

利用探勘技術發掘具有異常喊價傾向之競標者

陳垂呈*

(收稿日期：95 年 9 月 6 日；第一次修正：95 年 10 月 16 日；
接受刊登日期：95 年 11 月 16 日)

摘要

線上拍賣是電子商務經營模式中最受消費者歡迎的交易方式之一，雖然線上拍賣可以帶給參與者更多的購買樂趣，但在交易過程中也存有較多的風險，其中如何避免異常競標者哄抬拍賣商品的價格，是線上拍賣詐欺重要的問題之一。在本篇論文中，我們以競標者之交易資料為探勘的資料來源，每一筆交易資料包含競標者曾經喊價過與購買過的商品項目，我們設計兩個探勘方法，分別從以下兩方面發掘具有異常喊價傾向的競標者：一是設計一個兩階段的探勘方法，從交易資料中找出具有異常喊價傾向的競標者；二是以某一拍賣商品為探勘的目標，設計一個兩階段的探勘方法，找出對此一拍賣商品具有異常喊價傾向的競標者。我們根據所提出的方法，設計與建置一個偵測異常競標者的系統。此探勘結果，對拍賣商品避免被惡意地哄抬競價將可提供非常有用的資訊。

關鍵詞彙：資料探勘，線上拍賣，異常競標者

壹·導論

隨著電子商務 (e-commerce) 的盛行及物流金流的發達，消費者可以在不受時間及空間的限制下，很方便地在網路上瀏覽商品並進行購買。在電子商務眾多的經營形態中，線上拍賣是最受消費者歡迎的交易方式之一，在線上拍賣的網站上，消費者可以陳列寄賣的商品，而競標者可以享受到自由喊價的趣味性。隨著拍賣交易的熱絡，也隨之衍生了許多的交易問題，其中最常見的有競標者喊價之後，未能履行付款購買拍賣品的義務，或是某些競標者只是為了哄抬拍賣品的價格，而不是實際真正的競標者，因而影響真正想要喊價之競爭者。因此如何偵測出異常的競標者，使得拍賣交易能夠正常並公正的進行，即成為線上拍賣經營者必須考量的問題之一。

藉由資訊技術的支援，線上拍賣經營者可以很輕易地蒐集到競標者的交易記錄資料，並且快速的累積。若能從這些大量的交易資料中，找出競標者的消費模式，以協助偵測競標者之喊價是否異常，對於拍賣商品避免被惡意哄抬

* 作者簡介：陳垂呈，南台科技大學資訊管理系副教授。

價格，或事先預防無意購買拍賣商品之競標者的喊價行為，以遏止異常喊價行為之事件持續的發生，將可提供非常有用的資訊。

資料探勘 (data mining) 是從大量資料中找出潛在有用的知識與資訊，以提供決策分析之參考資訊，目前資料探勘技術已普遍地應用在許多的領域中 (Han and Kamber, 2006)。在本篇論文中，我們將利用探勘技術偵測那些競標者具有異常喊價的行為特徵，並根據所提出的方法，設計與建置一個偵測異常喊價競標者的系統。

在美國聯邦貿易委員會 (Federal Trade Commission) 所公佈之「掃蕩網路詐欺之犯罪報告」中 (<http://www.ftc.gov>, 2000)，將賣方偽裝成其他消費者，一同參與喊價進而哄抬拍賣商品的價格，視為網路拍賣詐欺重要類型之一。在本篇論文中，我們以競標者之交易資料為探勘的資料來源，每一筆交易資料包含競標者曾經購買過的商品項目與喊價過的商品項目，根據購買商品的忠誠性與關聯性，利用探勘技術分別從以下兩方面發掘具有異常喊價傾向的競標者：

1.發掘具有異常喊價傾向之競標者：

我們設計一個兩階段探勘的方法，以判斷出具有異常喊價傾向的競標者。在第一階段中，我們比對競標者曾經購買過的商品項目與喊價過的商品項目，若滿足所設定的「最小購買忠誠度」，則根據購物傾向的忠誠性，可判斷出競標者為非具有異常喊價傾向的競標者；否則，則必須進行下一階段的處理。在第二階段中，我們從競標者曾經購買過的商品項目中找出關聯規則 (association rules)，然後藉此計算競標者曾經喊價過之商品項目的關聯度，其關聯度表示競標者曾經喊價過之商品項目彼此間的關聯強度，以做為判斷競標者是否具有異常喊價傾向的依據。

2.發掘某一拍賣商品具有異常喊價傾向之競標者：

假設目前某一拍賣商品正進行競標，我們以此一拍賣商品為探勘的目標，設計一個兩階段探勘的方法，找出對此一拍賣商品而言具有異常喊價傾向的競標者。在第一階段中，我們從目前正在對此一拍賣商品進行喊價的競標者中，比對這些競標者曾經購買過的商品項目與喊價過的商品項目，若滿足所設定的「最小購買忠誠度」，或是在曾經購買過的商品項目中，若包含此一拍賣商品，則根據購物傾向的忠誠性、重覆性，可判斷出競標者為非具有異常喊價傾向的競標者；否則，則必須進行下一階段的處理。在第二階段中，我們從全部競標者曾經購買過的商品項目中，設計一個快速的方法，找出包含此一拍賣

商品的關聯規則，然後藉此計算這些競標者曾經喊價過之商品項目的關聯度，以做為判斷競標者是否具有異常喊價傾向的依據。

本文各章節介紹如下：第貳節介紹資料探勘技術、及其在偵測異常交易的相關研究；第參節說明發掘具有異常喊價傾向之競標者的探勘方法；第肆節說明發掘某一拍賣商品之競標者是否具有異常喊價傾向的探勘方法；第伍節為依據所提出的方法，設計與建置一個偵測異常競標者的系統；最後，在第陸節中做一結論。

貳·相關研究

資料探勘 (data mining) 是從大量資料中挖掘出潛在有用資訊與知識，發現專家尚且未知的新關係，以協助企業專業人員決策分析之參考。目前資料探勘技術已普遍地應用在各領域中，資料探勘可完成以下任務或是更多：關聯規則 (association rules)、分群 (clustering)、分類 (classification)、次序相關分析 (sequential pattern analysis) 等 (Chen et al., 1996; Han and Kamber, 2006)，其中又以關聯規則最常被使用做為表示商品項目之間的關聯性。

在眾多資料探勘應用的問題中，欺騙偵測 (fraud detection) 是其重要的應用之一 (Shao et al., 2002)，目前已有許多利用資料探勘技術分析交易異常與欺騙的相關研究 (翁頌舜、鄭富山, 2002; 陳怡婷, 2004; 詹培萱, 2002; 徐源隆, 2003; 葛煥元, 2003; 朱桓毅, 2004; Bonchi et al., 1999; Brause et al., 1999; Metwally et al., 2005; Chui and Tsai, 2004; Kou et al., 2004)。Brause et al. (1999) 曾利用類神經網路 (neural network) 技術偵測信用卡的交易是否異常; Chui and Tsai (2004) 則利用關聯規則做為偵測信用卡交易是否異常的方法依據; 翁頌舜和鄭富山 (2002) 曾利用資料探勘技術探討電信資料異常的研究; 陳怡婷 (2004) 曾利用資料探勘技術於犯罪資料的探勘，並實驗證明其探勘技術可在網路犯罪之偵查，確實可以提供有用的參考資訊。關於線上拍賣之消費行為及詐欺犯罪的相關研究，詹培萱 (2002) 曾利用資料探勘技術分析拍賣網站之消費者的購買行為，依據探勘結果將消費者分成三個消費族群，以提供拍賣網站之經營者或是賣方有用的行銷建議; 徐源隆 (2003) 曾探討網路拍賣詐欺的偵查模式，以協助警方偵辦網路拍賣詐欺的案件; 而葛煥元 (2003)、朱桓毅 (2004) 則在探討關於線上拍賣所衍生出的詐欺及欺騙等問題。

Agrawal et al. (1993) 首先提出探勘關聯規則以顯示出項目之間的關聯性，我們說明關聯規則的定義如下：假設 I 是所有項目的集合， T 是全部交易

資料的集合，一筆交易資料 T_j , $T_j \in T$ ，是由一些項目所形成的集合，稱之為項目組 (itemsets)，若一個項目組包含 k 個項目，稱之為 k -項目組 (k -itemsets)，以 $itemset_k$ 表示之， $k \geq 1$ 。在項目組 X 與 Y 之間有一關聯規則被表示成 $X \rightarrow Y$ ， $X, Y \subseteq I$ ，且 $X \cap Y = \emptyset$ 。有兩個參數 s 與 c 分別為支持度 (support) 與信賴度 (confidence)，做為決定關聯規則是否成立的依據，其中支持度 s 的定義為：在所有的交易資料集合中，同時包含 $X \cup Y$ 的比率值，即 $s = (\text{同時包含 } X \cup Y \text{ 的交易資料數量}) / (\text{總交易資料數量})$ ；信賴度 c 的定義為：在包含 X 的交易資料集合中，也同時包括 Y 的比率值，即 $c = (\text{同時包含 } X \cup Y \text{ 的交易資料數量}) / (\text{包含 } X \text{ 的交易資料數量})$ 。擷取出的關聯規則，其支持度與信賴度必須大於或等於所指定的最小支持度與最小信賴度，這樣的關聯規則才成立。

關聯規則的探勘過程，主要分成以下兩個階段：首先，找出滿足最小支持度的所有項目組，這些滿足最小支持度的項目組就稱之為高頻項目組 (frequent itemsets)，若某 k -項目組滿足最小支持度，即稱之為高頻 k -項目組 (frequent k -itemsets)，以 $frequent_k$ 表示之；然後，就根據前階段所找出的高頻項目組及以最小信賴度為條件，計算出所有符合的關聯規則。例如 ABC 為高頻 3-項目組，假如關聯規則 $AB \rightarrow C$ 滿足最小信賴度，則此關聯規則成立，擷取關聯規則的相關研究可參考 (Agrawal and Srikant, 1994; Srikant and Agrawal, 1995; Park et al., 1997)。

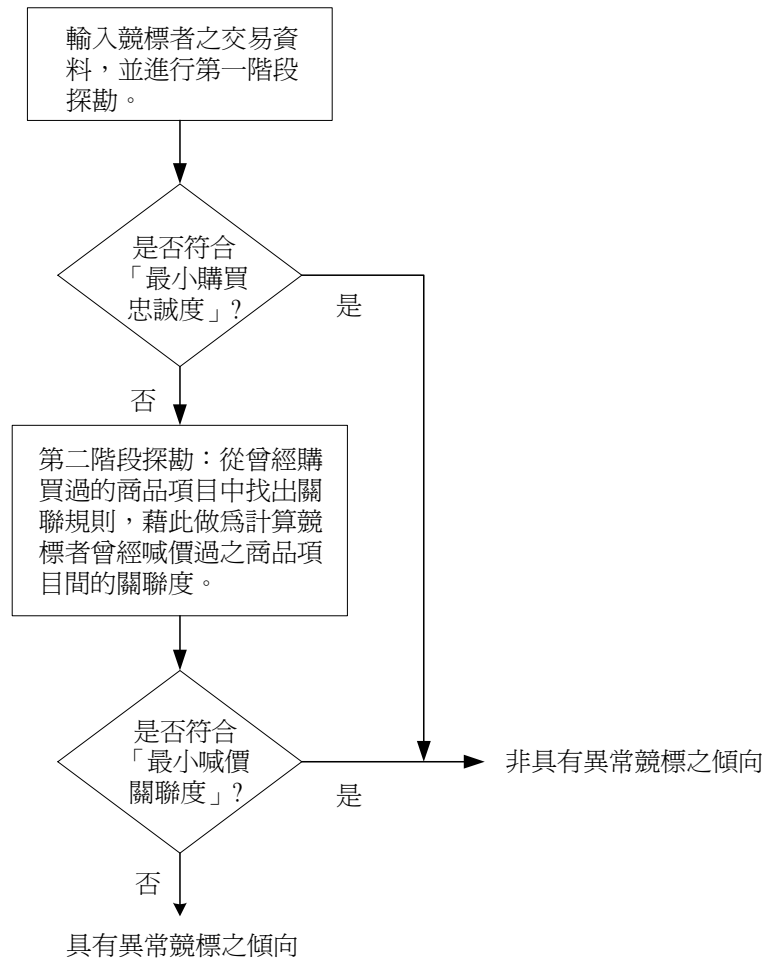
在本篇論文中，我們將探討線上拍賣之異常競標者的探勘問題，並以關聯規則做為偵測的方法依據，從以下兩方面偵測具有異常喊價傾向的競標者：一是從競標者之交易資料中找出具有異常喊價的傾向者；二是以目前某一拍賣商品為探勘的目標，找出對此一拍賣商品而言，具有異常喊價傾向的競標者。

參·發掘具有異常喊價傾向之競標者

我們以競標者之交易資料為探勘的資料來源，其交易資料包含喊價過的商品項目及購買過的商品項目，並且購買過的商品項目 \subseteq 喊價過的商品項目，設計一個兩階段的探勘方法，發掘具有異常喊價之行為特徵的競標者。接下來，我們說明偵測競標者是否具有異常喊價傾向的探勘方法，並以一實例說明探勘的過程。

一、發掘具有異常喊價傾向之競標者的探勘方法

我們以競標者之交易資料為探勘的資料來源，每一筆交易資料包含競標者曾經購買過及喊價過的商品項目，其中競標者曾經購買過的商品項目 \subset 喊價過的商品項目，設計一個兩階段的偵測方法，以發掘具有異常喊價之行為特徵的競標者，其探勘過程如圖一之流程圖。



圖一 探勘具有異常喊價傾向之競標者的流程圖

Butler and Peppard (1998) 以五個步驟說明消費者購物的決策模式：(1)需求的確認；(2)資訊的搜尋；(3)選擇的評估；(4)購買的執行；(5)購買後的行為。延伸此模式於網路購物之消費者行為的相關研究有 (朱桓毅，2004)，競標者網路拍賣之喊價行為屬於以上決策模式的第(4)步驟，因此可以確認競標者對

喊價之拍賣商品有其需求性。我們根據競標者對拍賣商品的需求性與其得標的記錄，定義「購買忠誠度」如下：

購買忠誠度=競標者曾經購買過之商品項目的個數/曾經喊價過之商品項目的個數。

若競標者曾經喊價過之商品項目的個數為 0 或 1，則忽略此競標者，其表示競標者尚未或初次進行競標的活動。從以上的定義中，其可以客觀地反應出競標者對其喊過價之拍賣商品的需求性及購買的迫切性。例如，某一競標者曾經購買過 ABC 等 3 項商品，而喊價過 ABCDE 等 5 項商品，則其購買忠誠度為 $3/5=60\%$ 。

接下來，我們說明發掘具有異常喊價傾向之競標者的探勘過程：在第一階段中，我們設定一個「最小購買忠誠度」，然後計算各競標者的購買忠誠度，若滿足所設定的「最小購買忠誠度」，則依據購買行為的忠誠度可判斷競標者為非具有異常喊價之傾向者；若未能滿足所設定的「最小購買忠誠度」，則必須進行下一階段的處理。

在第二階段中，我們從競標者曾經購買過的商品項目中找出關聯規則，因為購買過的商品項目是最能確實反應消費者之消費特徵、及商品項目間關聯性的資料，然後根據關聯規則所顯示出的消費傾向，計算競標者曾經喊價過之商品項目間的關聯性。在眾多探勘關聯規則的方法中，Apriori 演算法 (Agrawal and Srikant, 1994) 是最具代表性的方法之一，因此，我們利用 Apriori 演算法擷取商品項目之間的關聯規則。以下我們說明 Apriori 演算法擷取關聯規則的過程：

1. 從競標者曾經購買過的商品項目中找出 $frequent_{k-1}$ ， $k>1$ ，若為 \emptyset ，則停止執行。
2. 由步驟(1)中找出任兩個有 $k-2$ 項目相同的 $frequent_{k-1}$ 組合成 $itemset_k$ 。
3. 判斷由步驟(2)所找出的 $itemset_k$ ，其所有包含的 $itemset_{k-1}$ 之子集合是否都出現在步驟(1)中，假如成立就保留此 $itemset_k$ ，否則就刪除。
4. 檢查由步驟(3)所擷取的 $itemset_k$ 是否滿足最小支持度，假如符合就成為 $frequent_k$ ，否則就刪除。
5. 計算 $frequent_k$ 所形成的關聯規則，若滿足最小信賴度，則關聯規則成立。
6. 跳至步驟(1)找 $frequent_{k+1}$ ，直到無法產生高頻項目組為止。

從關聯規則中可顯示出競標者的購物傾向，然後依據關聯規則計算在競標者曾經喊價過的商品項目中，彼此之間的關聯性。在消費者購物的決策模式中，競標者對拍賣商品進行喊價的行為，已顯示出競標者對拍賣商品的需求性與興趣度，即使最後可能沒有得標購買，但也反應出競標者的消費行為傾向。我們根據以上演算法找出的關聯規則中，計算競標者曾經喊價過之商品項目彼此間的關聯性，定義競標者之「喊價關聯度」為：

$$\text{喊價關聯度} = (\max\{\text{關聯規則的項目個數}\}, \text{且此關聯規則中的項目} \subseteq \text{曾經喊價過的商品項目中}) / \text{曾經喊價過之商品項目的個數}。$$

其表示在競標者曾經喊價過的商品項目中，是否合乎關聯規則所顯示的關聯性，若競標者之喊價關聯度未能滿足所設定「最小喊價關聯度」，則顯示競標者具有異常喊價的傾向，否則為非具有異常喊價的傾向。例如，若找出關聯規則 $A \rightarrow B$ 、 $A \rightarrow C$ 及 $A \rightarrow BC$ ，某一競標者曾經喊價過的商品項目為 ABCD，根據以上喊價關聯度的定義，由於關聯規則 $A \rightarrow BC$ 為具有最大的項目個數，且 $ABC \subseteq ABCD$ ，因此，其喊價關聯度 = ABC 的個數 / $ABCD$ 的個數 = $3/4 = 75\%$ 。

二、實例說明

我們以表一之交易資料庫 D_1 為例， $\{A, B, C, D, E\}$ 為商品項目的集合， $\{T_1, T_2, T_3, T_4\}$ 為 4 筆競標者之交易資料的集合，設定最小支持度為 40% (即最小支持數量為 1.6)，最小信賴度為 70%，最小購買忠誠度為 60%，最小喊價關聯度為 50%。探勘具有異常喊價之行為特徵的競標者，其探勘的過程說明如下。

表一 交易資料庫 D_1

交易資料編號	購買過的商品項目	喊價過的商品項目
T_1	AD	ABCD
T_2	BE	ABDE
T_3	C	ABC
T_4	BE	BCE

首先，我們計算各競標者的購買忠誠度如下：

$$T_1 = AD \text{ 項目個數} / ABCD \text{ 項目個數} = 2/4 = 50\%$$

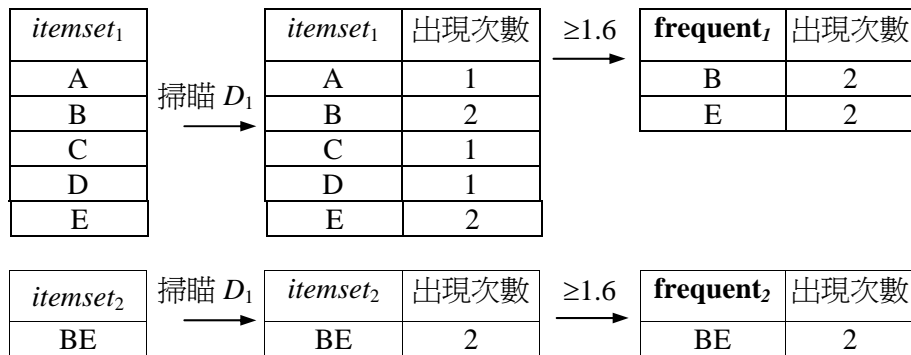
$$T_2 = BE \text{ 項目個數} / ABDE \text{ 項目個數} = 2/4 = 50\%$$

$$T_3 = C \text{ 項目個數} / ABC \text{ 項目個數} = 1/3 = 33\%$$

$$T_4 = BE \text{ 項目個數} / BCE \text{ 項目個數} = 2/3 = 67\%$$

經由上述的計算，發掘只有 T_4 的購買忠誠度滿足所設定的「最小購買忠誠度」，因此可判斷 T_4 為非具有異常喊價傾向的競標者，其餘競標者則必須進行第二階段的處理。

我們從競標者曾經購買過的商品項目中，利用 Apriori 演算法擷取商品項目之間的關聯規則，其擷取高頻項目組的過程如下：



無3-項目組。

擷取出的高頻2-高頻項目組有BE，計算其形成關聯規則 $E \rightarrow B$ 或 $B \rightarrow E$ 的信賴度，滿足「最小信賴度」。因此，根據關聯規則計算各競標者的喊價關聯度為：

T_1 之喊價關聯度為：因為 $BE \not\subseteq ABCD$ ，因此其關聯相似度為 $0/4=0$ 。

T_2 之喊價關聯度為：因為 $BE \subseteq ABDE$ ，因此其關聯相似度 $=2/4=50\%$ 。

T_3 之喊價關聯度為：因為 $BE \not\subseteq ABC$ ，因此關聯其相似度 $=0/3=0$ 。

經由以上相似度的計算，只有 T_2 滿足「最小喊價關聯度」，因此，具有異常喊價之行為特徵的競標者有 T_1 、 T_3 。

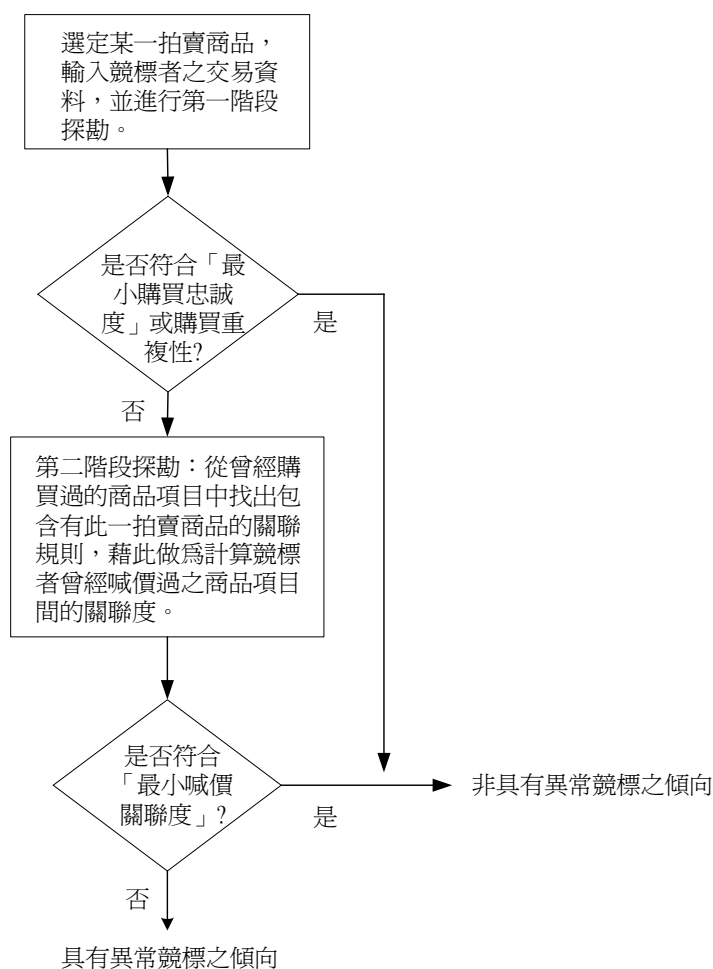
肆·發掘某一拍賣商品具有異常喊價傾向之競標者

我們仍以競標者之交易資料做為探勘的資料來源，並以某一拍賣商品之競標者為探勘的目標，對線上拍賣的經營者或是賣方而言，有時往往急於想知道此一拍賣商品之競標者是否具有異常喊價的行為特徵，因此，如何快速地偵測出此一拍賣商品之競標者是否具有異常喊價的傾向，即成為必須考量的問題

之一。接下來，我們說明偵測某一拍賣商品之競標者是否具有異常喊價傾向的探勘方法，並以一實例做說明。

一、發掘某一拍賣商品之競標者是否具有異常喊價傾向的探勘方法

我們以某一拍賣商品之競標者為探勘的目標，設計一個兩階段的探勘方法，判斷目前正在對此一拍賣商品進行喊價之競標者，是否具有異常喊價的傾向，其探勘過程如圖二之流程圖。



圖二 探勘某一拍賣商品具有異常喊價傾向之競標者的流程圖

在第一階段中，我們延用上一章節所描述之購買忠誠度的定義，計算這些競標者的購買忠誠度，若滿足所設定的「最小購買忠誠度」，則可判斷競標者為非具有異常喊價之傾向者。或是在這些競標者曾經購買過的商品項目中，若包含此一拍賣商品，則根據購物傾向的重複性，可判斷競標者為非具有異常喊價的傾向。若未能滿足所設定的「最小購買忠誠度」，則必須進行下一階段的處理。

在第二階段中，我們從所有競標者曾經購買過的商品項目中，修改 Apriori 演算法擷取商品項目間的關聯規則，其所組成的商品項目必須包含此一拍賣商品，假設此一拍賣商品項目為 A，其關聯規則形式必須為：

$A \rightarrow X$ (或 $X \rightarrow A$)， $A \cap X = \emptyset$ ，X 為一個或以上項目所形成的項目組。

從關聯規則所顯示出的消費傾向為：曾經購買商品項目 A 的消費者，也會購買商品項目 X 的傾向 (或是曾經購買商品項目 X 的消費者，也會購買商品項目 A 的傾向)。以下是我們為了配合探勘的需要，修改 Apriori 演算法，只擷取包含 A 的高頻項目組，稱之為 Apriori-1-Item 演算法，其過程說明如下：

1. 從所有競標者曾經購買過的商品項目中找出 $frequent_1$ ，必須包含 A，若為 \emptyset ，則停止執行。
2. 由步驟(1)中，A 與任一 $frequent_1$ 形成 $itemset_2$ ，檢查 $itemset_2$ 是否滿足最小支持度，若符合就成為 $frequent_2$ ，否則就刪除。
3. 找出所有的 $frequent_k$ ， $k > 1$ 。
4. 由步驟(3)中，組合任兩個有 $k-1$ 項目相同的 $frequent_k$ ，形成 $itemset_{k+1}$ 。
5. 判斷由步驟(4)所找出的 $itemset_{k+1}$ ，其所包含的 $\{A \cup itemset_{k-1}\}$ 之子集合是否都出現在步驟(3)中，若成立就保留此 $itemset_{k+1}$ ，否則就刪除。
6. 檢查由步驟(5)所找出的 $itemset_{k+1}$ 否滿足最小支持度，若符合就成為 $frequent_{k+1}$ ，否則就刪除。
7. 計算 $frequent_{k+1}$ 所形成的關聯規則，其形式必須為：
 $A \rightarrow X$ (或 $A \rightarrow X$)， $A \cup X$ 為 $frequent_{k+1}$
 若滿足最小信賴度，則關聯規則成立。
8. 跳至步驟(3)繼續找出 $frequent_{k+2}$ ，直到無法產生高頻項目組為止。

從以上演算法所擷取出的高頻項目組，必定包含此一拍賣商品 A，在探勘的過程中，我們只擷取包含 A 的項目組，因此能以較有效率地擷取出所要的關聯規則。計算高頻項目組所形成的關聯規則 $A \rightarrow X$ (或 $A \rightarrow X$)，若滿足最小信賴度，則關聯規則成立，從關聯規則中可顯示出競標者的購物傾向。然後依據關聯規則，計算在競標者曾經喊價過之商品項目中與此一拍賣商品間的關聯性。我們修改上一章節所描述之喊價關聯度的定義，定義競標者之「喊價關聯度」為：

拍賣商品 A 之喊價關聯度 = $(\max\{\text{關聯規則中 } X \text{ 的項目個數}\}) / (\text{曾經喊價過的商品項目} - \text{拍賣商品 A})$ 的個數。

其表示在競標者曾經喊價過的商品項目中，是否合乎包含商品 A 之關聯規則所顯示的關聯性。對目前正在對拍賣商品 A 進行喊價之競標者而言，若其喊價關聯度未能滿足所設定「最小喊價關聯度」，則顯示競標者具有異常喊價的傾向，否則為非具有異常喊價的傾向。例如，若找出關聯規則 $A \rightarrow B$ 、 $A \rightarrow C$ 及 $A \rightarrow BC$ ，某一競標者曾經喊價過的商品項目為 ABCD，根據以上喊價關聯度的定義，由於關聯規則 $A \rightarrow BC$ 為具有最大的項目個數，且 $BC \subseteq BCD$ ，因此，此一競標者之喊價關聯度 = BC 的個數 / BCD 的個數 = $2/3 = 67\%$ 。

二、實例說明

我們以表二之交易資料庫 D_2 為例， $\{A, B, C, D, E\}$ 為商品項目的集合， $\{T_1, T_2, T_3, T_4, T_5\}$ 為 5 筆競爭者之交易資料的集合，最小購買忠誠度為 60%，最小喊價關聯度為 60%，最小支持度為 40% (即最小支持數量為 2)，最小信賴度為 60%。假設目前一拍賣商品 A 正由競標者 $\{T_1, T_2, T_3, T_4, T_5\}$ 等 5 人進行喊價，探勘此一拍賣商品之競標者是否具有異常喊價的傾向，其探勘的過程說明如下。

表二 交易資料庫 D_2

交易資料編號	購買過的商品項目	喊價過的商品項目
T_1	BC	ABCD
T_2	DE	ADE
T_3	ABCE	ABCE
T_4	AB	ABC
T_5	ABC	ABCE

首先，我們計算各競標者的購買忠誠度如下：

$$T_1 = \text{BC 項目個數} / \text{ABCD 項目個數} = 2/4 = 50\%$$

$$T_2 = \text{DE 項目個數} / \text{ADE 項目個數} = 2/3 = 67\%$$

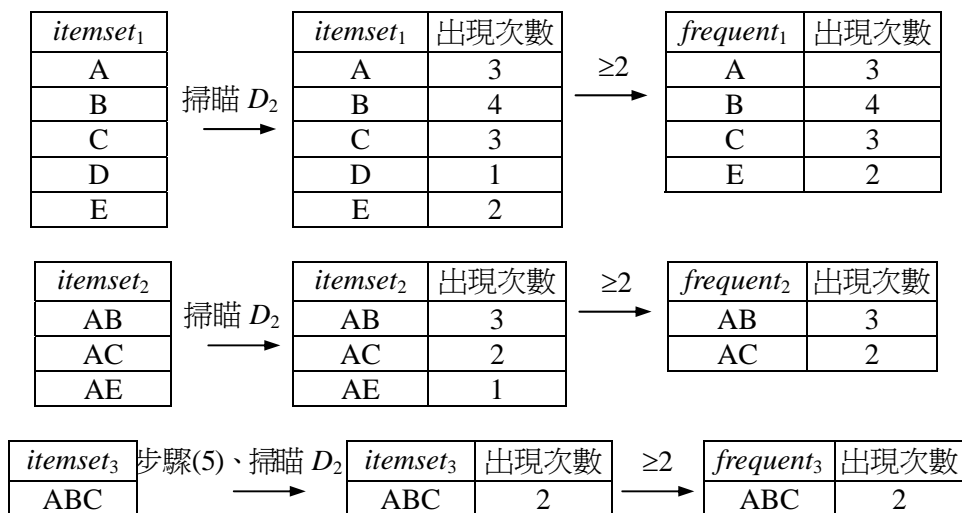
$$T_3 = \text{ABCE 項目個數} / \text{ABCE 項目個數} = 4/4 = 100\%$$

$$T_4 = \text{AB 項目個數} / \text{ABC 項目個數} = 2/3 = 67\%$$

$$T_5 = \text{ABC 項目個數} / \text{ABCE 項目個數} = 3/4 = 75\%$$

經由上述的計算，發掘 T_2 、 T_3 、 T_4 、 T_5 的購買忠誠度滿足所設定的「最小購買忠誠度」，因此可判斷為非具有異常喊價傾向的競標者，其餘競標者則必須進行第二階段的處理。

在第二階段中，我們利用 Apriori-1-Item 演算法，擷取包含商品項目 A 之高頻項目組的過程如下：



無4-項目組。

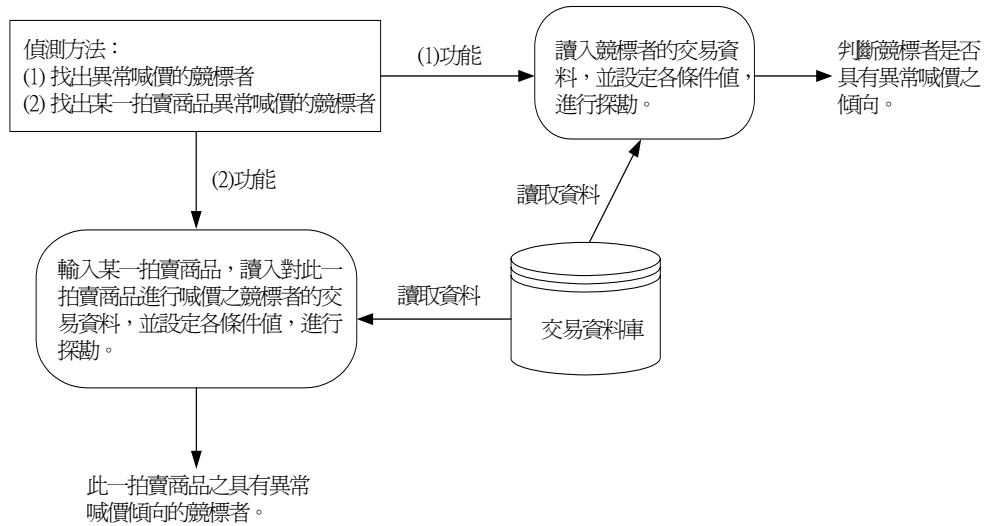
計算高頻 3-項目組 ABC 所形成的關聯規則 $A \rightarrow BC$ ，其信賴度為 $2/3=67\%$ ，滿足「最小信賴度」，因此關聯規則成立。我們計算競標者 T_1 之喊價關聯度為：

$$\text{喊價關聯度} = (\text{BC} \cap \text{BCD}) \text{ 的項目個數} / \text{BCD 的項目個數} = 2/3 = 67\%$$

滿足「最小喊價關聯度」，因此可判斷 T_1 為非具有異常喊價傾向的競標者。

伍・偵測異常喊價之競標者的系統實作

我們將前面章節所描述的探勘方法，設計與建置一個偵測具有異常喊價傾向之競標者的系統，系統探勘過程的模型如圖三。



圖三 系統探勘過程模型

我們以 C# 做為系統撰寫的程式語言，交易資料由 IBM Data Mining 網站 (<http://www.almaden.ibm.com/>) 下載資料模擬程式，以產生評估實驗中所需要的交易資料。為了評估探勘系統的執行效能，我們產生一個包含 50000 筆交易資料的交易資料庫 D ，其中每一筆交易資料所包含的項目視之為曾經喊價過的商品項目，而曾經購買過的商品項目再從中隨機產生，如表三。交易資料庫中的主要參數值其意義分別為： n 代表項目的數量、 $ntran$ 為交易資料筆數的數量、 np 為型樣組合的數量、 tl 為每筆交易資料的平均項目個數，其餘參數以預設值表示之。

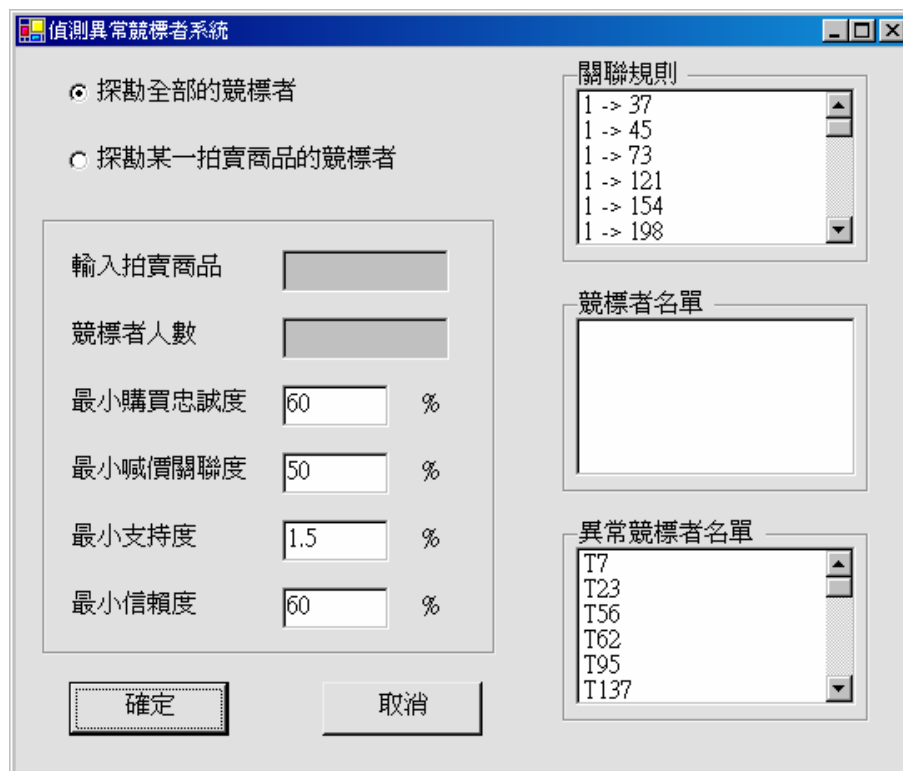
表三 交易資料庫與其參數

參數	n	$ntran$	np	tl
交易資料庫				
D	1000	50000	10000	10

我們分別以編號 1, 2, 3, ..., 1000 表示商品項目，以編號 T1, T2, T3, ..., T50000 表示交易資料。我們以交易資料庫 D 為探勘的資料來源，利用前面章

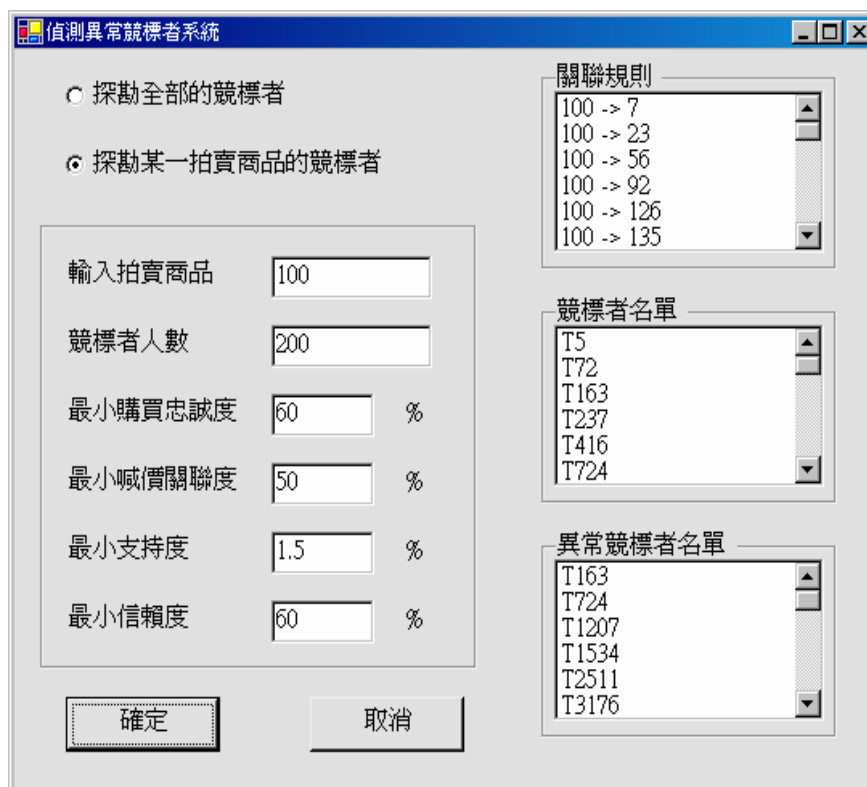
節所提出的方法，設計與建置一個探勘系統，以下為此一系統之探勘過程的操作介面。

圖四表示探勘的偵測畫面，其中包含兩項功能選項：「探勘全部的競標者」及「探勘某一拍賣商品的競標者」。假設目前點選「探勘全部的競標者」功能，並分別輸入「最小購買忠誠度」、「最小喊價關聯度」、「最小支持度」、「最小信賴度」等數值，經由第參章之探勘方法的計算過程，可在「關聯規則」欄位中顯示出找到的關聯規則，並在「具有異常喊價傾向」欄位中可顯示出找到的結果，如圖四。



圖四 偵測全部具有異常喊價傾向之競標者的執行畫面

圖五為點選「探勘某一拍賣商品的競標者」功能的偵測畫面，在「輸入拍賣商品」欄位中輸入欲探勘之拍賣商品項目的編號，在「競標者人數」欄位中輸入競標者的人數，然後從交易資料庫中隨機選定此數目的競標者，視之為目前對此一拍賣商品進行喊價者。經由第肆章之探勘方法的計算過程，可在「關聯規則」欄位中顯示出找到符合條件的關聯規則，並在「競標者名單」欄位中顯示目前正對此一拍賣商品進行喊價的競標者、及在「異常競標者」欄位中顯示出探勘的結果，如圖五。



圖五 偵測某一拍賣商品之競標者具有異常減價傾向的執行畫面

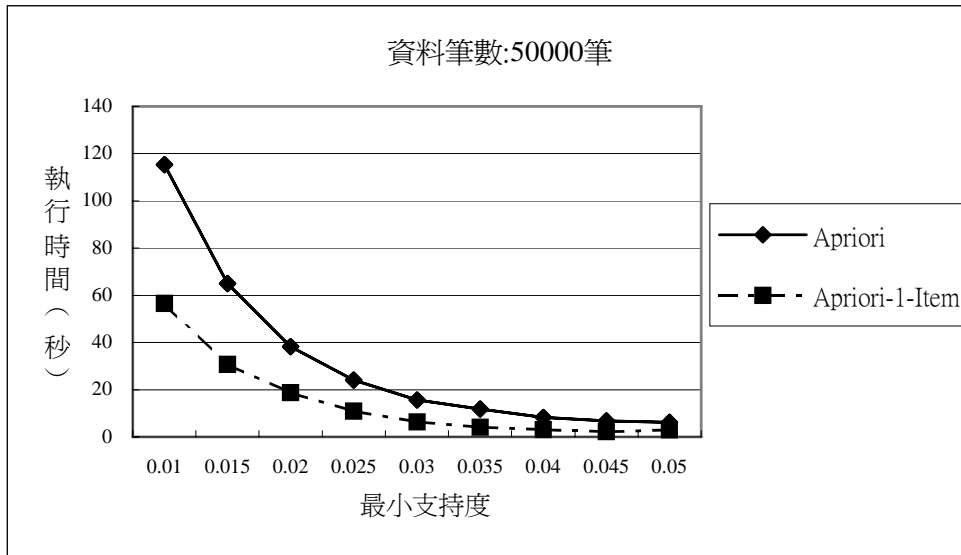
接下來，我們實驗評估前面章節所描述之演算法的執行效能，實驗平台說明如表四。

表四 效能實驗平台

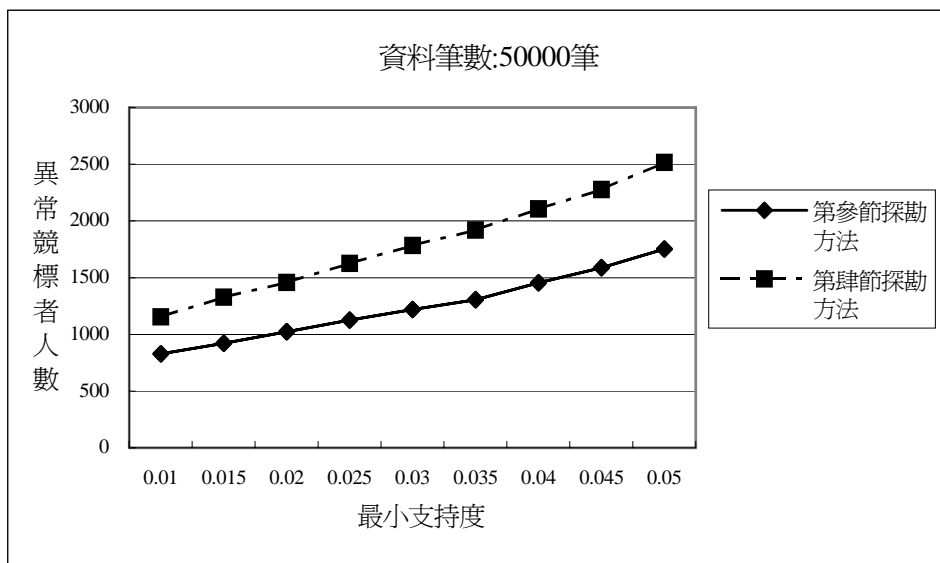
CPU	CPU-Pentium IV 2.4G Hz
Main memory	256 Mbytes
作業系統	Windows XP Professional
使用語言	C#
探勘資料來源	交易資料庫 <i>D</i>
最小購買忠誠度	20%
最小減價關聯度	35%
最小信賴度	60%

首先，在不失一般性的條件下，我們隨機以某一拍賣商品項目為探勘的評估目標，分別評估在不同最小支持度的條件下，使用 Apriori 演算法及 Apriori-1-Item 演算法的執行步驟，找到包含此一拍賣商品之關聯規則的執行

時間，其執行結果如圖六。從圖六中顯示出當最小支持度設定愈低時，Apriori-1-Item 演算法愈優於 Apriori 演算法的執行效率，找到所要的關聯規則。



圖六 Apriori 及 Apriori-1-Item 演算法探勘包含某一拍賣商品之關聯規則的執行效能

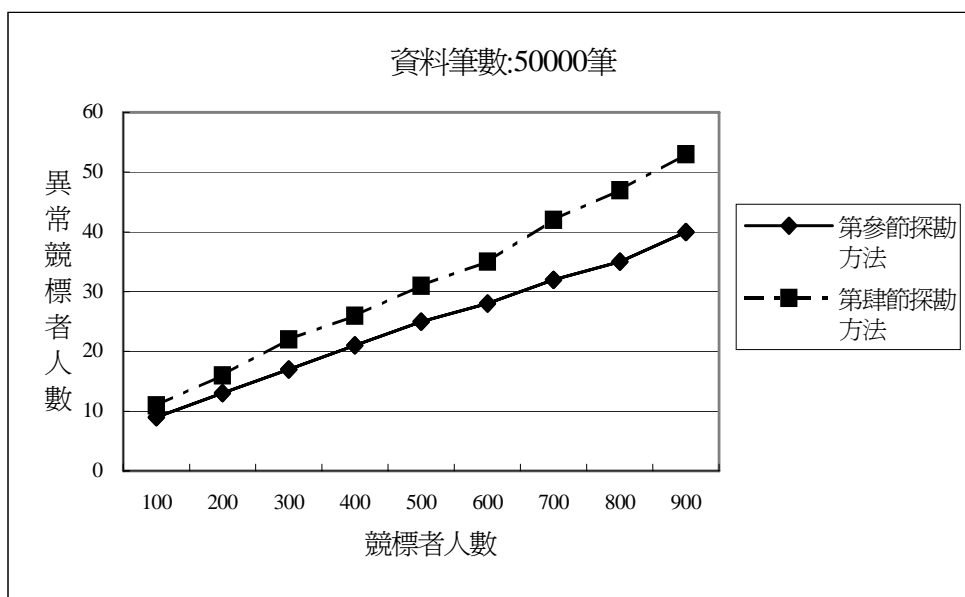


圖七 以全部競標者為偵測的對象，發掘具有異常喊價傾向的競標者

接下來，我們以 50000 筆交易資料之競標者為偵測的對象，分別利用第參節及第肆節所描述之探勘方法，評估在不同最小支持度的情況下，找出具有

異常喊價傾向的競標者。首先，我們評估利用第參節的探勘方法，其找出具有異常喊價傾向之競標者的數量，如圖七。然後，我們再以前面隨機挑選之拍賣商品項目為探勘的目標，假設以全部競標者為偵測的對象，評估利用第肆節的探勘方法，其找出對此一拍賣商品具有異常喊價傾向之競標者的數量，如圖七。從圖七中可顯示當最小支持度愈大時，由於找出的關聯規則較少，因此導致判斷出具有異常喊價傾向之競標者的人數也會較多。

我們以前面挑選之拍賣商品項目為探勘的目標，然後從交易資料庫中隨機挑選出 100, 200, 300, ..., 900 位競標者，視之為目前正在對此一拍賣商品項目進行喊價的競標者。我們以這些競標者做為偵測的對象，分別利用第參節及第肆節的探勘方法，並設定最小支持度為 0.05，找出具有異常喊價的競標者、及對此一拍賣商品具有異常喊價傾向的競標者，其結果如圖八。在第肆節的探勘方法中，由於只擷取包含此一拍賣商品項目的關聯規則，因此所找出的關聯規則數量會較少於第肆節的探勘方法。從圖八中可顯示利用第肆節的探勘方法，其判斷出具有異常喊價傾向之競標者的數量，會較多於第參節的探勘方法。



圖八 以隨機挑選之競標者為偵測的對象，發掘具有異常喊價傾向的競標者

陸・結論

在電子商務中，線上拍賣是最受消費者歡迎的交易方式之一，隨著交易規模日益擴大與熱絡，也衍生了許多的交易問題，其中賣方偽裝成其他競標者，一同參與喊價進而哄抬拍賣商品的價格，是線上拍賣詐欺重要問題之一。目前已有許多研究在探討線上拍賣所衍生出的欺騙及詐欺等問題（徐源隆，2003；葛煥元，2003；朱桓毅，2004），但鮮少研究利用探勘技術於此領域中。利用資料探勘技術於犯罪及欺騙的偵防，其有效性與實際應用性已被許多研究所證實（詹培萱，2002；徐源隆，2003；Metwally et al., 2005；Chui and Tsai, 2004；Brause et al., 1999）。

在本篇論文中，我們利用探勘技術分別從以下兩方面發掘具有異常喊價傾向的競標者：一是發掘具有異常喊價傾向之競標者；二是以某一拍賣商品為探勘的目標，找出對此一拍賣商品具有異常喊價傾向的競標者。從資料的蒐集、理論的闡述、探勘方法的設計及結果的分析，顯示出我們所提出之探勘過程有其實務應用及方法創新的學術價值。此探勘結果，對拍賣商品避免被惡意哄抬喊價，或事先預防無意購買拍賣品之競標者的喊價行為，將可提供非常有用的資訊。

參考文獻

- 朱桓毅，「拍賣網站消費與詐欺行為之研究」，銘傳大學傳播管理研究所碩士論文，2004 年。
- 美國聯邦貿易委員會，<http://www.ftc.gov/bcp/online/pubs/online/auctions.htm#fraud>。
- 徐源隆，「網路拍賣詐欺犯罪之偵查對策」，第七屆資訊管理學術暨警政資訊實務研討會，中央警察大學，2003 年，頁 89-97。
- 翁頌舜、鄭富山，「應用資料探勘探究電信資料異常之研究」，*輔仁管理評論*，第 9 卷第 3 期，2002 年，頁 77-110。
- 陳怡婷，「應用資料探勘技術於拍賣網站購買行為之研究」，實踐大學貿易經營研究所碩士論文，2004 年。
- 詹培萱，「利用文字探勘技術進行犯罪資料之發掘—以網路販售違禁品及網路賭博為例」，中央警察大學資訊管理研究所碩士論文，2002 年。
- 葛煥元，「線上拍賣欺騙行為偵測之研究」，國立中正大學資訊管理研究所碩士論文，2003 年。
- Agrawal, R., Imielinski, T. and Swami, A., "Mining Association Rules between Sets of Items in Very Large Database", *Proceedings of the ACM SIGMOD Conference on Management of Data*, 1993, pp. 207-216.

- Agrawal, R. and Srikant, R., "Fast Algorithms for Mining Association Rules in Large Database", Proceedings of the 20th International Conference on Very Large Data Bases, 1994, pp. 487-499.
- Bonchi, F., Giannotte, F., Mainetto, G. and Pedreschi, D., "A Classification-Based Methodology for Planning Audit Strategies in Fraud Detection", Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1999, pp. 175-184.
- Brause, R., Langsdorf, T., Hepp, M., "Neural Data Mining for Credit Card Fraud Detection", Proceedings of the 11th IEEE International Conference on the Tools with Artificial Intelligence, 1999, pp. 103-106.
- Butler, P. and Peppard, J., "Consumer Purchasing on the Internet: Processes and Prospects", *European Management Journal*, Vol. 16, 1998, pp. 600-610.
- Chen, M. S., Han, J. and Yu, P. S., "Data Mining: An Overview from a Database Perspective", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 8(6), 1996, pp. 866-883.
- Chui, C. C. and Tsai, C. Y., "A web Services-Based Collaborative Scheme for Credit Card Fraud Detection", Proceedings of the IEEE international conference on e-Technology, e-Commerce and e-Service, 2004.
- Han, J. and Kamber, M., "Data Mining: Concepts and Techniques", 2nd ed., Morgan Kaufmann, 2006.
- IBM, <http://www.almaden.ibm.com/software/projects/hdb/resources.shtml>.
- Kou, Y., and Lu, C. T., Sirwongwattana, S. and Huang, Y. P., "Survey of Fraud Detection Techniques", Proceedings of the 2004 IEEE International Conference on the Networking, Sensing and Control, 2004, pp. 749-754.
- Metwally, A., Agrawal, D. and Abbadi, A. E., "Using Association Rules for Detection in Web Advertising Networks", Proceedings of the 31st VLDB Conference, 2005, pp. 169-180.
- Park, J. S., Chen, M. S. and Yu, P. S., "Using a Hash-Based Method with Transaction Trimming for Mining Association Rules", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 9(5), 1997, pp. 813-825.
- Shao, H., Zhao, H. and Chang, G. H., "Applying Data Mining to Detect Fraud Behavior in Customs Declaration", Proceedings of the First International Conference on Machine Learning and Cybernetics, 2002, pp. 1241-1244.
- Srikant, R. and Agrawal, R., "Mining Generalized Association Rules", Proceedings of the 21st International Conference on Very Large Data Bases, 1995, pp. 407-419.

Using Mining Techniques to Discover Fraudulent Bidders in Online Auction

CHUI-CHENG CHEN*

ABSTRACT

The online auction is one of the most popular transaction ways in the business models of electronic commerce. Although online auction may provide participants more shopping enjoyment than traditional way, it may introduce potential risk in the transaction process. How to avoid the opportunity of fraud that allows bidders to manipulate the price of merchandise intentionally is an important issue in auction fraud. In this paper, bidders' transaction data are used as the source of mining, and each transaction data contains a bidder's transaction history and detail information. Two mining methods are presented for discovering the fraudulent bidders. The first task is to develop a mining method that contains two steps to find fraudulent bidders in transaction data. The second task is to treat one item as the target of mining, and to develop a mining method that contains two steps to find the fraudulent bidders for the item. Finally, we design and construct a detecting system to find the fraudulent bidders according to the above methods. The results of the mining can provide effective and useful mechanism to detect and avoid the price of merchandise to be manipulated intentionally.

Keywords: data mining, online auction, fraudulent bidders

* Chui-Cheng CHEN, Associate Professor, Department of Information Management, Southern Taiwan University of Technology.