

# 探勘全民健保各科別住院高醫療資源使用者之特徵

蔡秀滿

明新科技大學資訊工程學系

[pauray@must.edu.tw](mailto:pauray@must.edu.tw)

陳偉佳

明新科技大學電機工程所

[seiren0113@gmail.com](mailto:seiren0113@gmail.com)

## 摘要

近年來，健保財務收支赤字愈來愈嚴重，醫療資源使用不當是主要的原因之一。相關資料顯示，前 1% 的高使用者之醫療費用約佔全部醫療費用的 28%，因此，若能發掘高醫療資源使用者的特徵，將有助於了解健保資源使用的情形。本論文使用 C4.5 決策樹分類技術，對全民健康保險資料庫的住院資料進行探勘，發掘不同醫療機構的高醫療資源使用者之特徵，包括就醫科別、性別、年齡和住院次數及醫療費用的關係。研究結果將有助於制定更完善的醫療利用政策，以解決全民健保漏洞等問題。

關鍵詞：資料探勘、決策樹、全民健康保險、住院、醫療資源

# 探勘全民健保各科別住院高醫療資源使用者之特徵

## 壹、緒論

台灣自全民健保開辦以來，全體國民受到完善的醫療照顧，也由於健保申報制度的實施，使得許多相關醫療資料得以完整保存下來。這些在醫療資料庫中所累積的大量資料，可能隱藏許多重要的資訊，利用資料探勘的技術將可以從這些大量的歷史資料中挖掘出有價值的資訊，提供給管理人員作為決策的參考。

文獻[1]提到，在醫療資訊系統的廣泛使用下，資料庫中的資料量亦隨之快速累積，蘊藏在其中的資訊量也相對地提高，若能藉由資料探勘的技術，有效的挖掘出有用的知識，並應用在醫院的經營或醫師的治療行為上，將有助於醫院醫療成本的控制及醫療品質的提升。他們以決策樹分類方法推論症狀與疾病兩者之間的關係，讓患者可藉由症狀來瞭解其可能發生的疾病。更重要的是可藉由資訊系統的發展，提供醫師一些額外的資訊，以避免錯誤的產生並且提供快速查閱病患相關資料來減少時間的浪費。例如：以藥物間的交互作用提示來避免用藥錯誤發生，並且以檢驗和檢查報告查詢取代書面病歷的翻閱，進而提高問診的速度。文獻[2]則是應用資料探勘中的序列樣式探勘技術，從醫療紀錄中找出隱性知識，發掘不同慢性疾病之間的發生順序，一方面提供醫療專業人員進行醫療照護時，對慢性疾病的發生順序作為參考，另一方面作為慢性病患自身健康管理的依據，透過慢性疾病的發生順序與家族成員的關聯，讓自身健康管理的方式能更加落實。

文獻[3]的目的是建置一個資料探勘工具，使用全民健康保險資料庫的系統架構操作程序與模式，將全民健保資料庫帶進新的應用方式，以探索門診高就診次者的特性，不僅能夠對政策提出建議，也可以提供後續對全民健保資料庫在資料探勘相關議題上之研究參考。文獻[4]則是以接受健檢服務之高階顧客群的問卷作為分析探討的資料，進行兩部分的研究：第一部份是將客戶基本資料以分群技術進行群集分析，具有相似特性的客戶將被分在同一群，再針對後續需求進行探討；第二部分則是對分群顧客特性的分析，利用決策樹進行分類，進行滿意度預測並找出重要關鍵指標，以便能將有限的資源重新組合，投入在高階客戶需要的服務重點上。

文獻[5]是利用類神經網路建立顧客關係管理中病患流失的分析模式，以探討病患組成結構和流失特徵。文獻[6]則是利用資料探勘技術中的購物籃分析工具，考慮年齡、檢傷分級、科別、來院方式、離院後動向、系統疾病分類等變數，找出滯留急診超過24小時的病患屬性及其就醫屬性的組合。

本論文的研究目的則是使用決策樹分類技術來進行資料探勘，希望能從全民健康保險資料庫的住院資料中，針對不同醫療機構，找出不同就醫科別、年齡、性別的患者和住院天數及住院醫療費用的關聯性，以發掘住院高醫療資源使用者的特徵。

## 貳、 相關研究

近年來已有許多研究探討住院醫療資源耗用的相關問題。文獻[7]探討癌症住院患者醫療資源耗用的因素，主要對象是研究結腸直腸癌患者。研究結果發現，病患個人特質和醫院屬性是造成結腸直腸癌病患住院醫療資源耗用差異之影響因素，因此建議衛生主管機關在修訂疾病免費篩檢基準時，宜適時將病患和疾病特性納入考量，以符合醫療資源分配之公平原則。此外，對醫療院所的建議則是強調預防疾病的發生，應該對民眾宣導多做早期診斷與篩檢可控制疾病嚴重程度的重要觀念，如果已發生疾病，則要給予積極有效之治療。

文獻[8]指出過去龐大的健保支出中，大部份的住院醫療費用與住院日數長短有密切的關係，許多國外的研究指出診斷關係群的實施顯著降低美國醫療保險的住院醫療費用，而醫療費用和住院日數的關係是密不可分的，因此，若要控制醫療費用的成長，住院日數的控制將是短期可行的方案。研究結果發現，醫師隨著年資的增加，可以判定的診斷數也較多，而處置乃跟隨診斷而來，因此醫療費用也自然較高。研究發現在迴歸模式中達顯著變項者為年齡、科別、診斷數及處置數，醫務管理者應能針對顯著的變項中可控制的部份進一步深入分析，掌握更明確的資訊，以訂定住院日數的常模，提供一套模式來解釋這些特質與醫療資源耗用的關係，如此對於財務狀況日益拮据的醫療院所而言，應能發揮管理上的利用價值。

文獻[9]利用資料探勘技術探討超長天住院病人之相關屬性，從分析結果發現，「三十日以上超長住院率」從未有明顯下降的情形。為了能充份了解超長天住院病人之特質及就醫屬性，他們利用資料探勘的技術從超長天住院病人個案中，分析出超長住院潛在之相關屬性及其關聯性，突顯出隱藏其中的問題。此研究驗證了資料探勘技術與一般傳統統計分析結果相似，並且發現資料探勘工具確實能夠挖掘出潛在之特性組合，對於異常資料之稽核也有極大的效用。

文獻[10]以全民健保84年急性病床住院資料進行分析，以住院超過30日定義為超長住院，超長住院人日佔總住院人日24.4%，若能避免超過30日的住院，則估計可空出10.1%之病床，此部份費用約佔健保住院費用之13.1%。以公立私立別區分，公立醫院超長住院率為7.5%，遠較私立醫院27%高。以層級別區分，醫院層級愈高，超長住院愈嚴重，公立與私立醫學中心的超長住院率分別為8.9%與5.6%。依專科別區分，內科病床超長住院率10.4%，雖遠較復健科32.9%為低，但由於內科病患最多，故超長住院人次居最冠。由前述分析可以發現，超長住院是迫切需要被正視的問題，建議未來應朝落實出院計畫、鼓勵發展各類長期照護服務、提高住院部份負擔比例、拉大醫院層級部份負擔差距、以及提高自行負擔金額上限等方向努力。

文獻[11]指出健保局住院品質指標「30日以上超長住院率」自2004年監控以來，未能有效改善，而老人為住院醫療耗用與超長期住院中最大族群，所以研究高齡患者超長期住院醫療利用影響因子有其必要性。研究結果發現，在長期住院的高齡患者中，具有女性、重大傷病身分、住院期間有手術、私立醫院照護、呼吸系統疾病型態等因子者，其住院日數與住院醫療費較高，且達統計上的顯著差異。

本論文將使用Weka系統中的決策樹分析工具來進行資料探勘。Weka是利用Java程式語言發展而成的自由軟體，由於在Java的環境下運作，因此具備可以提供跨平台執行的功能。它是由紐西蘭Waikato大學機器學習研究團隊所開發出來的資料探勘應用軟體，主要提供機器學習演算法的分析歸納功能，並提供各種資料探勘的重要分析工具，包括資料預處理、關聯規則、分類預測等功能，而且也具備了視覺化的圖形介面。

本研究使用Weka工具中的C4.5演算法來進行資料的分類探勘。C4.5演算法是一種建立決策樹的方法，決策樹分析適用於所有的分類問題，它是使用樹狀結構來產生一些規則，以建立分類的模型。決策樹中的每一個內部節點會對應到某項屬性的測試，每一個分支代表被測試屬性的一個可能的值，而每一個葉節點則會對應到一個目標屬性的值。常見的方法包括ID3和C4.5，都是由同一位學者 Quinlan 所提出的。C4.5演算法是針對ID3演算法造成決策樹的結果傾向於選擇屬性值較多的屬性之缺點所做的修正。

## 參、研究方法

資料探勘是資料庫的知識發掘(Knowledge Discovery in Databases, 簡稱 KDD)中很重要的一個步驟。知識發掘的過程包括「資料收集」、「資料前置處理」、「資料倉儲建立」、「資料探勘」、「樣式評估」等階段。資料庫中累積的大量歷史資料可能隱藏許多有意義的資訊，資料探勘的目的就是從這些大量的歷史資料中挖掘出有價值的資訊，以作為制定管理決策時的重要參考。這一節，我們將介紹在進行資料探勘之前非常重要的兩個階段：「資料收集」和「資料前置處理」。

### 一、資料收集

本研究是以全民健康保險研究資料庫為對象<sup>1</sup>，根據2005年承保抽樣歸人檔在2009年R201承保資料片的DD檔(住院醫療費用清單明細檔)，作為進行住院高醫療資源使用者特徵分析的資料來源，原始資料筆數共有4698筆。歷年來，雖然有許多關於住院高醫療資源利用的研究報告出現，但尚未有研究探討不同醫療機構型態的使用者特徵差異。由於醫療機構的型態不同，住院高醫療資源使用者之特徵可能有所差別，若能發掘其中的差異性，將能提出更完善的配套措施來節制醫療資源浪費的情形。

### 二、資料前置處理

本研究的目的是發掘各科別住院高醫療資源使用者的特徵，我們將住院高醫療資源的利用分為兩方面進行研究：(1)分析就醫科別、年齡、性別和年度住院天數的關係；(2)分析就醫科別、年齡、性別和年度住院醫療費用的關係。下面將說明本研究資料前置處理所包含的兩個階段：資料整合和資料轉換。

#### (一) 資料整合

處理步驟如下：

- (1) 原始DD資料檔的副檔名為DAT檔，裡面資料的每一個欄位值尚未做區隔，所以我們先將DAT檔轉成excel檔案，並且利用程式分析，將每個欄位值分隔出來。

---

<sup>1</sup> 本研究資料來源為衛生署中央健康保險局提供，財團法人國家衛生研究院管理之「全民健康保險研究資料庫」。文中任何闡釋或結論並不代表衛生署中央健康保險局或財團法人國家衛生研究院之立場。

- (2) 在原始資料中，總共有70個欄位，我們先把不需要的欄位刪除，只留下必要的欄位，包括：醫事機構代號、身分證統一編號、出生年月日、就醫科別、入院年月日、出院年月日、急性病床天數、慢性病床天數、醫療費用、性別欄位。
- (3) 「醫事機構代號」的後2碼為權屬別，權屬別共分為五類醫療機構，包括：公立醫院、私立醫院、公立診所、私立診所、財團法人其他醫療機構，其中只有公立醫院、私立醫院、私立診所三類有包含住院資料。
- (4) 根據醫事機構代號的權屬別，將同一人在同一科別的多筆資料彙整成一筆，最後再進行一次欄位刪除，只留下就醫科別、年齡、性別、住院天數和住院醫療費用等欄位。

## (二) 資料轉換

我們將使用決策樹分類技術來發掘各科別住院高醫療資源使用者的特徵，雖然C4.5演算法可以處理數值型態的屬性，但是我們的研究中發現，將年齡數值先轉換為離散型態，決策樹所產生的規則集可以更精簡，而且分類正確率和年齡數值不轉換時的情況是相當的。因此，我們將數值型態的資料先轉換成非數值型態的資料。

- (1) 分析就醫科別、年齡、性別和年度住院天數的關係之前，我們根據圖1和圖2的統計資料以及平均年度住院天數，將年齡和年度住院天數分別進行轉換，如表1和表2所示。

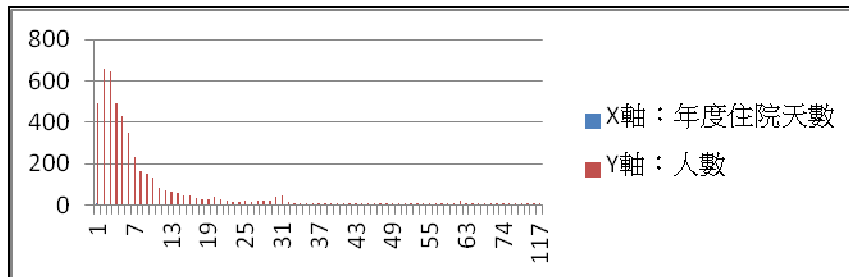


圖1 年度住院天數和人數的關係

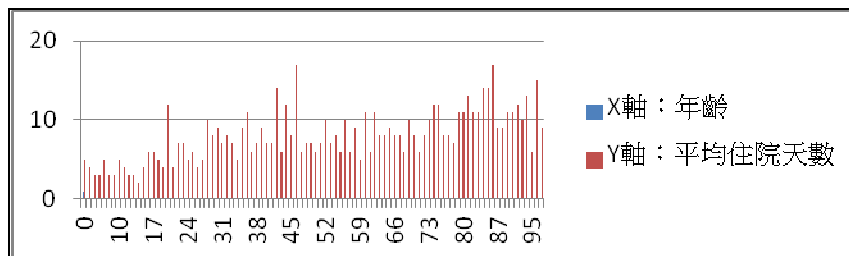


圖2 年齡對應該年齡平均年度住院天數的關係

- (2) 分析就醫科別、年齡、性別和年度住院醫療費用的關係之前，我們先根據圖3和圖4的統計資料，以及年度住院醫療費用的平均值，將年度住院醫療費用進行轉換，如表3所示，年齡的轉換方式和表2相同。

表1 住院天數轉換代碼

住院天數	轉換後代號
1~5天	D1
6~10天	D2
11~20天	D3
21~30天	D4
31天(含)以上	D5

表2 年齡轉換代碼

年齡(2009年為基準)	轉換後代號
1~15	A1
16~30	A2
31~45	A3
46~60	A4
61~75	A5
76(含)以上	A6

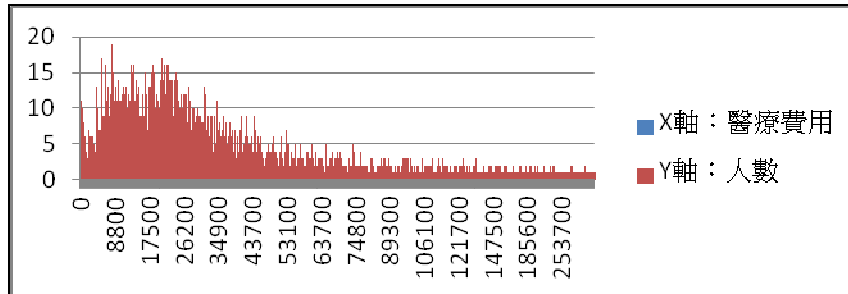


圖3 人數和年度住院醫療費用的關係

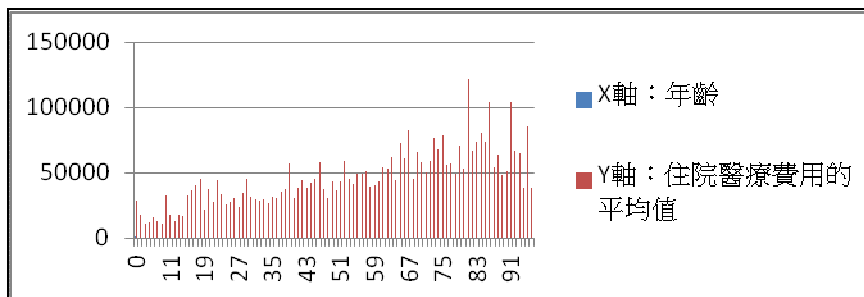


圖4 年度住院醫療費用的平均值和年齡的關係

表3 年度住院醫療費用所對應的轉換代號

醫療費用	轉換後代號
58,000(含)元以下	C1
58,001~107,500(含)元	C2
107,500元以上	C3

#### 肆、資料分析

在這一節中，我們將使用Weka的C4.5決策樹工具分析就醫科別、年齡、性別和年度住院天數的關係，以及就醫科別、年齡、性別和年度住院醫療費用的關係。我們先對「不分醫療機構」的所有資料進行分析，接著分別考慮三大類醫療機構，包括「公立醫

院」、「私立醫院」和「私立診所」，從中探討「不分醫療機構」以及三類不同醫療機構的高資源使用者特徵是否有所差異。

#### 一、分析就醫科別、年齡、性別和年度住院天數的關係

在進行探勘時，我們發現資料中，年度住院天數D1所占的資料比例極高，造成所產生的決策樹只有單一個節點。在刪除D1之後所進行的探勘結果，因D2所占的資料比例相對較高，所產生的決策樹仍只有單一個節點。在就醫科別的部分，我們考慮資料筆數的代表性，將就醫科別筆數小於10者先進行刪除。根據統計資料，各科別的年度平均住院天數為9.24天，由於我們所考慮的對象是高資源使用者，因此我們將D1和D2資料在探勘前先進行刪除，只考慮年度住院天數10天以上者。

##### (1) 不分醫療機構

決策樹分類正確率為60%，根據決策樹模型，我們可以得到下列二十一條規則：

- 1 規則一：若就醫科別為家醫科，則住院天數為11~20天。(正確率=56.2%)
- 1 規則二：若就醫科別為內科且性別為男性，則住院天數為11~20天。(正確率=68%)
- 1 規則三：若就醫科別為內科、性別為女性且年齡為31~60歲，則住院天數為21-30天。(正確率=54.5%)
- 1 規則四：若就醫科別為內科、性別為女性且年齡為61-75歲，則住院天數為11~20天。(正確率=70%)
- 1 規則五：若就醫科別為內科、性別為女性且年齡為76歲(含)以上，則住院天數為21-30天。(正確率=46.6%)
- 1 規則六：若就醫科別為外科，則住院天數為11~20天。(正確率=69.3%)
- 1 規則七：若就醫科別為骨科，則住院天數為11~20天。(正確率=82.6%)
- 1 規則八：若就醫科別為神經外科，則住院天數為11~20天。(正確率=75.5%)
- 1 規則九：若就醫科別為神經科，則住院天數為11~20天。(正確率=64.7%)
- 1 規則十：若就醫科別為精神科，則住院天數為31天(含)以上。(正確率=56.6%)
- 1 規則十一：若就醫科別為復健科，則住院天數為21-30天。(正確率=73.9%)
- 1 規則十二：若就醫科別為整型外科，則住院天數為11~20天。(正確率=68.7%)
- 1 規則十三：若就醫科別為腸胃內科，則住院天數為11~20天。(正確率=69.8%)
- 1 規則十四：若就醫科別為心臟血管內科，則住院天數為11~20天。(正確率=71.4%)
- 1 規則十五：若就醫科別為胸腔內科，則住院天數為11~20天。(正確率=49.6%)
- 1 規則十六：若就醫科別為腎臟內科，則住院天數為11~20天。(正確率=65.7%)
- 1 規則十七：若就醫科別為血液腫瘤科，則住院天數為11~20天。(正確率=47.3%)
- 1 規則十八：若就醫科別為內分泌科且性別為男性，則住院天數為11~20天。(正確率=100%)
- 1 規則十九：若就醫科別為內分泌科且性別為女性，則住院天數為21-30天。(正確率=50%)
- 1 規則二十：若就醫科別為感染科，則住院天數為11~20天。(正確率=47.8%)
- 1 規則二十一：若就醫科別為直腸外科，則住院天數為11~20天。(正確率=83.3%)

##### (2) 公立醫院

決策樹分類正確率為54.4%，根據決策樹模型，我們可以得到下列十九條規則：

- 1 規則一：若就醫科別為內科且性別為女性，則住院天數為21-30天。(正確率=58.3%)
- 1 規則二：若就醫科別為內科且性別為男性，則住院天數為11-20天。(正確率

=75%)

- 1 規則三：若就醫科別為外科，則住院天數為11-20天。(正確率=68.1%)
  - 1 規則四：若就醫科別為骨科，則住院天數為11-20天。(正確率=66.6%)
  - 1 規則五：若就醫科別為神經外科，則住院天數為11-20天。(正確率=68.4%)
  - 1 規則六：若就醫科別為神經科，則住院天數為11-20天。(正確率=71.4%)
  - 1 規則七：若就醫科別為精神科且年齡為16-30歲，則住院天數為31天(含)以上。(正確率=45.8%)
  - 1 規則八：若就醫科別為精神科且年齡為31-45歲、性別為女性，則住院天數為11-20天。(正確率=64.7%)
  - 1 規則九：若就醫科別為精神科、年齡為31-45歲且性別為男性，則住院天數為31天(含)以上。(正確率=53.4%)
  - 1 規則十：若就醫科別為精神科且年齡為46-60歲，則住院天數為31天(含)以上。(正確率=66.6%)
  - 1 規則十一：若就醫科別為精神科、年齡為61-75歲且性別為女性，則住院天數為21-30天。(正確率=66.6%)
  - 1 規則十二：若就醫科別為精神科、年齡為61-75歲且性別為男性，則住院天數為31天(含)以上。(正確率=80%)
  - 1 規則十三：若就醫科別為精神科且年齡為76歲(含)以上，則住院天數為31天(含)以上。(正確率=60%)
  - 1 規則十四：若就醫科別為腸胃內科，則住院天數為11-20天。(正確率=71.4%)
  - 1 規則十五：若就醫科別為胸腔內科，則住院天數為11-20天。(正確率=48.9%)
  - 1 規則十六：若就醫科別為腎臟內科，則住院天數為11-20天。(正確率=76.9%)
  - 1 規則十七：若就醫科別為血液腫瘤科，則住院天數為11-20天。(正確率=46.6%)
  - 1 規則十八：若就醫科別為感染科且年齡為46-60歲，則住院天數為11-20天。(正確率=100%)
  - 1 規則十九：若就醫科別為感染科且年齡為61歲(含)以上，則住院天數為21-30天。(正確率=50%)
- (3) 私立醫院

決策樹分類正確率為65%，根據決策樹模型，我們可以得到下列十九條規則：

- 1 規則一：若就醫科別為家醫科，則住院天數為11-20天。(正確率=63.6%)
- 1 規則二：若就醫科別為內科，則住院天數為11-20天。(正確率=63.8%)
- 1 規則三：若就醫科別為外科，則住院天數為11-20天。(正確率=70.3%)
- 1 規則四：若就醫科別為骨科，則住院天數為11-20天。(正確率=89.1%)
- 1 規則五：若就醫科別為神經外科，則住院天數為11-20天。(正確率=80%)
- 1 規則六：若就醫科別為神經科，則住院天數為11-20天。(正確率=60%)
- 1 規則七：若就醫科別為精神科，則住院天數為31天(含)以上。(正確率=67.7%)
- 1 規則八：若就醫科別為復健科，則住院天數為21-30天。(正確率=76.4%)
- 1 規則九：若就醫科別為整型外科，則住院天數為11-20天。(正確率=72.7%)
- 1 規則十：若就醫科別為腸胃內科，則住院天數為11-20天。(正確率=68.7%)
- 1 規則十一：若就醫科別為心臟血管內科，則住院天數為11-20天。(正確率=84.6%)
- 1 規則十二：若就醫科別為胸腔內科，則住院天數為11-20天。(正確率=50%)
- 1 規則十三：若就醫科別為腎臟內科且年齡為16-30歲，則住院天數為21-30天。(正確率=100%)
- 1 規則十四：若就醫科別為腎臟內科且年齡為31-75歲，則住院天數為11-20天。(正確率=75%)
- 1 規則十五：若就醫科別為腎臟內科、年齡為76歲(含)以上且性別為女性，則住



院天數為21-30天。(正確率=83.3%)

- 1 規則十六：若就醫科別為腎臟內科、年齡為76歲(含)以上且性別為男性，則住院天數為11-20天。(正確率=66.6%)
- 1 規則十七：若就醫科別為血液腫瘤科，則住院天數為11-20天。(正確率=47.8%)
- 1 規則十八：若就醫科別為內分泌科且性別為女性，則住院天數為21-30天。(正確率=50%)
- 1 規則十九：若就醫科別為內分泌科且性別為男性，則住院天數為11-20天。(正確率=100%)

#### (4) 私立診所

決策樹分類正確率為84.6%，根據決策樹模型，我們發現在私立診所部分，年度住院天數1-5天占大部分，所以只有一個節點。

雖然C4.5演算法可以直接處理數值型態的資料，但是分析結果發現，將年齡屬性先進行離散化，決策樹的分類正確率平均情況和直接使用數值型年齡是相當的，並且可以得到更精簡的規則集，如表4所示。

表4 年齡型態在分析「年度住院天數」時的影響

醫療機構類型	分類正確率(使用數值型年齡)	分類正確率(使用離散型年齡)	規則數(使用數值型年齡)	規則數(使用離散型年齡)
不分醫療機構	57.5%	60.1%	41	21
公立醫院	58.2%	54.5%	26	19
私立醫院	64.5%	65.2%	30	19

## 二、分析就醫科別、年齡、性別和年度住院醫療費用的關係

在進行探勘時，我們發現資料中年度住院天數C1所占的比例極高，造成所產生的決策樹只有單一個節點。由於我們所考慮的對象是高醫療資源使用者，且各科別的平均住院醫療費用為58,000元，因此我們將C1資料在探勘前先進行刪除，只考慮年度住院醫療費用在58,000元以上者，並且考慮資料筆數的代表性，先將就醫科別筆數小於10者進行刪除。

### (1) 不分醫療機構

決策樹分類正確率為58.7%，根據決策樹模型，我們可以得到下列三十七條規則：

- 1 規則一：若就醫科別為內科且年齡1-60歲，則住院醫療費用為58,001~107,500(含)元。(正確率=74%)
- 1 規則二：若就醫科別為內科且年齡61-75歲、性別為女性，則住院醫療費用為107,500元以上。(正確率=62.5%)
- 1 規則三：若就醫科別為內科且年齡61-75歲、性別為男性，則住院醫療費用為58,001~107,500(含)元。(正確率=100%)
- 1 規則四：若就醫科別為內科且年齡76歲(含)以上，則住院醫療費用為107,500元以上。(正確率=69.5%)
- 1 規則五：若就醫科別為外科且年齡1-30歲，則住院醫療費用為58,001~107,500(含)元。(正確率=100%)
- 1 規則六：若就醫科別為外科且年齡31-45歲、性別為女性，則住院醫療費用為58,001~107,500(含)元。(正確率=66.6%)

- 1 規則七：若就醫科別為外科且年齡31-45歲、性別為男性，則住院醫療費用為107,500元以上。(正確率=60%)
- 1 規則八：若就醫科別為外科且年齡46-60歲，則住院醫療費用為58,001~107,500(含)元。(正確率=50%)
- 1 規則九：若就醫科別為外科且年齡61歲(含)以上，則住院醫療費用為107,500元以上。(正確率=61.9%)
- 1 規則十：若就醫科別為小兒科，則住院醫療費用為58,001~107,500(含)元。(正確率=60%)
- 1 規則十一：若就醫科別為婦產科，則住院醫療費用為58,001~107,500(含)元。(正確率=66.6%)
- 1 規則十二：若就醫科別為骨科且年齡1-45歲，則住院醫療費用為58,001~107,500(含)元。(正確率=100%)
- 1 規則十三：若就醫科別為骨科且年齡46-60歲，則住院醫療費用為107,500元以上。(正確率=52.6%)
- 1 規則十四：若就醫科別為骨科且年齡61-75歲、性別為女性，則住院醫療費用為107,500元以上。(正確率=61.2%)
- 1 規則十五：若就醫科別為骨科且年齡61-75歲、性別為男性，則住院醫療費用為58,001~107,500(含)元。(正確率=58.3%)
- 1 規則十六：若就醫科別為骨科且年齡76歲(含)以上，則住院醫療費用為58,001~107,500(含)元。(正確率=50%)
- 1 規則十七：若就醫科別為神經外科，則住院醫療費用為107,500元以上。(正確率=60%)
- 1 規則十八：若就醫科別為泌尿科，則住院醫療費用為58,001~107,500(含)元。(正確率=70.5%)
- 1 規則十九：若就醫科別為耳鼻喉科，則住院醫療費用為58,001~107,500(含)元。(正確率=84.6%)
- 1 規則二十：若就醫科別為神經科，則住院醫療費用為107,500元以上。(正確率=69.5%)
- 1 規則二十一：若就醫科別為精神科，則住院醫療費用為58,001~107,500(含)元。(正確率=73.2%)
- 1 規則二十二：若就醫科別為復健科，則住院醫療費用為58,001~107,500(含)元。(正確率=70%)
- 1 規則二十三：若就醫科別為整型外科，則住院醫療費用為58,001~107,500(含)元。(正確率=60%)
- 1 規則二十四：若就醫科別為腸胃內科，則住院醫療費用為58,001~107,500(含)元。(正確率=55.3%)
- 1 規則二十五：若就醫科別為心臟血管內科，則住院醫療費用為107,500元以上。(正確率=63.3%)
- 1 規則二十六：若就醫科別為胸腔內科，則住院醫療費用為107,500元以上。(正確率=64.2%)
- 1 規則二十七：若就醫科別為腎臟內科，則住院醫療費用為58,001~107,500(含)元。(正確率=56.4%)
- 1 規則二十八：若就醫科別為血液腫瘤科，則住院醫療費用為58,001~107,500(含)元。(正確率=61.2%)
- 1 規則二十九：若就醫科別為內分泌科，則住院醫療費用為58,001~107,500(含)元。(正確率=80%)

- 1 規則三十：若就醫科別為感染科且年齡16-30歲，則住院醫療費用為58,001~107,500(含)元。(正確率=100%)
  - 1 規則三十一：若就醫科別為感染科且年齡31-60歲，則住院醫療費用為58,001~107,500(含)元。(正確率=100%)
  - 1 規則三十二：若就醫科別為感染科且年齡61歲(含)以上，則住院醫療費用為107,500元以上。(正確率=83.3%)
  - 1 規則三十三：若就醫科別為直腸外科，則住院醫療費用為58,001~107,500(含)元。(正確率=65%)
  - 1 規則三十四：若就醫科別為心臟血管外科，則住院醫療費用為107,500元以上。(正確率=69.2%)
  - 1 規則三十五：若就醫科別為胸腔外科，則住院醫療費用為58,001~107,500(含)元。(正確率=60%)
  - 1 規則三十六：若就醫科別為消化外科且性別為女性，則住院醫療費用為58,001~107,500(含)元。(正確率=100%)
  - 1 規則三十七：若就醫科別為消化外科且性別為男性，則住院醫療費用為107,500元以上。(正確率=75%)
- (2) 公立醫院

決策樹分類正確率為60.7%，根據決策樹模型，我們可以得到下列十七條規則：

- 1 規則一：若就醫科別為內科且年齡1-60歲，則住院醫療費用為58,001~107,500(含)元。(正確率=87.5%)
- 1 規則二：若就醫科別為內科且年齡61歲(含)以上，則住院醫療費用為107,500元以上。(正確率=71.4%)
- 1 規則三：若就醫科別為外科且年齡1-45歲，則住院醫療費用為58,001~107,500(含)元。(正確率=100%)
- 1 規則四：若就醫科別為外科且年齡46-60歲，則住院醫療費用為107,500元以上。(正確率=62.5%)
- 1 規則五：若就醫科別為外科且年齡61-75歲，則住院醫療費用為58,001~107,500(含)元。(正確率=50%)
- 1 規則六：若就醫科別為外科且年齡76歲(含)以上，則住院醫療費用為107,500元以上。(正確率=70%)
- 1 規則七：若就醫科別為骨科且年齡1-45歲，則住院醫療費用為58,001~107,500(含)元。(正確率=100%)
- 1 規則八：若就醫科別為骨科且年齡46-60歲，則住院醫療費用為107,500元以上。(正確率=66.6%)
- 1 規則九：若就醫科別為骨科且年齡61歲(含)以上，則住院醫療費用為107,500元以上。(正確率=58.6%)
- 1 規則十：若就醫科別為神經外科，則住院醫療費用為107,500元以上。(正確率=75%)
- 1 規則十一：若就醫科別為精神科，則住院醫療費用為58,001~107,500(含)元。(正確率=76.5%)
- 1 規則十二：若就醫科別為腸胃內科，則住院醫療費用為58,001~107,500(含)元。(正確率=57.1%)
- 1 規則十三：若就醫科別為心臟血管內科，則住院醫療費用為107,500元以上。(正確率=68%)
- 1 規則十四：若就醫科別為胸腔內科，則住院醫療費用為107,500元以上。(正確率=60%)

- 1 規則十五：若就醫科別為腎臟內科，則住院醫療費用為58,001~107,500(含)元。  
(正確率=85.7%)
  - 1 規則十六：若就醫科別為血液腫瘤科，則住院醫療費用為58,001~107,500(含)元。  
(正確率=54.5%)
  - 1 規則十七：若就醫科別為感染科，則住院醫療費用為107,500元以上。(正確率=72.7%)
- (3) 私立醫院
- 決策樹分類正確率為54.6%，根據決策樹模型，我們可以得到下列二十四條規則：
- 1 規則一：若就醫科別為內科，則住院醫療費用為58,001~107,500(含)元。(正確率=58%)
  - 1 規則二：若就醫科別為外科且年齡為1-15歲，則住院醫療費用為107,500元以上。(正確率=100%)
  - 1 規則三：若就醫科別為外科且年齡為16-30歲，則住院醫療費用為58,001~107,500(含)元。(正確率=100%)
  - 1 規則四：若就醫科別為外科且年齡為31-45歲，則住院醫療費用為107,500元以上。(正確率=66.6%)
  - 1 規則五：若就醫科別為外科且年齡為46-60歲，則住院醫療費用為58,001~107,500(含)元。(正確率=75%)
  - 1 規則六：若就醫科別為外科且年齡為61-75歲，則住院醫療費用為107,500元以上。(正確率=80%)
  - 1 規則七：若就醫科別為外科且年齡為76歲(含)以上，則住院醫療費用為58,001~107,500(含)元。(正確率=66.6%)
  - 1 規則八：若就醫科別為骨科，則住院醫療費用為58,001~107,500(含)元。(正確率=53.3%)
  - 1 規則九：若就醫科別為神經外科，則住院醫療費用為107,500元以上。(正確率=52.9%)
  - 1 規則十：若就醫科別為泌尿科，則住院醫療費用為58,001~107,500(含)元。(正確率=84.6%)
  - 1 規則十一：若就醫科別為耳鼻喉科，則住院醫療費用為58,001~107,500(含)元。  
(正確率=80%)
  - 1 規則十二：若就醫科別為神經科且年齡為16-60歲，則住院醫療費用為58,001~107,500(含)元。(正確率=100%)
  - 1 規則十三：若就醫科別為神經科且年齡為61歲(含)以上，則住院醫療費用為107,500元以上。(正確率=75%)
  - 1 規則十四：若就醫科別為精神科，則住院醫療費用為58,001~107,500(含)元。(正確率=66.6%)
  - 1 規則十五：若就醫科別為復健科，則住院醫療費用為58,001~107,500(含)元。(正確率=68.7%)
  - 1 規則十六：若就醫科別為整型外科，則住院醫療費用為58,001~107,500(含)元。  
(正確率=61.5%)
  - 1 規則十七：若就醫科別為腸胃內科，則住院醫療費用為58,001~107,500(含)元。  
(正確率=53.8%)
  - 1 規則十八：若就醫科別為心臟血管內科，則住院醫療費用為107,500元以上。(正確率=61.8%)
  - 1 規則十九：若就醫科別為胸腔內科，則住院醫療費用為107,500元以上。(正確率=61.8%)

率=66.2%)

- 1 規則二十：若就醫科別為腎臟內科，則住院醫療費用為107,500元以上。(正確率=60%)
- 1 規則二十一：若就醫科別為血液腫瘤科，則住院醫療費用為58,001~107,500(含)元。(正確率=66.6%)
- 1 規則二十二：若就醫科別為直腸外科，則住院醫療費用為58,001~107,500(含)元。(正確率=63.6%)
- 1 規則二十三：若就醫科別為心臟血管外科，則住院醫療費用為107,500元以上。(正確率=75%)
- 1 規則二十四：若就醫科別為胸腔外科，則住院醫療費用為58,001~107,500(含)元。(正確率=50%)

#### (4) 私立診所

決策樹分類正確率為100%，根據決策樹模型，我們發現在私立診所部分，年度住院醫療費用全為58,000(含)元以下，所以只有一個節點。

在分析「年度住院醫療費用」時我們也發現，將年齡屬性先進行離散化，決策樹的分類正確率平均情況和直接使用數值型年齡是相當的，並且可以得到更精簡的規則集，如表5所示。

表5 年齡型態在分析「年度住院醫療費用」時的影響

醫療機構類型	分類正確率(使用數值型年齡)	分類正確率(使用離散型年齡)	規則數(使用數值型年齡)	規則數(使用離散型年齡)
不分醫療機構	62.4%	58.8%	41	38
公立醫院	62.3%	60.7%	23	17
私立醫院	55.1%	57.2%	25	25

## 伍、結論

本研究是根據2005年承保抽樣歸人檔在2009年R201承保資料片，對住院高醫療資源使用者進行探勘。我們先在不分醫療機構的情況下進行探勘，然後根據醫療機構代號，將有住院資料的醫療機構分為三類，包括「公立醫院」、「私立醫院」和「私立診所」，從中探討「不分醫療機構」以及三類不同醫療機構的高資源使用者特徵是否有所差異。我們使用Weka的C4.5決策樹分類法，分別分析就醫科別和病患特性，與住院天數和住院醫療費用的關係，這些研究成果，可以提供健保局在制定醫療資源分配之相關政策時，有更多的參考資料，以提升病患就醫的權益和醫療照護的品質。根據現階段的研究成果，我們將持續進行下列更深入的研究。

- 1 探討「住院天數」30天以上和「住院醫療費用」107,500元以上的病患之間是否有關聯。
- 1 增加2005年承保抽樣歸人檔在2009年R202的承保資料，比較探勘結果是否有所差異，以提升研究結果的可信度和準確率。

## 參考文獻

1. 陳俊龍、龔榮源，民 95，「以資料探勘技術建立病患就醫導引—以胃腸科病患為例」，醫療資訊雜誌，第十五卷第一期，台灣醫學資訊學會，17~34。
2. 林宇健，民 97，資料探勘技術應用於慢性疾病健康照護管理系統，靜宜大學資訊管理學系碩士論文。
3. 林明昌，民 95，利用資料探勘技術探討中央健康保險局中區分局資料庫門診高就診次者之特性，國立中正大學資訊管理研究所碩士論文。
4. 王建智、徐正容、陳銘樹，民 97，「應用資料探勘方法中的分群分析技術來探究高階健檢客戶之行態組成以進行客戶關係管理—以台北某醫學中心之高階健檢客戶為例」，健康管理學刊，第6卷，第一期，49~59。
5. 范牧蘭、楊瓊珠、劉崇祥、趙崇翎，2008，「基層醫療診所病患流失之預警—倒傳遞神經網路之運用」，醫管期刊，286~298。
6. 周歆凱、翁林仲、張怡秋、黃興進、蔡明足、蘇喜，2007，「運用購物籃分析技術探討滯留急診超過24小時病患特性」，醫管期刊，216~230。
7. 黃倩盈，民 98，影響結腸直腸癌住院患者醫療資源耗用之因素分析，義守大學醫務管理學系研究所碩士論文。
8. 王靜怡，民 94，「影響病人住院日數之因素分析」--醫師檔案分析之應用，國立陽明大學醫務管理研究所碩士論文。
9. 詹前隆、鄒淑英，民 95，利用資料探勘技術探討超長天住院病人之相關屬性，元智大學資訊管理所
10. 周歆凱，2004，利用資料探勘技術探討急診高資源耗用者之特性，國立台灣大學醫療機構管理研究所碩士論文。
11. 王勝廣、方文輝、江昇達、周稚傑、高東煒、高森永、張耀文，2011，高齡患者長期住院醫療利用影響因子之探討，台灣老年醫學暨老年學雜誌，86~104。

## Mining Characteristics of High Medical Utilization Inpatients among Different Divisions

Pauray S.M. Tsai  
Department of Computer Science and Information Engineering,  
Minghsin University of Science and Technology  
pauray@must.edu.tw

Wei-Jia Chen  
Institute of Electrical Engineering,  
Minghsin University of Science and Technology  
seiren0113@gmail.com

#### Abstract

Related information shows that the medical expenditure of the top 1% high utilization patients occupies 28% of the total medical expenditure. In recent years, due to the inadequate usage of medical resources, the deficit of national health insurance increases severely. In this paper, we use C4.5 decision tree technique to analyze data on the national health insurance research database. The relationships of division, gender and age with the length of stay and inpatient expenditure are analyzed, respectively. The characteristics of high medical utilization inpatients among different medical institutions are discovered, which will be useful for establishing adequate medical policies.

Keywords: data mining, decision tree, national health insurance, inpatient service, medical utilization