

基於網路點擊序列預測交易決策之推薦機制

陳蓉慧

朝陽科技大學資訊管理系
e-mail: hueichen25@gmail.com.tw

薛夙珍

朝陽科技大學資訊管理系
e-mail: schsueh@cyut.edu.tw

摘要

電子商務網站蓬勃發展，網路上銷售的商品五花八門，該如何給予使用者們真正所需之商品和資訊，是現今網站經營者的重大挑戰。由於網站數量眾多而導致顧客的忠誠度不高，網路行銷管理者如何針對目標顧客作更精準的目標行銷(Target marketing)，將瀏覽者轉變成為忠誠的消費者，將有助於網站銷售的提升。本論文提出之以點擊序列(Click-sequence)資料進行分析的推薦機制，準確地預測顧客的交易決策，再分別進行不同的推薦策略。在論文中所提出的方法，對現有顧客或是新顧客都共同具備的網頁點擊資料作序列樣式分析，例如先點擊網頁 A 之後再點擊網頁 B 的循序行為。因此，論文利用已存在之點擊資料來做為訓練資訊，接著再瞭解哪些點擊路徑為有效購買路徑，並且依據已分析出之有效購買路徑來比對潛在新顧客的瀏覽路徑，作有效地商品資訊推薦。希望能依據點擊序列反映出顧客真實行為，掌握顧客動態的偏好特性，以達到推薦效益最佳化與提升顧客忠誠度。

關鍵詞：點擊序列、推薦機制、序列探勘、資料探勘、目標行銷。

Abstract

The proliferation of e-commerce website produces a variety of products and diversified information. The provision of products and information that fits users' need becomes a challenge for successful e-commerce. The customer loyalty is also decreased for the various websites providing similar products. Website administrators have difficulties of separating effective purchasing behaviors from the large amount of user-clicks so that successful target marketing by transforming website surfers to real customers cannot be easily created. In this paper, we propose a mechanism to mine the click-sequences, predict the purchasing decisions, and suggest corresponding recommendations to

increase customer transactions. By extracting effective purchasing click-sequences, we may add promotions in the potential clicking-paths to attract users for target marketing. The click-sequences can also be analyzed to obtain the preferences of customers so that recommendations can be optimized and customer loyalty can be raised.

Keywords: Click-sequence, Recommendation, Sequence mining, Data mining, Target marketing.

1. 前言

在於一個資訊爆炸的時代，以往的資訊不足已經不是目前大環境所面臨的問題，現在所面臨到的問題則是要如何從這些大型資源中發現並且挖掘出真正所需要的資訊，而資料探勘(Data mining)技術就是從大型的資料集中，分析、發現、擷取並且預測的一種技術，目前也被廣泛的使用在醫療、金融、電信、客戶關係管理及零售業，在零售業的應用當中，最耳熟能詳的為美國最大的零售連鎖店 Wal-Mart 的市場菜籃分析(Market basket analysis)，使用關聯法則(Association rules)[4,5]發現在禮拜五的晚上，啤酒和尿布的組合搭配這項規則，此規則就是前面所提到從大型資料集中發掘顧客所購買之商品間的關聯性，零售業者可針對所發掘出的資訊進行行銷上的應用以提升營業銷售，例如將尿布與啤酒進行搭配的促銷活動或者是更變商品櫃位放置顯眼的地方。

然而這樣的資訊應用在電子商務網站可以提供網站管理者了解顧客的購買行為，來設計適合得行銷活動之外，也可以透過商品間的關聯性來將網頁呈現的方式有所進行設計[8]，除了透過關聯法則發掘出關聯性外，也可透過網路點擊序列(Web click-sequence)來了解使用者於網站瀏覽的點擊順序為何，以便提供顧客瀏覽路徑的建議或改善網站的設計，使用網路點擊序列的分析當中所發掘出的不只關聯性並且還有順序性，而點擊序列若使用關聯法則(Association rules)發掘出的只有網頁間的關聯性卻忽略了其順序性，在本質上就是一個

有順序性的排序雖此問題增加了問題切割上的複雜度但卻也更精確的看出顧客於網站上的所有行為，然而，目前電子商務的競爭越演越烈，現今的網站管理者不能只考慮這兩項因素，在競爭白熱之際，該如何更有效的運用目前所擁有的資訊以利進行顧客關係管理(Customer relationship management, CRM)策略，以提升競爭優勢且增加客戶黏著度以利增加營業收益，然而目前電子商務網站多數都推出推薦系統(Recommender Systems)的服務，它不只推薦顧客潛在最有可能購買的商品之外也協助顧客進行交易上的決策，並且針對顧客進行個人化的推薦服務而不再是制式化的推薦方式導致推薦質量較差且不精準，而個人化(Personalized recommender systems)推薦系統則區分為內容導向過濾與協同過濾[6]，前者是基於顧客的偏好與購物資訊進而推薦符合顧客之事物，此方法有過於限定(over-specialization)的限制，因此技術是透過依照顧客的偏好檔案跟購物資訊進而推薦出相似的產品沒有變通性，也沒有考慮到顧客可能會隨著時間的改變而改變個人偏好，並且也對於沒有任何偏好資料及購物資料者無法進行推薦，而後者則是強調使用者間的相似程度，計算相似度後進行推薦，然而因為必須有購物資訊、個人資訊來或商品資訊進行相似度的計算，當資料稀疏(Sparse)時此方法無足夠的資料量來進行計算導致無法較精準計算出顧客間的相似程度。

在於上述所說明的兩種技術之外也在於過去的研究當中，大多數的研究方法都是使用購買資訊來進行分析並且提出不同的方法，並且對於新顧客於電子商務環境中的推薦機制尚無較合適的方法而選擇針對此客戶群進行制式化的推薦，要解決上述兩種侷限於購買資訊的分析及新客戶群的應對並不是現階段的技術無法達到，而是在過去的研究中忽略了顧客於電子商務環境中的行為樣式(Behavioral pattern)，例如顧客於網頁停留的時間、瀏覽點擊的路徑順序、瀏覽的數量、最後購買的決策及放入購物籃中的商品，然而上述所列舉之影響因素皆為一連串的連帶變動影響的因果關係，若就過去所是用之方法只由購買紀錄的資料來判定顧客偏好或特性進行推薦，會導致可納入分析的因素較於狹隘、獨斷具絕對性、無任何因果可追蹤並且分析此顧客的消費行為及特性、且被動性的獲取到最後的顧客購買資訊、資訊獲取緩慢以及顧客可能會因為隨著時間而改變行為樣式，而上述的種種因素可能是

經常造成推薦上的不準確或是選擇使用制式化推薦的原因，而會造成此現象的發生原因在於過去的推薦系統中都並無將完整的顧客行為樣式及變動因素納入考量而錯失資訊以致無法立即的調整推薦機制，因此若能即時並且掌握完整的顧客資訊從顧客於電子商務的環境當中一開始就持續並且快速的掌握特性資料且考慮顧客的變動性，取而代之於顧客結束整個交易行為後才掌握購買資料，那便能改善新顧客群於推薦機制上的應用，更能立即而準確並且根據顧客特性之變動進行推薦商品上的調整利於推薦質量上的提高。

有鑑於此，本論文為了解決上述問題以及對於新顧客群於本篇論文中介界定於此新顧客為加入此電子商務網站為期一個月以內之新顧客推薦上得棘手問題，提出以網路的點擊序列來進行分析並且預測顧客的交易決策以便提供網站管理者能清楚的區分出顧客的購買意願進而針對顧客進行目標行銷(Target marketing)，然而為了掌握完整的顧客行為樣式，從現有的顧客資料中不論是新顧客或是舊有顧客網頁管理者都共同所有的資訊為點擊資料，以至以點擊序列資料進行蒐集和應用從一開始的商品瀏覽的序列到最後的購買決策以及購買了什麼產品，便能進行完整的追蹤根據並且也需考慮到了顧客可能因時間而改變的行為樣式或是個人偏好等因素，因此本篇論文所使用的點擊序列方法會直接且明確的反應出顧客改變的行為樣式與個人偏好，點擊序列於電子商務環境中所代表著與現實生活中真實的消費行為改變，然而，提出了為了解決新顧客於網頁上得推薦機制因而利用點擊資料來進行分析與改善，而此方法也針對顧客資料的不足性與產品資訊的稀疏性等限制有所解決，在本篇所提出的方法中使用過去一段時間已有的其他顧客點擊資料當作訓練集資料(Training data)，來預測此新顧客的購買決策為何針對此結果進行未來顧客的推薦機制上的不同，在訓練集資料的選擇上為了避免年代過於久遠資料導致預測結果的不準確性發生，因此使用最接近目前的月數來進行分析未來幾個月內的推薦。

因此本篇論文所提出基於於網路點擊的分析中利用點擊序列，來觀察顧客瀏覽的行徑間的順序間的關聯性之外並且分析此路徑最後的購買決策，以此概念達到充分掌握顧客行為以提升推薦之效度與準度改善過去之推薦上的不準確性以及錯失資料，並且減少無謂的

虛擬行銷成本並且增加網站之收益，而本論文所提出之推薦方法以利決策者進行精確的目標行銷外也具擴展性，並不只有本篇所界定之新顧客之外也可以使用此推薦方法來提高推薦之準確度及取而代之制式化之推薦。

2. 文獻探討

在於目前網路技術的蓬勃發展，現今的電子商務網站可說是越來越多，並且在眾多電子商務網站中不外乎的就是要增加網站的曝光度進而增加瀏覽人數，然而只讓瀏覽者人數變多並不是每個網站管理者的最終目的，最主要的目的在於該如何把網站瀏覽者變成購買者甚至最後變成傳播者，所以從一開始到底要怎樣吸引使用者的目光到最後完成交易的過程當中除了增加黏著度之外也使顧客於此電子商務網站在搜尋所需之商品便利且快速達到整個交易流程是順暢且滿意的，所以網站中必須有個領導者，帶領著顧客從巨量的資訊當中找出顧客真正所需要並且提供所謂的個人化服務機制[13]，而此機制則為「推薦系統」。

推薦系統於目前使用的相當廣泛，而這當中用在推薦機制上常用資料探勘的技術，像是內容導向過濾、協同過濾[1,2,6]、關聯法則[3,5,9]或序列樣式[6,7,12]等等。

內容導向過濾方法並須先建立使用者偏好檔案，也就是使用者的一些喜好上的收集之後，系統在於進而推薦符合此使用者偏好上的需求，此方法有受限制的地方，就是當使用者的偏好資料不足時則無法進行推薦以及推薦的商品相似度很高並無考慮到其他商品的關聯性。

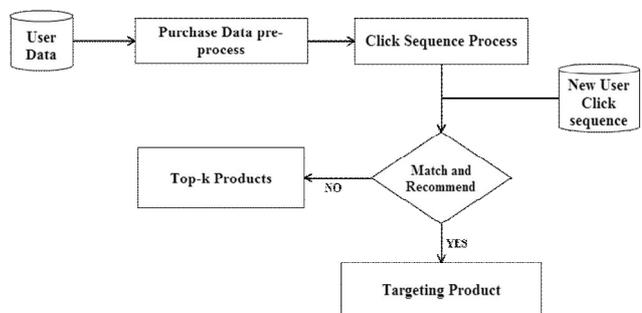
協同過濾則是計算使用者之間的相似度，依據計算出之相似度高的使用者在偏好的商品上會有一定相似程度因而進行做推薦，更有學者指出其實這原理就如同自動化的口碑行銷，系統所推薦的商品是基於其他口味類似的使用者之偏好而決定的[1]，然而此協同過濾技術對於新商品、新使用者與資料稀疏性等有所限制，對於新的商品可能評價上的資料不足透過此技術會無法廣泛的推薦增加曝光度，而對於新的使用者可能還未對眾多商品做評價，因為無法很精準的計算出此使用者與其他使用者之間的相似程度，而所謂之稀疏性是指計算之評價資料的不足，像於電子商務網頁初期一開始使用者人數不多並且商品數目也有限，即可計算之資料就會顯得稀少且不足，然而這些評價不足、獲取不易等問題，[6]此學者

提出已隱性(Implicit)評價資料的計算方式來取代傳統顯性(Explicit)評價等問題，不過使用此隱性評價的方式此顧客必須擁有購買紀錄，不然此方式依然取代顯性評價的問題。

關聯法則可以說是使用最普遍的並且由[4,5]學者所提出，此法則就是在尋找當一堆交易資料當中所購買的商品間的關聯性，哪些商品會被一起放到購物來中，然而[9]學者即使用關聯法則來挖掘點擊流資料進行做推薦，當中關聯法則只強調商品間的關聯性，並沒有去探討商品被放進購物籃中的先後順序，在於實際的商務環境中，放入購物籃中的先後必然是有順序性的本質，更可能因為先放入商品 A 之後才發現進而放入商品 B，有鑑於此，本篇所以提出基於網路點擊序列預測顧客交易決策以提升推薦效益之方法，呈現於電子商務網頁中的真實點擊序列行為並且反應於最後交易決策，並對於新的使用者使用點擊資料取而代之一開始上未取得與較不容易取得之購買資料與個人偏好克服新使用者問題。

3. 研究方法與步驟

以下將介紹本篇論文所提出基於網路點擊序列預測新顧客交易決策以提升推薦效益之方法，將由圖一來說明整個研究架構包括研究概念與建構處理方式。



圖一、研究方法架構圖

3.1 研究架構概念

首先於圖一所顯示的顧客資料(User Data)是指網站管理者所擁有的顧客資料中以本篇所提之方法需使用顧客的點擊資料與交易資料來進行處理，而必須對顧客資料中的交易資料進行資料前處理，因要預測其購買決策的部分，所以必須對顧客交易資料中，篩選出分別有確定購買的交易資料以及完整此交易的點擊路徑，接下來的處理程序之點擊序列處理(Click Sequence Process)必須對上一步驟前處理的資料彙整顧客點擊路徑及購買資料進行分析後處理成點擊序列，如顧客於電子商務網

站進行瀏覽網頁，然而瀏覽眾多的網頁中，於此瀏覽網頁當中有一點擊序列(Web click-sequence)為 ABCE，並且於網頁 C 購買商品 2 和 3 以及於 E 購買商品 5，而此路徑將於資料前處理的步驟中被選取出來並且經由點擊序列處理成為建置樹狀結構並且表示為 $WCS\langle ABC\{2,3\}E\{5\}\rangle$ ，而路徑 ABCE 則視為有效購買路徑，此處理程序完成後會都依據此路徑處理方式建構成為點擊序列購買決策樹(Click-sequence purchase decision tree, CSPD-tree)，而此 CSPD-tree 包含有效購買路徑，其它無效購買路徑也就是此路徑最後並沒有進行購買交易行為則將不納入樹狀結構中，最後 CSPD-tree 完成後整個分析的程序就此完成，而新顧客的點擊序列資料進入後就可以開始比對此新顧客目前的路徑是否為有效購買路徑，一經比對(Match)後發現此顧客路徑為有效購買路徑即可對此新顧客進行目標產品(Targeting product)的行銷推薦，然而如果對於比對後的路徑為無效購買路徑則可對此顧客推薦 Top-k 商品上的推薦行銷手法。

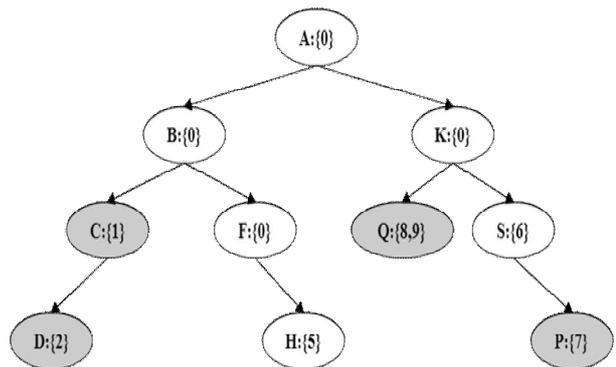
3.2 研究方法建構與處理

此研究的建構與處理方式會分為兩個步驟一為先對過去已有的點擊資料進行處理並且建構成樹狀的資料結構，然而二則在新顧客的點擊序列進來後，與建構完成的樹狀結構進行比對，針對不同的比對結果進行行銷策略上的差異。

3.3.1 購物資料前處理 (Purchase data pre-process)

當網站管理者擁有眾多的交易資料與點擊資料時，必須先找出所需並且針對這些有用的資料進行篩選，來節省時間與空間避免花費在不需要的資料上，由於本篇所提出的方法必須先找出到底哪些是有效購買路徑，因此必須先對已擁有之交易資料與點擊資料進行篩選，篩選出哪些路徑最後顧客是有購買商品，而如果此路徑顧客最後無購買任何商品那則視為無效購買路徑則不篩選進入下個階段分析處理程序，以圖二來做說明，此圖為資料庫中點擊路徑與交易資料以樹狀結構圖呈獻，其中點擊路徑為 AKSP 當中，此顧客於 P 網頁買了商品 7，此路徑則視為有效購買路徑並且於前處理程序中被篩選，反之點擊路徑為 ABFH 因整個路徑的過程中顧客都無於此路徑購買任何商品，因此此路徑則視為無效購買路徑並且於前處理程序中不被篩選進入下個階段分

析處理程序。



圖二、點擊路徑與交易資料樹狀圖

3.3.2 點擊序列購買決策樹(Click-sequence purchase decision tree, CSPD-tree)建構

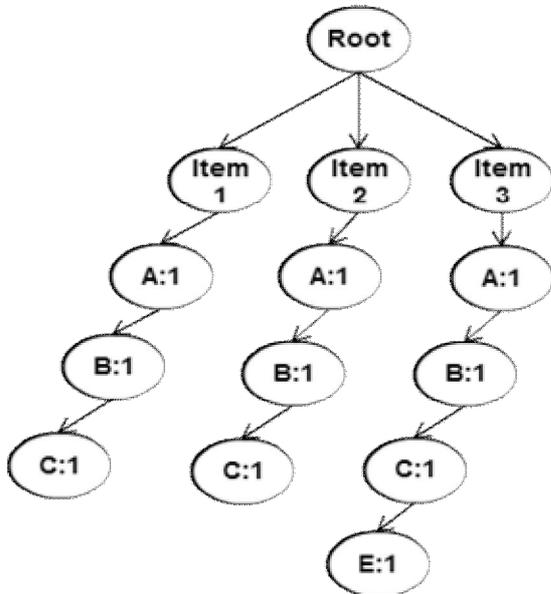
首先此步驟先須對過去已有的 $i, \forall i = 1, 2, \dots, N$ 個月數顧客點擊資料與交易資料進行處理成點擊序列並且建構成 CSPD-tree，而 CSPD-tree 則是以二元樹的概念建構，並且是由多個點擊網頁 Page $p_1, p_2, \dots, p_k, \forall p = 1, 2, \dots, N$ 所建構出節點，並由這些節點分別建構出不同商品的分支最後組成一顆完整的 CSPD-Tree，然而每筆顧客交易點擊序列資料都會包含瀏覽編號(Traversal identifier, TID)、瀏覽點擊路徑與購買的商品，如表一所示 TID 為 1 之瀏覽點擊路徑為 ABCE，並且分別在 C、E 網站各購買商品 1、2 與 3。

表一、顧客交易與點擊資料

TID	WCS
1	ABC{1,2}E{3}
2	ACD{3}
3	AEG{6}BC{2}

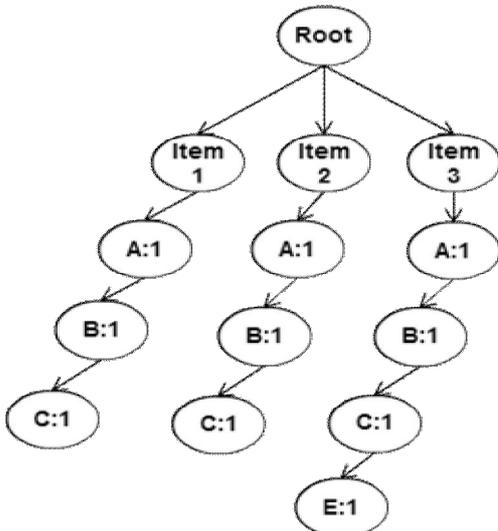
以下將舉表一之範例說明，假設目前已進行完成上一步驟之資料前處理，所篩選出之有效購買路徑為表一中三筆資料，而當中第一筆資料也就是 TID 為 1，則將此路徑先分別建構成購買商品 1 的樹狀支點為 ABC、購買商品 2 的樹狀支點為 ABC 與購買商品 3 的樹狀點擊路徑為 ABCE 以圖三表示，然而每個節點當中儲存了兩個欄位，分別為 Page、p_sup，Page 為顧客所瀏覽之網頁而 p_sup 則為瀏覽此網頁的支持度，而在根節點(Root)下之第一層節點則分別為各個商品項目，此節點往下延展才為點擊序列，而在此路徑當中 ABCE 進行樹狀切割時。並不是切割成為 {ABC}{E} 的原因在於

因此顧客的此交易路徑應視為同一個 Session，也就是為同一個 TID 的路徑組合，故不分別切割，然後在 Item3 中的葉節點支持度 (p_sup) 都為 1，也是原因在於都是出自同一個 TID 因此都為同一個 Session 的瀏覽，故 p_sup 都為一。

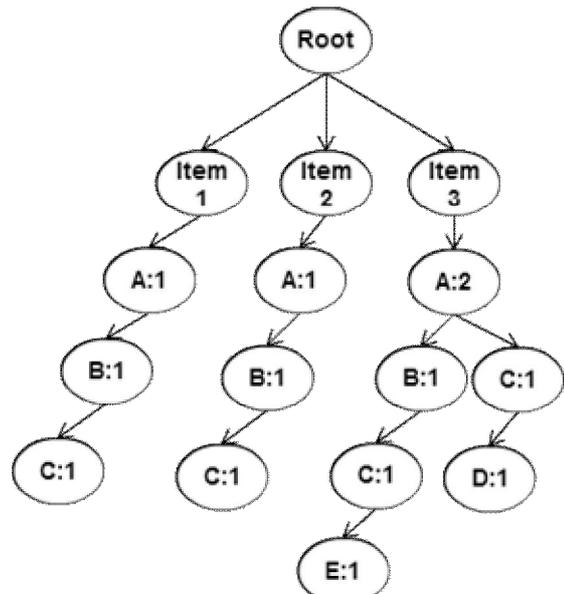


圖三、交易點擊序列之 CSPD 建構

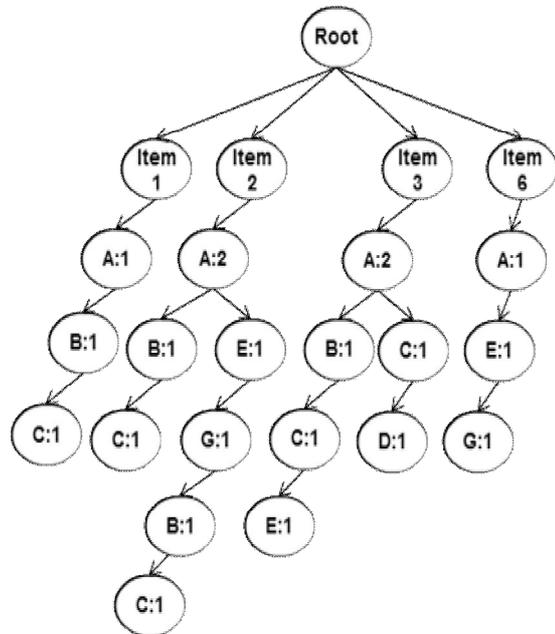
接下來示範整個處理的流程，也以表一之資料為例進行建構，由圖四至圖六呈現。



圖四、處理 WCS<ABC{1,2}E{3}>



圖五、處理 WCS<ACD{3}>



圖六、處理 WCS<AEG{6}BC{2}>

3.3.3 預測交易決策與推薦

經由點擊序列處理建構完整的 CSPD-tree 後，接下來將可以針對我們目標顧客的點擊路徑進行比對，判斷此目標顧客的點擊路徑到底是否會購買商品，如一經比對之後此目標顧客是有效購買路徑則進行目標行銷，並且推薦目標商品給此目標顧客，反之如發現此目標顧客經比對之後為無效購買路徑即對目標客戶進行 Top-k 的商品推薦，此流程呈現於圖七以演算法表示如下。

Input: CSPD-tree, new user click sequence, a user-defined i months to be training data, $\forall i = 1, 2, \dots, N$ and a user-defined j months to predict months, $\forall j = 1, 2, \dots, N$

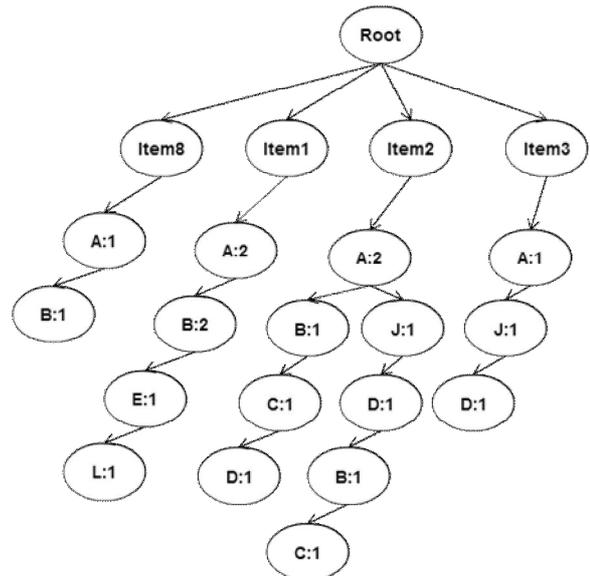
Output: recommendable strategy

```

foreach new user sequences  $\in$  CSPD-tree do
  if new user sequence  $\notin$  CSPD-tree then
    recommend the top-k product for user
  else
    recommend the targeting product from the CSPD-tree that product node in the second layer
  end if
end for

```

圖七、預測交易決策與推薦之演算法



圖八、建構完成圖

3.3.4 修剪 CSPD-tree 樹狀結構

為了防止 CSPD-tree 無限量的增長和擴大，因而造成執行速度上的緩慢或是記憶體佔用容量過多，因此定義儲存此顆樹狀結構的記憶體上限容量，當超過此容量或者是樹根節點個數大於 α ，就進行修剪的機制，可由節點支持度最小開始依序進行刪除。

4. 範例說明

此章節舉一範例說明本篇所提出之方法流程，將如下述詳細之說明。

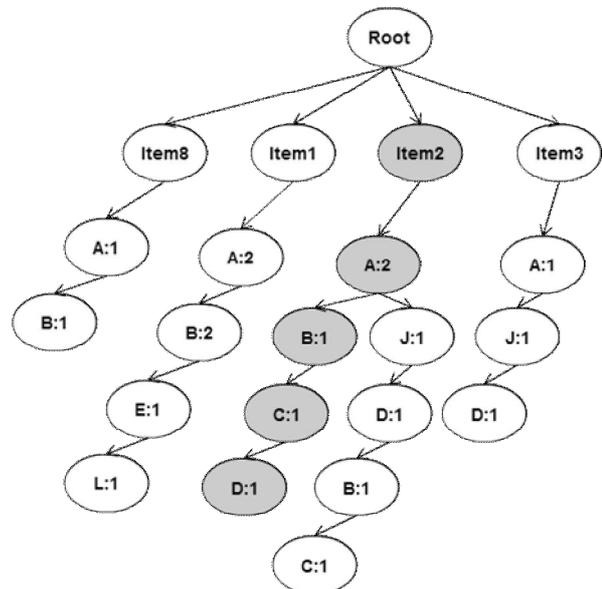
最先開始網站管理者所擁有之紀錄點擊資料及於點擊過程當中是否有發生任何交易紀錄之檔案，這當中我們所需要擷取並且分析的就只有有進行購買商品的點擊路徑，如表二所示，指有 TID 為一之 WCS 是不會被納入分析，因為此 WCS 並無購買任何商品故視為無效購買路徑，其它三個 WCS 則為有效購買路徑，因此納入資料分析當中。

表二、購物資料前處理

TID	紀錄檔之點擊路徑與購買商品	是否擷取分析
1	WCS<AVODJGB>	否
2	WCS<AB{8}EL{1}>	是
3	WCS<AB{1}CD{2}EK>	是
4	WCS<AJD{3}BC{2}>	是

再者進行 CSPD-tree 建構，如圖八為建構完成圖。

接下來將針對目標顧客之點擊路徑進行比對，如目前目標顧客之點擊路徑為 ABCD，在進行推薦時則選擇商品 2 為目標行銷如圖九，以此類推，但如點擊路徑無比對成功則選擇電子商務網站之網站 Top-k 商品推薦。



圖九、目標行銷比對

5. 結論

經由本篇論文所提出之基於網路點擊序列預測顧客交易決策以提升推薦效益，此方法不只提高推薦上的質量，並且也解決了一般推薦系統上難以解決的問題，那就是當新的顧客或是原有顧客的偏好資料、交易資料或者評價資料等等以上與顧客有關之有效分析資料不足時對於推薦機制執行上的推薦效益程度有

所影響甚至是採用較早的制式化推薦，故本篇採用點擊序列資料來進行顧客分析克服顧客資料的稀疏性與不足性，並且也對於電子商務網站初期的資料不足性有所解決的辦法，也對於顧客隨著時間而改變自身喜好或者習慣上，經由點擊瀏覽序列上有直接且實質上的反應，可快速掌握顧客偏好資料，增加顧客於電子商務網站之黏著度，始用此方法之對象並不局限於新加入電子商務網站之新顧客，對於已擁有長久之顧客也可採用此方法，故本篇題出此方法具擴展性。

參考文獻

- [1] 羅健銘，”協同過濾於網站推薦之研究”，台北科技大學商業自動化與管理研究所碩士論文，1991。
- [2] 吳肇銘、金志聿與林怡秀，”協同過濾技術在商品推薦系統上之應用與成效評估”，第十五屆 ICIM 國際資訊管理學術研討會，12 頁，2004-05-29
- [3] 李御璽與顏秀珍，”網路交易型態探勘技術之研究”，電子商務學報，10 卷，4 期，2008-12-01，Pages 989-1008
- [4] Agrawal, R., & Srikant, R. (1994). Fast algorithm for mining association rules. In Proceedings of the twentieth VLDB conference (pp. 487–499). Santiago.
- [5] Agrawal, R., Imieliński, T. and Swami, A., "Mining association rules between sets of items in large databases", Proceedings of the 1993 ACM SIGMOD international conference on Management of data - SIGMOD '93, pp. 207-216
- [6] Keunho Choi, Donghee Yoo, Gunwoo Kim and Yongmoo Suh, "A hybrid online-product recommendation system: Combining implicit rating-based collaborative filtering and sequential pattern analysis", Electronic Commerce Research and Applications, Volume 11, Issue 4, July–August 2012, Pages 309–317
- [7] Bi-Ru Dai, Hung-Lin Jiang and Chih-Heng Chung, "Mining Top-K Sequential Patterns in the Data Stream Environment", 2010 International Conference on Technologies and Applications of Artificial Intelligence, pp. 142-149
- [8] Goksedef, M., & Gunduz-Oguducu, S. (2010). Combination of web page recommender systems. Expert Systems with Applications, 37(4), 2911–2922
- [9] Yongsoo Kim and Bongjin Yum, "Recommender system based on click stream data using association rule mining", Expert Systems with Applications, Volume 38, Issue 10, September 2011, Pages 13320-13327
- [10] Kim, D., & Yum, B.-J. (2005). Collaborative filtering based on iterative principal component analysis. Expert Systems with Applications, 28(4), 823–830
- [11] Kim, Y. S., Yum, B.-J., Song, J., & Kim, S.-M. (2005). Development of a recommender system based on navigational and behavioral patterns of customers in ecommerce sites. Expert Systems with Applications, 28(2), 381–393
- [12] Hua-Fu Li, Suh-Yin Lee and Man-Kwan Shan, "DSM-PLW: Single-pass mining of path traversal patterns over streaming Web click-sequences", Computer Networks, Volume 50, Issue 10, 14 July 2006, Pages 1474–1487
- [13] Resnick, P., "Recommender Systems", Communications of the ACM, 40(3), 1997, pp. 56-58