

以ID3演算法進行學生休退學規則之探勘 — 以南部某科大為例

陳廷斌

崑山科技大學資訊管理系

teddy@mail.ksu.edu.tw

高淑珍

崑山科技大學資訊管理系

kaosc@mail.ksu.edu.tw

摘要

台灣大學之普遍林立以及少子化的趨勢，再加上逐年增高的休退學率，使得學校經營面臨危機並衍生教育資源浪費以及社會問題，因此休退學之預防已成為逐漸被重視的課題，有鑑於資料探勘（Data Mining）的技術逐漸發展，倘若能將此技術應用於休退學學生的預測問題上，將可幫助學校行政人員洞悉先機以降低休退學的發生率。因此本研究針對學生入學之部份人口統計變數，以ID3（Iterative Dichotomiser 3）演算法，產生可能辦理休退學之學生預測規則231筆，而所產生的預測規則經過驗證可以達到95.37%的預測準確率。因此，研究中所產生的休退學學生之預測規則確實能夠在未來提供予學校單位參考，以期能收洞悉先機、防範及減少休退學之效。

關鍵詞：休退學預測、資料探勘、ID3演算法

壹、緒論

一、 研究背景及動機

隨著教育的普及，學子可經由許多不同的管道來選擇自己心目中理想的學校，但是大學的普遍林立以及台灣少子化的趨勢（圖 1），使得許多學校開始面臨經營的危機。然而，近年來學生就學期間辦理休退學的人數有逐年增加的趨勢，經教育部統計，大專校院於94學年度的休退學率為4.23%，然至95學年度已攀升至4.65%，與90年相較之下，休退學人數已成長了1.2倍，而97學年度又比96學年度增加了8600人，增加比率達5.38%。以南部某私立科大為例，根據該校統計資料顯示該科大98學年度全校總休退學人數為632人，然至99學年度已達710人，增加比率高達12.34%。



資料來源：私校第四屆第3次理監事聯席會資料，2008

圖 1 97至114年18歲學齡人口推估圖

就學校經營者而言，學生的休退學不但容易造成教育資源的浪費，更有可能因為成本過高導致學校陷入難以經營的窘境，此外，休退學的學生因為與學校生活脫節，再加上終日無所事事，進而沉淪網咖等不良場所，乃至結夥搶劫、吸毒、加入幫派、作奸犯科，造成社會動盪，影響社會層面甚廣。以休退學學生與犯罪行為之關連性為主，楊士隆（2001）曾就幫派青少年進行調查，結果顯示其中有高達76%曾經辦理休退學。黃明鎮（2009）也曾對54個犯罪者進行調查，其中有22名曾經為休退學的學生，佔了總比率的41%。因此學生的休退學比率逐年升高，不但是社會的隱憂，近年來也成為學校單位重視的議題。

1.2 研究目的

然如研究背景及動機所述，針對學生休退學之預測實有其必要性，目前有關休退學學生相關研究仍然無法達到很好的預測效果，例如：詹益敏（2008）針對學生的「在學資料」及「學籍資料」來預測學生辦理休退學，其中「在學資料」即是指學生入學後可變動的資料，如選課、操行、成績及缺曠，而「學籍資料」是較不會隨時間變動的資料，如學號、系別、生日、性別、身分證等，然因「在學資料」變數變異性大，因此研究仍然無法有效地針對高危險群學生提出預警。此外，章勝傑與周慧婷（2008）以中輟量表

來進行休退學生的預測，此法需時過長，耗費的成本過高，因而也使得預測的可行性大為降低。

另一方面，近年來資料庫知識發掘技術(Knowledge Discovery in Database, KDD)以及資料探勘(Data Mining)技術被廣泛使用，因此本研究採用ID3演算法(Iterative Dichotomiser 3)，利用資料探勘的技術針對學生的人口統計變數，預測可能辦理休退學之高危險群學生的預測模型。希望所預測之結果能給予學校機關以為參考，事先預防以減少學生辦理休退學的人數、減少教育資源之浪費及降低社會問題的發生。

貳、文獻探討

2.1 學生休退學預測之相關研究

近年來有部份研究開始針對休退學預測進行探討與改善，所謂預測是防範於未然的一種方法，于宗先(1972)認為預測(Forecasting)乃是人們對未被觀察到事實的一種說明；Makridakis(1986)則認為預測是人們為了想了解未來尚無法確定之事件而進行的一種活動；John(1989)則指出預測是藉由已發生過之歷史資料，來推斷未來不確定之事；Martino(1994)認為預測是指為了未來所要關注的事件而提早進行的計算或估計。因此經由彙整以上學者的論述，本研究認為所謂的預測就是藉著收集歷史的資料並利用某些數學演算法來預計未來可能發生之結果的一門藝術。學生休退學預測因其重要性，故目前已有部份研究針對此議題加以探討，章勝傑與周慧婷(2008)以中輟量表來進行休退學生的預測，研究中針對台東市學區內12所學校導師選定的1231名學生進行預測，之後再針對教師所預測的結果再由14所學校的30名有擔任導師一年半以上經驗的導師，加以檢驗其信度，此法須投入過多時間，然卻可能因近幾年環境的變更，而徒勞無功，難以達成預期之效果。因此休退學的預測雖然逐漸受到重視，但研究中所使用的方法為了能夠正確預測學生休退學的發生機率，所花費的時間過長，耗費的資金也較高，而且容易因為抽樣對象的不同而導致預測的錯誤。因此針對休退學生的預測確實存在改善的空間，例如：減少預測時間的耗費、更正確地分析出屬於休退學之高危險群學生等等，便於決策者擬定相關策略以因應。

2.2 資料庫知識發掘與資料探勘

2.2.1 定義

資料探勘意指如何於極大數量之資料中，抽取潛藏值得探索之有用資訊，以便於決策者制定決策之依據。所謂資料庫知識發掘(Knowledge Discovery in Database, KDD)，Fayyad(1996)曾將其定義為：從既有資料中探勘其為有益的、奇特的、潛在有用的、並將其轉化成淺而易懂的知識之過程。換言之，即是於資料庫裡探求對決策者有價值的資訊，而將其轉化為決策者易懂的知識之過程。此外，Fayyad(1996)也提出資料庫知識發掘之流程(圖2)是經由目標選取、前置作業、資料轉換、資料探勘以及資料編譯，從資料庫中擷取資料再轉換知識。

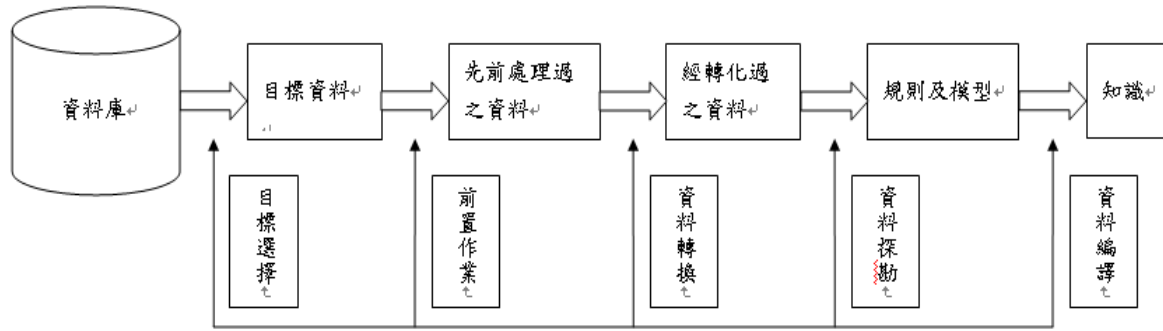


圖2 資料庫知識發掘之流程

資料來源：Fayyad et al., 1996

「資料探勘」(Data Mining) 在KDD的技術中扮演著核心的角色，Berry and Linoff (1997) 將「資料探勘」定義為：為了取得有用的法則或模組，而於龐大資料中以自動或半自動的方式來探勘資料。Kleissner (1998) 則將「資料探勘」定義為一種持續循環之決策分析過程，以及於雜亂無章的混合資料中，探勘出有用資訊之知識，而其結果可供企業專業人士之決策參考，簡而言之，「資料探勘」就是將資料庫中潛藏有用的資訊加以萃取以供決策者制定決策之依據。

2.2.3 資料探勘技術

資料探勘技術包括：分類 (classification)、分群 (clustering)、關聯法則 (association rules)、預測 (predication)、次序相關分析 (sequential pattern analysis)、趨勢分析 (trend analysis)、時間序列分析 (time serial analysis) 或更多層面的技術 (Berry and Linoff 1997; Fayyad 1996; 簡禎富 2005)，這些技術探勘之結果，可供決策者制定決策之參考。

(1) 分類 (classification)：由已知之資料中，經由決策者所製定之屬性條件進行分類。例如購買筆電時，根據消費者歷史資料來分類，當價格2萬時，將消費者的購買意願分為「偏高」或「偏低」，再經由產品其他屬性，如此即可了解影響消費者購買意願之屬性強弱。

(2) 分群 (clustering)：從未經分類的資料中選取相類似之資料為一群組，使資料彼此間有較大之共通性，例如購買筆店時，當價格於15000-20000、20001-25000、25001-30000時，筆電之銷售數量分群。可得各價位時間，筆電之銷售率。

(3) 關聯法則 (association rules)：指一屬性發生時另一屬性經常伴隨發生，例如購買筆電時通常會購買隨身碟，決策者可經由消費者此一消費習性，將隨身碟放置於筆電較近或顯而易見處以增加銷售數量。

(4) 預測 (predication)：它是經由已發生之歷史性資料來預測未來可能發生之結果，例如從歷史性資料中發現台機電近二年股票價格的分佈，都落在77-79元之間，而去預測台機電在6個月後某日價格會在78-82元之間。

(5) 次序相關分析 (sequential pattern analysis)：從一群有次序之交易資料中，尋找經常有次序性重覆出現的交易項目之組合，用以瞭解顧客之長期購買行為，例如，購買麵包之後再購買果醬的有次序性的重覆購買行為。

(6) 趨勢分析 (trend analysis)：是指藉由蒐集資料，並藉由製作圖表來分析未來趨勢，而其基本假設為過去曾經發生過，將來也會發生，例如利用股票走勢圖表來預測股價未來的表現。

(7)時間序列分析(time serial analysis)：乃藉由時間數列所能反映的現象，如社會經濟的發展過程及其規律性，用以預測未來發展趨勢之方法。例如：100年10月12日至100年10月20日的每日宏達電的股票交易資料。藉由觀察及分析宏達電股票，以預測未來走勢或價格。

2.2.4 ID3 (Iterative Dichotomiser 3) 演算法

Quinlan (1979) 提出決策樹 (Decision Tree) 的ID3 (Iterative Dichotomiser 3) 演算法，它是一種在進行預測過程中選擇最佳屬性為其節點，中間節點代表預測之屬性，分枝為屬性測試之結果，一直延伸，似樹狀之圖形，而每個分枝也代表著一種方案，與屬性交織盤根而錯節，用以解答預測過程時所有事件發生的可能性，其藉著取得各類別的訊息量、各個屬性得訊息量及取得各屬性訊息獲取量之步驟以決定最佳之屬性，並以最多之訊息獲取量之屬性為其根，次者為下個節點，再次者為第三節點，以建構完整之決策樹，並用於解決預測之問題。

2.2.5 資料探勘應用

近年來資料探勘已成功地應用在許多方面，其中在行銷預測部份包括：胡建勳與王建驊(2005)利用關聯法則 (Association Rules) 技術於探討線上遊戲玩家的消費行為；許惠婷(2005)則運用決策樹 (Decision Trees) 探勘技術找出交易資料所隱藏的規則以作為外包供應商之管理依據；廖樹賢、蕭培元與周繕緣(2011)則利用關聯法則探討兩岸股票市場指數之連動性；而醫學預測模型的建立部份包括：洪冠群(2004)運用關聯法則 (Association Rules) 技術以血液透析來探究病患之住院預測；姚志成(2005)運用決策樹探勘技術於建構脂肪肝預測模式；而在其學校行政及學生課業預測支援的部份包括：蔡秀滿、林文宗與劉紫瑩(2008)，應用關聯法則及決策樹技術於預測學生特質與學業成績間之關係；陳垂呈與陳幸暉(2011)則利用資料探勘的分群化 (Clustering) 技術於建置圖書館書籍推薦系統；詹益敏(2008)針對學生的「在學資料」及「學籍資料」利用決策樹及時窗分析 (Time Windows) 等資料探勘技術來預測休退學的學生並提出預警。

而在ID3演算法應用方面，林余任(2002)曾以ID3決策樹預測最大總利潤，發現採行差異化行銷手法，即針對高消費群、中消費群及低消費群，分別給予不同的價位或贈品的方式，能獲取最大總利潤；林志弘(2003)運用ID3演算法於預測中輟生之系統上，得到74%的準確性結果；高淑珍(2004)採用ID3演算法於顧客回應之模式上，發現財力、繳費方式、年齡與保險銷售間之關係；蕭世全(2006)曾運用決策樹ID3演算法於建立精密模具之應用上，發現不同屬性須有不同解決對策，須建立知識庫及將模具老師傅之內隱知識轉成外顯知識以獲得更高之成效；李明修(2007)應用於分析圖書館閱覽者之行為，而理出了一套不同系所學生借閱書籍的法則等等。

參、研究方法

本研究為了獲得休退學學生之預測模型，將以ID3演算法針對歷史資料中已辦理休退學之學生其入學時的人口統計變數，以所蒐集資料的三分之二用來制定休退學規則及產生預測模型，接下來再以蒐集資料的三分之一以印證預測模型之預測準確度。至於使用的七項人口統計變數說明如下：

1.入學管道：分為(1)聯招生 (2)推甄生 (3)申請入學生 (4)技優生 (5)身心障礙生。

- 2.是否應屆：分為(1)應屆畢業 (2)非應屆畢業。
- 3.戶籍地：分為(1)北區 (2)中區 (3)南區 (4)花東離島地區。
- 4.入學前學制：分為(1)高中 (2)高職。
- 5.性別：分為(1)男 (2)女。
- 6.特殊身份：分為(1)一般生 (2)原住民 (3)身障生 (4)退伍軍人 (5)僑生及離島生。
- 7.畢業科系：分為(1)工程學科 (2)商業管理學科 (3)綜合高中及普通科。

之後本研究依循下列步驟進行模型建立：

步驟一：計算出休退學類別訊息量熵值(entropy)

Equation (A) : $I(nc_1, nc_2, \dots, nc_n)$

nc_i : 結論是屬於 c_i 這一類的觀察紀錄數目, $i=1, 2 \dots n$

N : 總觀察紀錄數目

步驟二：取得各個屬性的訊息量

Equation (B) : $E(A) = \sum_{i=1}^m \left(\frac{n_{vi}}{N} \right) I(n_1 v_1 c_1, n_2 v_2 c_2, \dots, n_n v_n c_n)$

$E(A)$: 屬性A的訊息量

m : 屬性A的可能值的數目

n_{vi} : 屬性A的值是 v_i 的觀察紀錄數目

$n_i v_i c_i$: 屬性A的值是 v_i 且其結論是屬於 c_i 的觀察紀錄數目。

步驟三：取得屬性訊息獲取量

Equation (C) : $\text{Gain}(A) = \text{Equation (A)} - \text{Equation (B)}$

步驟四：產生規則

而在規則驗證的部份，本研究以之前的步驟所產生的預測規則再利用蒐集資料的三分之一進行印證，計算出符合預測結果的筆數，並進一步取得預測的準確率。

肆、研究結果與討論

4.1 研究結果

依據上述研究方法，本研究採用國內南部某科技大學從2005年7月至2010年10月，共計3112筆，剔除近幾年已廢除之招生管道及資料不完整難以探勘之資料，剩餘2722筆，先以其中1815筆資料中之學生入學前人口統計變數共七項屬性來產生規則及模型，得到如表1之各屬性訊息量結果，以及如表2之屬性訊息獲取量。

表1 屬性訊息量結果

屬性	戶籍地	入學管道	特殊身份	畢業科系	性別	入學前學制	是否應屆
屬性訊息量	0.247971	0.24846	0.248592	0.249266	0.249416	0.249604	0.24977

表2 屬性訊息獲取量排序表

屬性	戶籍地	入學管道	特殊身份	畢業科系	性別	入學前學制	是否應屆
屬性訊息獲取量	0.751624	0.751135	0.751003	0.750329	0.750179	0.749991	0.749825
順序	1	2	3	4	5	6	7

根據各屬性訊息獲取量排序產生130條預測規則，由於有部份規則雖然有相同的屬性卻同時歸屬於休學及退學的類別，因此針對此情形本研究採取以支持比例來描繪其佔了休退學總數中之比重，並將其分列為兩種規則，依此方法，本研究最後得到了231條預測規則，因礙於篇幅所限，在此僅列出20條預測規則（如表3）加以說明。由表3欄位中編號代表其規則之編號，戶籍地、入學管道、特殊身份、畢業科系、性別、入學前學制、是否應屆為研究之七大屬性。編號1及編號2其戶籍地為中區、入學管道為申請入學、特殊身份為一般生、畢業科系綜高及普通科、性別為女、入學前學制屬於高中、是否應屆屬於應屆，其探勘結果有可能辦理休學或退學，其支持比例代表其辦理休學或退學的比例各佔50%；編號9中代表戶籍地為北區、入學管道是技優生、特殊身份屬一般生、畢業科系為商業管理學科、性別是男、入學前學制屬高職、是否應藉中屬於應屆，其只能辦理退學，故其支持比例為100%。

表3 部份的預測規則結果

編號	戶籍地	入學管道	特殊身份	畢業科系	性別	入學前學制	是否應屆	結果	支持比例
1	中區	申請入學	一般生	綜高及普通科	女	高中	應屆	休學	50.00%
2	中區	申請入學	一般生	綜高及普通科	女	高中	應屆	退學	50.00%
3	中區	推甄生	一般生	商業管理學科	女	高職	應屆	退學	62.16%
4	中區	推甄生	一般生	商業管理學科	男	高職	應屆	休學	44.44%
5	中區	聯招生	一般生	工程學科	男	高職	非應屆	休學	50.00%
6	中區	聯招生	一般生	工程學科	男	高職	應屆	退學	43.75%
7	北區	申請入學	一般生	綜高及普通科	男	高中	應屆	退學	75.00%
8	北區	技優生	一般生	工程學科	女	高職	應屆	休學	50.00%
9	北區	技優生	一般生	商業管理學科	男	高職	應屆	退學	100.00%
10	北區	身心障礙生	身障生	工程學科	男	高職	應屆	休學	50.00%
11	北區	身心障礙生	身障生	綜高及普通科	女	高中	應屆	退學	50.00%
12	北區	推甄生	一般生	工程學科	男	高職	應屆	休學	50.00%

13	北區	推甄生	一般生	商業管理學科	男	高職	應屆	退學	77.78%
14	北區	聯招生	一般生	工程學科	男	高職	非應屆	退學	100.00%
15	北區	聯招生	一般生	綜高及普通科	男	高中	應屆	退學	60.00%
16	花東離島地區	申請入學	一般生	綜高及普通科	男	高中	應屆	退學	100.00%
17	花東離島地區	聯招生	一般生	商業管理學科	女	高職	應屆	退學	100.00%
18	南區	申請入學	一般生	綜高及普通科	男	高中	非應屆	退學	71.43%
19	南區	身心障礙生	身障生	商業管理學科	男	高職	應屆	休學	62.50%
20	南區	聯招生	退伍軍人	工程學科	男	高職	非應屆	退學	20.00%

經由上列得到231條規則之後，本研究再以剩餘907筆學生資料來驗證結果的準確度，結果發現規則可以正確預測的筆數為865筆，正確率為95.37%。

4.2 討論

從研究所獲得的休退學之預測規則中，我們發現下列現象：

(1)從表3中編號第16的規則中，我們可以看出屬性中戶籍地位於花東離島地區、入學管道為申請入學、特殊身份為一般生、畢業科系為綜高及普通科、性別為男性、是否應屆中屬於應屆生，其較偏向於辦理退學，原因是因為綜高及普通科的學生學習較著重於學術概念，難以適應技職體系之教學方式，故而容易改換學習跑道。

(2)從表3中編號第17的規則中，我們可以看出屬性中戶籍地位於花東離島地區、入學管道為聯招生、特殊身份為一般生、畢業科系為商業管理學科、性別為女性、入學前學制為高職、是否應屆中屬於應屆生，其較偏向於辦理退學，原因是因為花東離島地區距離家裡太遠，學生會想選擇離家裡較近之地方就讀。

伍、結論與研究限制

由於資料探勘已臻完善，再加上決策樹之ID3演算法簡單而明確，誠如本研究採行國內某私立科技大學之休退學學生歷史性資料作為樣本，資料範圍自2005年7月至2010年10月，共計3112筆，剔除近幾年已廢除之招生管道及資料不完整難以探勘之資料，剩餘2722筆，先以1815筆資料根據七大屬性來作探勘並以ID3演算法求出其訊息獲取量，根據屬性訊息獲取量高低排序，接著再分為休退學兩種屬性並求得支持比重，最後產生休退學之規則，共計231條，另外再用剩餘之907筆資料來加以印證，得到規則的預測準確率為95.37%，因此利用ID3演算法運用於休退學之預測上，確實能有效預測出屬於休退學之高危險群學生，並產生有效之規則與模型，透過這些預測規則及模型將使決策者更容易判斷未來可能辦理休退學之潛在學生而防範於未然。

本研究雖然已經將休退學學生之七項人口統計變數資料納入研究範圍，但是仍有其他值得探勘之人口統計變數尚未被考慮，例如：經濟因素、家庭狀況，未來如果能夠加入這些人口統計變數，相信將能使得探勘所獲得的預測模型更加準確，更加具有預測力。

參考文獻

1. 李明修，2007，『應用資料探勘技術分析圖書館閱覽者行為』，朝陽科技大學工業工程與管理系碩士論文。
2. 林余任，2002，『以ID3預測總利潤最大化之研究』，大葉大學工業工程研究所碩士論文。
3. 林志弘，2003，『中輟生預測系統之探索-資料挖掘之應用』，國立中央大學資訊管理研究所碩士論文。
4. 林清文，2005，『中途輟學防治輔導』，教育研究月刊第一三二期：154~155 頁。
5. 林樹，2006，『我國中輟生之問題與對策』，教育暨外國語文學報，第二期：1-13 頁。
6. 胡建勳、王建驊，2005，『利用資料探勘技術探討線上遊戲玩家消費行為』，電子商務經營管理研討會，逢甲大學電子商務研究中心主辦。
7. 姚志成，2005，『運用資料探勘技術建構脂肪肝預測模式』，中原大學資訊管理學系碩士論文。
8. 洪冠群，2004，『多重最小支持度關聯規則探勘演算法之醫療檢驗應用-以血液透析病患之住院預測為例』，國立東華大學資訊工程學系碩士在職專班論文。
9. 高淑珍，2004，『應用資料探勘於顧客回應模式之研究-以國內 A 壽險公司為例』，國立成功大學企業管理學系博士論文。
10. 陳垂呈、陳幸暉，2011，『建置圖書館書籍推薦系統：資料探勘之應用』，工程科技與教育學刊，第八卷·第一期：469~478 頁。
11. 許惠婷，2005，『運用資料探勘於外包供應商管理之研究』，國立台灣科技大學工業管理系碩士論文。
12. 黃明鎮，2009，『失翼的一群-輟學生問題與青少年犯罪』，
網址：<http://life.fhl.net/Desert/980903/010.htm>。
13. 章勝傑、周慧婷，2008，『國小導師預測學生國中輟學可能性之甄別研究』，臺北市立教育大學學報，第三卷·第一期：35~70 頁。
14. 楊士隆，2001，『幫派少年成長歷程與副文化之調查研究』，國立中正大學犯罪防治研究所。
15. 詹益敏，2008，『資料探勘技術應用於學生休退學預警系統之建構』，朝陽科技大學資訊管理系碩士論文。
16. 廖樹賢、蕭培元、周繕緣，2011，『資料探勘應用於 ECFA 對兩岸股票市場指數連動性之研究』，科技整合管理研討會，淡江大學主辦。
17. 蔡秀滿、林文宗、劉紫瑩，2008，『資料探勘技術應用於學生特質與學業表現之研究』，資訊科技國際研討會，朝陽科技大學主辦。
18. 翁慈宗，2009，『資料探勘的發展與挑戰』，科學發展期刊，第四四二期：32~39 頁。
19. 簡禎富，2005，『決策分析與管理』，177~253、557~574 頁，台北：雙葉書廊。
20. 蕭世全，2006，『資料探勘於建立精密模具試模知識研究』，國立雲林科技大學資訊管理系碩士論文。

21. Berry, M., and Linoff, G, "Data Mining Techniques for Marketing," Sales and Customer Support, Wiley, New York., 1997.
22. Fayyad, U., Piatetsky-Shapiro, G., and Smith, P., "The KDD Process for Extracting Useful Knowledge from Volumes of Data," Communication of ACM, Vol. 39, No. 11, November 1996, pp. 27-34.
23. Quinlan, J. R., Discovering rules from large collections of examples: A case, in D. Michie, ed., "Expert Systems in the Micro Electronic Age," Edinburgh University Press, 1979.

The ID3 Algorithm for Mining Rules of Students' Suspension or Withdrawal—Taking a University of Technology in South Taiwan for Example

Chen, Ting-Pin
Kun-Shan University Information Department
teddy@mail.ksu.edu.tw

Kao, Shu-Chen
Kun-Shan University Information Department
kaosc@mail.ksu.edu.tw

Abstract

The number of colleges and universities has been expanding rapidly, but the birth rate has been decreasing annually in Taiwan. Furthermore, the amount of students' suspension or withdrawal also has been becoming more than before. The manager of universities encounters more difficulties respecting to these problems. Thus, the problems of studying suspension and withdrawal not only result in the waste of educational resources, but also produce such more social problems. Therefore the problems about students' suspension or withdrawal have been becoming an important topic recently. In this study, ID3 (Iterative Dichotomization), which is a well known data mining technology, is used to build the prediction model for resolving the problem of students' suspension or withdrawal. With the collected data, 231 rules are generated with ID3 algorithm and the precise prediction rate of that calculation reaches is 95.37%. This result will provide the manager of university with valuable information to prevent the situation of suspension or withdrawal.

Keywords: prediction of suspension, data mining, ID3 algorithm