a) 加入交易 T_{12}
(CTL_{12}：T_3~T_{12})

時間點：t_{13}

e	t_s	*f*	*sum*
A	1	10	67
B	1	4	20
C	6	2	14
D	11	1	11
E	7	4	39
AB	1	3	8
AC	6	2	14
AD	11	1	11
AE	7	3	30

(b) 加入交易 T_{13}
(CTL_{13}：T_4~T_{13})

圖 3.1.4 平均時戳法之摘要資料結構（時間點 t_{12}~t_{13}）

在時間點 12 及時間點 13 時，分別加入交易 T_{12} 及 T_{13} 的結果如圖 3.1.4(a)及圖 3.1.4(b)所示，兩個時間點皆沒有過時的資料集。

時間點：t_{14}

e	t_s	*f*	*sum*
A	1	10	67
B	1	5	34
C	6	2	14
D	11	1	11
E	7	4	39
AB	1	3	8
AC	6	2	14
AD	11	1	11
AE	7	3	30

(a) 加入交易T_{14}
(CTL_{14}：T_5~T_{14})

時間點：t_{14}

e	t_s	*f*	*sum*
A	1	10	67
B	1	5	34
C	6	2	14
D	11	1	11
E	7	4	39
AC	6	2	14
AD	11	1	11
AE	7	3	30

(b) 刪除過時資料項集
(CTL_{14}：T_5~T_{14})

圖 3.1.5 平均時戳法之摘要資料結構（時間點 t_{14}）

在時間點 14，加入交易 T_{14} 的摘要資料結構中後如圖 3.15(a)所示。此時 CTL_{14}^{first} 為 5，資料項集 AB 之平均時戳 avg_{14}(AB) 為 (3/8) 小於 3（即 $1+(5-1)/2=3$），表示為過時資料項集，因此將 AB 從摘要資料結構內刪除，刪除後結果如圖 3.1.5(b)所示。

時間點：t_{15}

e	t_s	*f*	*sum*
A	1	11	83
B	1	5	34
C	6	2	14
D	11	1	11
E	7	5	54
AC	6	2	14
AD	11	1	11
AE	7	4	45

(a) 加入交易T_{15}
(CTL_{15}：T_6~T_{15})

時間點：t_{15}

e	t_s	*f*	*sum*
A	1	11	83
B	1	5	34
E	7	5	54
AE	7	4	45

(b) 刪除非常見資料項集
(CTL_{15}：T_6~T_{15})

圖 3.1.6 平均時戳法之摘要資料結構（時間點 t_{15}）

在時間點 15 時，加入交易 T_{15} 後摘要資料結構中沒有過時的資料項集，處理結果如圖 3.1.6(a)所示。此時間點為 5 的倍數，因此檢查並刪除非常見資料項集 C、D、AC 及 AD 及其摘要資訊，刪除後結果如圖 3.1.6(b)所示。

若在時間點 15 時要探勘最近常見資料項集，則從圖 3.1.6(b)所示之摘要資料結構中，估算出資料項集 A 之最近支持度為 0.73（即 11/15=0.73）；資料項集 B 之最近支持度為 0.33(即 5/15=0.33)；資料項集 E 之最近支持度為 0.5(5/10=0.3)；資料項集 AE 之最近支持度為 0.4(4/10=0.4)，因此可得到常見資料項集 A 及 E。

3-2 出現頻率改變點法

如上一節所述，我們所建立的資料項集之出現摘要資訊中，對每個資料項集 p 所記錄計數值 $p.f$ 是從 p 加入摘要資料結構中的開始時間點 $p.t_s$，到目前時間點 t 所累計的出現次數。假設我們皆不會將 p 從摘要資料結構中刪除，則 $p.t_s$ 表示 p 在資料流中第一次的出現時間，因此可分析如以下兩種情況：

(1) $p.t_s \geq CTL_t^{first}$ ：則 $p.f$ 所記錄的是 p 在 CLT_t 中精確的支持度計數值，因此 $R\sup_t^{DS}(p)$ 可由 $(p.f/w)$ 精確的計算出來。

(2) $p.t_s < CTL_t^{first}$ ：在此情況下，$p.f$ 所記錄的計數值還包括 p 在 CTL_t 之前所累計的支持度計數值，因此有可能影響 p 在 CTL_t 中是否為常見資料項集的判斷。

我們進一步分析 p 在 $p.t_s$ 到 $(CTL_t^{first}-1)$ 之間，p 在 CTL_t 中，以及 p 在 $p.t_s$ 到目前時間 t 之間是否為常見資料項集的情況，如表 3.2 所示。

表 3.2 資料項集 p 在各區間出現狀況分析

情況	$p.t_s$ 到 $(CTL_t^{first}-1)$	CTL_t 中	$p.t_s$ 到目前時間 t
情況 1	p 為常見	p 為常見	p 為常見
情況 2	p 為常見	p 為非常見	p 可能為常見或非常見
情況 3	p 為非常見	p 為常見	p 可能為常見或非常見
情況 4	p 為非常見	p 為非常見	p 為非常見

由表 3.2 所示之分析情況，以 p 從 $p.t_s$ 到目前時間點 t 期間的出現頻率，來判斷 p 在 CTL_t 是否為最近常見資料項集時，可能發生誤判的情況為情況 2 或情況 3。在情況 2 中，錯誤發生的原因是 p 在 CTL_t 中的出現分佈變的稀疏，但其累計次數平均下來仍為常見，因此誤判為最近常見資料項集。在情況 3 中，則是 p 在 CTL_t 外（即從 $p.t_s$ 到 $CTL_t^{first}-1$）的出現次數很稀疏，造成從 $p.t_s$ 到 t 判斷為非常見資料項集。

由情形 2 跟 3 可知，資料項集出現頻率發生變化，會造成以 p 的累計出現次數判斷 p 是否是最近常見資料項集的誤判情形。因此，出現頻率改變點法所記錄的摘要資料結構，會記錄資料項集發生出現頻率可能轉變趨向不為常見的時間點，也就是此次出現和前一出現時間點的間隔大於$(1/S_{min})$之時間點，稱為一個出現頻率改變點。此外，並將此時間點之前到開始紀錄時間點或上一次頻率改變點間的出現次數記錄下來。當資料項集的開始計數時間 $p.t_s$ 已不落在目前交易視窗中，便可依所記錄頻率改變點重設 $p.t_s$，使其儘可能調整到最接近 CTL_t^{first}，並從 $p.f$ 中減去其記錄的過時出現次數，使近似探勘最近常見資料項集的結果更精準。出現頻率改變點法所用的摘要資料結構對每個資料項集記錄以下資訊：

(1) e：資料項集內容；

(2) t_s：開始時間，即此結構中目前記數 p 之支持度計數值的開始時間點；

(3) f：支持度計數值，即從 t_s 到目前時間點間 p 之出現次數；

(4) t_e：最後時間，即 p 最近一次出現的時間點；

(5) C_d：累計過時計數值，即 p 從 t_s 到最近一次頻率改變點之間的出現次數（不含最近一次改變點）；

(6) *Rqueue*：由頻率改變點對(t_r, C_r)組成的佇列，儲存資料項集 p 的頻率改變點 t_r，及 p 的上一次頻率改變點到 t_r 之前的出現次數 C_r（不含 t_r 這一次）。

和平均時戳法類似，每當目前時間點 t 新輸入一筆交易 T_t，會先執行新增的步驟，再執行去除過時資訊的調整。但在新增資料項集時必須判斷是否需記錄為其頻率改變點，而對過時資訊的調整則根據頻率改變點所記錄的資訊來分析處理。

步驟 1) 加入新輸入交易 T_t。

將 T_t 的所有子集 p 找到對應於摘要資料結構中的資料項集位置，將其支持度計數值 f 加上 1。並計算此資料項集與上次出現相隔$t - p.t_e$ 是否超過$(1/S_{min})$個時間點，若是則執行步驟 1.1。最後將 $p.t_e$ 更新為 t。

步驟 1.1) 記錄出現頻率改變點。

資料項集 p 在時間點 t 時出現一個頻率改變點，將頻率改變點 t_r 為 t，t_r 到上一個改變點間的計數值（不包含 t_r 這一次計數值）C_r 設為 $(p.f - 1) - p.C_d$，將頻率改變對(t_r, C_r)加入 $p.Rqueue$ 佇列中，最後將 $p.C_d$ 設為 $p.f - 1$。

如該子集在摘要資料結構中沒有儲存對應的資料項集，則將此資料項集

加入摘要資料結構，其支持度計數值 f 設為 1，開始時間點 t_s 及最後時間點 t_e 皆設為 t，C_d 設為 0，且 *Rqueue* 為空佇列。

當 $t \geq w$，表示目前交易視窗已有 w 筆交易後，便可繼續進行以下步驟。

步驟 2) 調整開始時間點。

檢查摘要結構中是否有資料項集 p，其開始時間點 $p.t_s$ 小於 CTL_t^{first}。對於符合此情況的資料項集 p，若 $p.Rqueue$ 為空，表示在其計數範圍中 p 一直為常見資料項集（間隔出現時間點皆小於 $1/S_{min}$），如同分析情況 1，不會導致對最近常見資料項集誤判的結果，所以不需調整。當 *Rqueue* 不為空佇列時，表示 p 有發生頻率改變點，則檢查第一個頻率改變點對 (t_r, C_r)，在下述三種情況下，此方法會調整開始時間點：

(1) 頻率改變點 t_r 小於等於 CTL_t^{first}：表示此 t_r 之前的出現次數 C_r 都落在 CTL_t 外，因此可調整 $p.t_s$ 由此頻率改變點 t_r 開始計數。

(2) 頻率改變點 t_r 大於 CTL_t^{first} 且 C_r 等於 1：表示 t_r 之前的出現次數只有開始時間點 $p.t_s$ 一次，且已知開始時間點 $p.t_s$ 小於 CTL_t^{first}，已不屬於 CTL_t，因此也可調整 $p.t_s$ 由此頻率改變點 t_r 開始計數。

(3) 頻率改變點 t_r 大於 CTL_t^{first} 且 C_r 不等於 1，但 t_r 前一次出現 p 的時間點 t_e' 小於 CTL_t^{first}：t_e' 的最晚時間點可由 t_r 往前估算：既然 t_r 為一頻率改變點，因此 $t_r - t_e' > 1/S_{\min}$，可推得 $t_e' < t_r - (1/S_{\min})$。亦可由

$p.t_s$ 往後推估算：由於 $p.t_s$ 到 t_r 前有出現 C_r 次，且其間沒有發生頻率改變點，因此每兩次出現的間隔 $\leq 1/S_{\min}$，可推得 $t_e' \leq p.t_s + (C_r - 1) \times (1/S_{\min})$，也就是 $t_e' < p.t_s + (C_r - 1) \times (1/S_{\min}) + 1$。綜合兩個估算值，$t_e'$ 的最晚時間點 t_e'_last = $\min(t_r - (1/S_{\min}), p.t_s + (C_r - 1) \times (1/S_{\min}) + 1) - 1$。當 t_e'_last 小於 CTL_t^{first}，表示 t_r 之前上一次出現 p 的時間點已不在 CTL_t 中，因此可調整 $p.t_s$ 由此頻率改變點 t_r 開始計數。

如符合上述三個調整的情況，則執行步驟 2.1，並將此頻率改變點對(t_r, C_r)從 *p.Rqueue* 中移除。否則直接執行步驟 3。

步驟 2.1) 調整開始時間點及支持度計數值。

調整資料項集 p 的開始時間點 $p.t_s$ 的方式是，將支持度計數值 $p.f$ 扣掉過時計數值 C_r，將開始時間 $p.t_s$ 設為此頻率改變點 t_r，並將累計過時計數值 $p.C_d$ 減掉 C_r。

步驟 3) 刪除過時資料項集。

從摘要資料結構中檢查是否有資料項集 p，其支持度計數值 $p.f$ 為 0 或最近一次出現的時間點 $p.t_e$ 小於 CTL_t^{first}，這些情況皆表示此資料項集沒有出現在目前交易視窗 CTL_t 中，因此將其資料項集從摘要資料結構中刪除。

不符合上述步驟 2 中三個調整的情況是當頻率改變點 t_r 大於 CTL_t^{first}，C_r 大於 1，且 t_r 前一次出現 p 的最晚時間點 t_e'_last 大於或等於 CTL_t^{first}。由於 $p.t_s$ 到 t_r 間沒有出現頻率改變點，表示從 $p.t_s$ 到 $(CTL_t^{first}-1)$ 間，p 為一個常見資料項集（間隔出現時間點皆小於 $1/S_{min}$）。因此即使在此狀況下沒有調整 $p.t_s$ 及 $p.f$。如果 p 在 CTL_t 中為常見資料項集（如同分析情況 1），可以保證其採用 $p.t_s$ 到 t 間的計數值 $p.f$，也會判斷 p 為最近常見資料項集，不會造成誤判。只有當 p 在 CTL_t 中為非常見資料項集時（如同分析情況 2），有可能被誤判為最近常見資料項集。因此採用此出現頻率改變點摘要資訊之調整方法，可以保證 CTL_t 中的最近常見資料項集可以完全被找出來。

同樣為了避免摘要資料結構過大，可以定期或視需要情況，刪除摘要資料結構中最近估算支持度小於 ε 之資料項集。每個資料項集 p 的最近支持度估算方式同平均時戳法：當資料項集 $p.t_s$ 大於 CTL_t^{first}，其最近支持度計數值的最大值 $RC_t^{\max}(p)$ 由 $p.f+\lfloor(p.t_s-CTL_t^{first})\times\varepsilon\rfloor$ 計算得之，而其最近支持度 $R\sup_t^{DS}(p)$ 便由 $RC_t^{\max}(p)$ 除以 $|CTL_t|$ 計算得之；否則其最近支持度 $R\sup_t^{DS}(p)$ 由 $p.f$ 除以 $t-p.t_s+1$ 計算得之。

當需要探勘最近常見資料項集時，只要從摘要資料結構中找出最近支持度 $R\sup_t^{DS}(p)$ 大於等於 S_{min} 之資料項集 p 即可。

[範例 3.2]

以表 3.1 所示之資料流範例（一），每個時間點輸入一筆新交易處理，每 5 個時間點執行一次刪除非常見資料項集的步驟。若最小支持度門檻值 S_{min} 設為 0.5、最大支持度誤差值 ε 設為 0.25，且視窗大小 w 設為 10，則建立出現頻率改變點摘要資訊的處理步驟說明如下：

時間點：t_1

e	f	t_s	t_e	C_d	$Rqueue$
A	1	1	1	0	{}
B	1	1	1	0	{}
AB	1	1	1	0	{}

(a) 加入交易T_1
(CTL_1：T_1)

時間點：t_2

e	f	t_s	t_e	C_d	$Rqueue$
A	2	1	2	0	{}
B	2	1	2	0	{}
AB	2	1	2	0	{}

(b) 加入交易T_2
(CTL_2：T_1~T_2)

圖 3.2.1 出現頻率改變點法之摘要資料結構（時間點 t_1~t_2）

在時間點 1 時，資料流輸入第一筆交易 T_1（AB），將 T_1 的所有子集{A, B, AB}之(e, t_s, t_e, f, $Rqueue$)資訊加入摘要資料結構中，新增(A, 1, 1, 1, 0, {})、(A, 1, 1, 1, 0, {})及(A, 1, 1, 1, 0, {})三個資料項集摘要資訊，加入結果如圖 3.2.1(a)所示。接著資料流輸入第二筆交易 T_2（AB），將 T_2 的所有子集{A, B, AB}找出其在摘要資料結構中對應的資料項集，皆沒有發生頻率改變點，因此資料項集摘要資訊(A, 1, 1, 1, 0, {})更新為(A, 2, 1, 2, 0, {})，(B, 1, 1, 1, 0, {})更新為(B, 2, 1, 2, 0, {})，(AB, 1, 1, 1, 0, {})更新為(AB, 2, 1, 2, 0, {})，更新結果如圖 3.2.1(b)所示。

時間點：t_3

e	*f*	t_s	t_e	C_d	*Rqueue*
A	2	1	2	0	{}
B	2	1	2	0	{}
AB	2	1	2	0	{}
D	1	3	3	0	{}

(a) 加入交易T_3
(CTL_3：T_1~T_3)

時間點：t_4

e	*f*	t_s	t_e	C_d	*Rqueue*
A	3	1	4	0	{}
B	2	1	2	0	{}
AB	2	1	2	0	{}
D	1	3	3	0	{}

(b) 加入交易T_4
(CTL_4：T_1~T_4)

圖 3.2.2 出現頻率改變點法之摘要資料結構（時間點 t_3~t_4）

在時間點 3 時，加入交易 T_3 後之摘要資料結構如圖 3.2.2(a)所示。在時間點 4 時，加入交易 T_4 後之摘要資料結構如圖 3.2.2(b)所示。

時間點：t_5

e	*f*	t_s	t_e	C_d	*Rqueue*
A	4	1	5	0	{}
B	3	1	5	2	{(5,2)}
AB	3	1	5	2	{(5,2)}
D	1	3	3	0	{}

(a) 加入交易T_5
(CTL_5：T_1~T_5)

時間點：t_5

e	*f*	t_s	t_e	C_d	*Rqueue*
A	4	1	5	0	{}
B	3	1	5	2	{(5,2)}
AB	3	1	5	2	{(5,2)}

(a) 刪除非常見資料項
(CTL_5：T_1~T_5)

圖 3.2.3 出現頻率改變點法之摘要資料結構（時間點 t_5）

在時間點 5 時，加入 T_5（AB），其中資料項集 AB 與上一次出現時間點 2 相隔 3 個時間點（即 5 – 2 = 3），大於 2(即 $1/S_{min}$=1/0.5=2)，因此將頻率改變點對(5,2)存進 AB.*Rqueue* 中，處理後結果如圖 3.2.3(a)所示。此時間點為 5 的倍數，因此刪除其最近支持度為 0.2（即 1/5=0.2）的資料項集 D，刪除後結果如圖 3.2.3(b)所示。

時間點：t_6

e	*f*	t_s	t_e	C_d	*Rqueue*
A	5	1	6	0	{}
B	3	1	5	2	{(5,2)}
C	1	6	6	0	{}
AB	3	1	5	2	{(5,2)}
AC	1	6	6	0	{}

(a) 加入交易T_6
(CTL_6：T_1~T_6)

時間點：t_7

e	*f*	t_s	t_e	C_d	*Rqueue*
A	6	1	7	0	{}
B	3	1	5	2	{(5,2)}
C	1	6	6	0	{}
E	1	7	7	0	{}
AB	3	1	5	2	{(5,2)}
AC	1	6	6	0	{}
AE	1	7	7	0	{}

(b) 加入交易T_7
(CTL_7：T_1~T_7)

時間點：t_8

e	*f*	t_s	t_e	C_d	*Rqueue*
A	7	1	8	0	{}
B	3	1	5	2	{(5,2)}
C	2	6	8	0	{}
E	1	7	7	0	{}
AB	3	1	5	2	{(5,2)}
AC	2	6	8	0	{}
AE	1	7	7	0	{}

(c) 加入交易T_8
(CTL_8：T_1~T_8)

時間點：t_9

e	*f*	t_s	t_e	C_d	*Rqueue*
A	7	1	8	0	{}
B	3	1	5	2	{(5,2)}
C	2	6	8	0	{}
E	2	7	9	0	{}
AB	3	1	5	2	{(5,2)}
AC	2	6	8	0	{}
AE	1	7	7	0	{}

(d) 加入交易T_9
(CTL_9：T_1~T_9)

圖 3.2.4 出現頻率改變點法之摘要資料結構（時間點 t_6~t_9）

從時間點 6 到時間點 9，加入該時間點新交易資料後之摘要資料結構分別如圖 3.2.4(a)到 3.2.4(d)所示。

時間點：t_{10}

e	*f*	t_s	t_e	C_d	*Rqueue*
A	8	1	10	0	{}
B	3	1	5	2	{(5,2)}
C	2	6	8	0	{}
E	3	7	10	0	{}
AB	3	1	5	2	{(5,2)}
AC	2	6	8	0	{}
AE	2	7	10	1	{(10,1)}

(a) 加入交易T_{10}
(CTL_{10}：T_1~T_{10})

時間點：t_{11}

e	*f*	t_s	t_e	C_d	*Rqueue*
A	9	1	11	0	{}
B	3	1	5	2	{(5,2)}
C	2	6	8	0	{}
D	1	11	11	0	{}
E	3	7	10	0	{}
AB	3	1	5	2	{(5,2)}
AC	2	6	8	0	{}
AD	1	11	11	0	{}
AE	2	7	10	1	{(10,1)}

(b) 加入交易T_{11}
(CTL_{11}：T_2~T_{11})

圖 3.2.5 出現頻率改變點法之摘要資料結構（時間點 t_{10}~t_{11}）

在時間點 10 時，加入交易 T_{10}（AE），其中資料項集 AE 與上一次出現時間點 7 相隔 3 個時間點大於 2，將頻率改變點對(10,1)存進 AE.*Rqueue* 中，處理結果如圖 3.2.5(a)所示。在時間點 10 後的每個時間點，皆要檢查是否有資料項集 $p.t_s$ 小於 CTL_t^{first} ，可以調整開始時間點及支持度計數值。在時間點 11 時，加入交易 T_{11} 後，處理後結果如圖 3.2.5(b)所示，沒有需要調整者。

時間點：t_{12}

e	*f*	t_s	t_e	C_d	*Rqueue*
A	9	1	11	0	{}
B	4	1	12	3	{(5,2)(12,1)}
C	2	6	8	0	{}
D	1	11	11	0	{}
E	3	7	10	0	{}
AB	3	1	5	2	{(5,2)}
AC	2	6	8	0	{}
AD	1	11	11	0	{}
AE	2	7	10	1	{(10,1)}

(a) 加入交易 T_{12}
(CTL_{12}：T_3~T_{12})

時間點：t_{12}

e	*f*	t_s	t_e	C_d	*Rqueue*
A	9	1	11	0	{}
B	2	5	12	1	{(12,1)}
C	2	6	8	0	{}
D	1	11	11	0	{}
E	3	7	10	0	{}
AB	1	5	5	0	{}
AC	2	6	8	0	{}
AD	1	11	11	0	{}
AE	2	7	10	1	{(10,1)}

(b) 調整開始時間點
(CTL_{12}：T_3~T_{12})

圖 3.2.6 出現頻率改變點法之摘要資料結構（時間點 t_{12}）

在時間點 12 時，加入交易 T_{12}（B），其中資料項 B 與上一次出現時間點 5 相隔 7 個時間點大於 2，將頻率改變點對(12,1)存進 B.*Rqueue* 中，處理後結果如圖 3.2.6(a)所示。接著檢查開始時間點小於 CTL_{12}^{first} =3 且 *Rqueue* 不為空佇列的資料項集 B 及 AB。從 B.*Rqueue* 中取出第一個改變對(5,2)，符合調整情形(3)，其 $\min(5-2,1+(2-1)\times 2+1)-1=2$ 小於 CTL_{12}^{first}，將 B 的開始時間點 B.t_s 設成 5 且其計數值 B.*f* 減 2；資料項集 AB 亦同 B 之情況，將其開始時間點 AB.t_s 設成 5，且其計數值 AB.*f* 減掉 2，調整結果如圖 3.2.6(b)所示。

時間點：t_{13}

e	f	t_s	t_e	C_d	*Rqueue*
A	10	1	13	0	{}
B	2	5	12	1	{(12,1)}
C	2	6	8	0	{}
D	1	11	11	0	{}
E	4	7	13	3	{(13,3)}
AB	1	5	5	0	{}
AC	2	6	8	0	{}
AD	1	11	11	0	{}
AE	3	7	13	2	{(10,1)(13,1)}

(a) 加入交易T_{13}
(CTL_{13}：T_4~T_{13})

時間點：t_{14}

e	f	t_s	t_e	C_d	*Rqueue*
A	10	1	13	0	{}
B	3	5	14	1	{(12,1)}
C	2	6	8	0	{}
D	1	11	11	0	{}
E	4	7	13	3	{(13,3)}
AB	1	5	5	0	{}
AC	2	6	8	0	{}
AD	1	11	11	0	{}
AE	3	7	13	2	{(10,1)(13,1)}

(b) 加入交易T_{14}
(CTL_{14}：T_5~T_{14})

圖 3.2.7 出現改變點法之摘要資料結構（時間點 t_{13}~t_{14}）

在時間點 13 時，加入 T_{13}（AE），其中資料項集 E 及 AE 各產生一個頻率改變點對(13,3)及(13,1)，處理後結果如圖 3.2.7(a)所示。在時間點 14 時，加入交易 T_{14}，處理後結果如圖 3.2.7(b)所示。

時間點：t_{15}

e	f	t_s	t_e	C_d	*Rqueue*
A	11	1	15	0	{}
B	3	5	14	1	{(12,1)}
C	2	6	8	0	{}
D	1	11	11	0	{}
E	5	7	15	3	{(13,3)}
AB	1	5	5	0	{}
AC	2	6	8	0	{}
AD	1	11	11	0	{}
AE	4	7	15	2	{(10,1)(13,1)}

(a) 加入交易T_{15}
(CTL_{15}：T_6~T_{15})

時間點：t_{15}

e	f	t_s	t_e	C_d	*Rqueue*
A	11	1	15	0	{}
B	2	12	14	0	{}
C	2	6	8	0	{}
D	1	11	11	0	{}
E	5	7	15	3	{(13,3)}
AB	1	5	5	0	{}
AC	2	6	8	0	{}
AE	4	7	15	2	{(10,1)(13,1)}

(b) 調整開始時間點
(CTL_{15}：T_6~T_{15})

時間點：t_{15}

e	f	t_s	t_e	C_d	*Rqueue*
A	11	1	15	0	{}
B	2	12	14	0	{}
C	2	6	8	0	{}
D	1	11	11	0	{}
E	5	7	15	3	{(13,3)}
AC	2	6	8	0	{}
AE	4	7	15	2	{(10,1)(13,1)}

(c) 刪除過時資料項集
(CTL_{15}：T_6~T_{15})

時間點：t_{15}

e	f	t_s	t_e	C_d	*Rqueue*
A	11	1	15	0	{}
B	2	12	14	0	{}
E	5	7	15	3	{(13,3)}
AE	4	7	15	2	{(10,1)(13,1)}

(d) 刪除非常見資料項集
(CTL_{15}：T_6~T_{15})

圖 3.2.8 出現頻率改變點法之摘要資料結構（時間點 t_{15}）

在時間點 15 時，加入交易 T_{15} 之結果如圖 3.2.8(a)所示。此時W_{12}^{first} 為 6，資料項集 B 的開始時間點小於 6 且可從 B.*Rqueue* 中取出改變對(12,1)，符合調整情形(2)，將開始時間點 B.t_s 調整為 12，且計數值 B.f 減 1，調整結果如圖 3.2.8(b)所示。接著執行步驟 3，刪除最後時間點 t_e 小於 6 的過時資料項集 AB，刪除後結果如圖 3.2.8(c)所示。此時間點為 5 的倍數，因此檢查並刪除非常見資料項集 C、D 及 AC，刪除後結果如圖 3.2.8(d)所示。

若在時間點 15 時要探勘最近常見資料項集，則從圖 3.2.8(d)所示之摘要資料結構中，可估算出資料項集 A 之最近支持度為 0.73（即 5/15=0.73）；資料項集 B 之最近支持度為 0.3（即 3/15=0.33）；資料項集 E 之最近支持度為 0.5（5/10=0.5）；資料項集 AE 之最近支持度為 0.4（4/10=0.4），因此可找出常見資料項集有 A 及 E。