

目錄

前言	1
研究背景	1
研究現況	2
研究目的	2
文獻探討	4
資料挖掘的定義	4
資料挖掘過程的四大步驟	4
重要的資料挖掘技術	4
<i>Market Basket Analysis</i>	5
<i>Memory-Based Reasoning (MBR)</i>	5
<i>Cluster Detection</i>	5
<i>Link Analysis</i>	6
<i>Decision Trees and Rule Induction</i>	6
<i>Artificial Neural Networks</i>	6
<i>Genetic Algorithms</i>	6
資料與方法	8
投保中斷之資料挖掘	8
一、投保中斷之定義	8
二、資料來源	11
三、資料轉換流程	13
繳費異常之資料挖掘	14
一、繳費異常之定義	14
二、資料來源	16
三、資料轉換流程	19
資料挖掘技術	20
決策樹(<i>Decision Trees</i>)	21
關聯規則(<i>Association Rule</i>)	24
<i>Missing Data</i> 的處理方式	25
結果	26
中斷	26
第一類欠費	33
第六類欠費	36

承保中斷/繳費異常資料挖掘系統.....	40
目的.....	40
軟硬體環境.....	40
資料載入方式.....	41
系統架構與功能.....	41
使用說明.....	44
待預測資料檔格式.....	50
參考文獻.....	53
附錄.....	54
圖表.....	54
投保中斷相關記錄之描述性統計.....	54
第一類繳費異常相關紀錄之描述性統計.....	60
第六類繳費異常相關之描述性統計.....	65
行職業分類中英對照表.....	71
投保單位-郵遞區號-通訊地址(ZIP_CODE)對照.....	71
期末報告審查意見回應書.....	72

圖目錄

圖 1、健保局處理中斷資料流程之一.....	9
圖 2、健保局處理中斷資料流程之二.....	10
圖 3、投保中斷分析之資料轉換及處理流程.....	13
圖 4、健保局處理繳費事宜之流程.....	15
圖 5、繳費異常分析之資料轉換與處理.....	19
圖 6、決策樹.....	21
圖 7、刪剪樹對減少錯誤率的效用.....	22
圖 8、中斷資料重要屬性之分佈-1.....	26
圖 9、中斷資料重要屬性之分佈-2.....	27
圖 10、中斷資料重要屬性之分佈-3.....	27
圖 11、一類欠費重要屬性之分佈-1.....	33
圖 12、一類欠費重要屬性之分佈-2.....	33
圖 13、一類欠費重要屬性之分佈-3.....	34
圖 14、六類欠費重要屬性之分佈-1.....	37
圖 17、系統功能圖.....	41
圖 16、系統架構圖.....	41
圖 18、系統主畫面.....	45
圖 19、模式建立介面.....	45
圖 20、資料預測介面.....	48
圖 21、資料預測之參數設定介面.....	48
圖 22、資料預測之執行與狀態顯示介面.....	49
附圖 1、轉出/入未中斷與中斷記錄之比數.....	54
附圖 2、性別在未中斷/中斷資料中之分佈.....	54
附圖 3、年齡在未中斷/中斷資料中之分佈.....	55
附圖 4、投保金額在未中斷/中斷資料中之分佈.....	55
附圖 5、前轉出/入次數在未中斷/中斷資料中之分佈.....	56
附圖 6、投保單位通訊地在未中斷/中斷資料中之分佈.....	56
附圖 7、行業別在未中斷/中斷資料中之分佈.....	56
附圖 8、投保單位隸屬機關未中斷/中斷資料中之分佈.....	57
附圖 9、投保單位眷屬人數在未中斷/中斷資料中之分佈.....	57
附圖 10、投保單位被保險人在保人數在未中斷/中斷資料中之分佈.....	57
附圖 11、投保單位保險類目在未中斷/中斷資料中之分佈.....	58
附圖 12、投保單位被保險人眷屬人數在未中斷/中斷資料中之分佈.....	58
附圖 13、眷屬加退保次數人數在未中斷/中斷資料中之分佈.....	58
附圖 14、眷屬平均年齡在未中斷/中斷資料中之分佈.....	59

附圖 15、前欠費次數在第一類資料中之分佈	60
附圖 16、應繳費用在第一類資料中之分佈	60
附圖 17、中斷人數異動次數在第一類資料中之分佈	61
附圖 18、投保單位異動次數在第一類資料中之分佈	61
附圖 19、投保單位通訊地在第一類資料中之分佈	61
附圖 20、投保單位行業別在第一類資料中之分佈	62
附圖 21、眷數在保人數在第一類資料中之分佈	62
附圖 22、眷數在保人數增減在第一類資料中之分佈	62
附圖 23、被保險人在保人數在未欠費資料中之分佈	63
附圖 24、被保險人在保人數在欠費資料中之分佈	63
附圖 25、被保險人在保人數增減在未欠費資料中之分佈	63
附圖 26、被保險人在保人數增減在欠費資料中之分佈	64
附圖 27、保險類目在未欠費資料中之分佈	64
附圖 28、保險類目在欠費資料中之分佈	64
附圖 29、前欠費次數在第六類未欠費資料中之分佈	65
附圖 30、前欠費次數在第六類欠費資料中之分佈	65
附圖 31、性別在第六類未欠費資料中之分佈	65
附圖 32、性別在第六類欠費資料中之分佈	66
附圖 33、年齡在第六類未欠費資料中之分佈	66
附圖 34、性別在第六類欠費資料中之分佈	66
附圖 35、投保金額在第六類未欠費資料中之分佈	67
附圖 36、投保金額在第六類欠費資料中之分佈	67
附圖 37、應繳金額在第六類未欠費資料中之分佈	67
附圖 38、應繳金額在第六類欠費資料中之分佈	68
附圖 39、通訊地在第六類未欠費資料中之分佈	68
附圖 40、通訊地在第六類欠費資料中之分佈	68
附圖 41、中斷次數在第六類資料中之分佈	69
附圖 42、轉出/入次數在第六類未欠費資料中之分佈	69
附圖 43、轉出/入次數在第六類欠費資料中之分佈	69
附圖 44、轉出/入次數之增減在第六類未欠費資料中之分佈	70
附圖 45、轉出/入次數之增減在第六類欠費資料中之分佈	70

表目錄

表 1、資料挖掘主要技術.....	4
表 2、投保中斷分析之原始資料表.....	11
表 3、投保中斷分析相關屬性之定義及來源表格之一.....	11
表 4、投保中斷分析相關屬性之定義及來源表格之二.....	12
表 5、第一類繳費異常分析之原始資料表格.....	16
表 6、第一類繳費異常分析之相關屬性定義與來源表格之一.....	16
表 7、第一類繳費異常分析之相關屬性定義與來源表格之二.....	17
表 8、第六類繳費異常分析之原始資料表格.....	17
表 9、第六類繳費異常分析基本屬性之定義與來源表格.....	18
表 10、第六類繳費異常分析衍生屬性之定義與來源表格.....	18
表 11、易中斷規則中各類被保險人之分佈-1.....	27
表 12、易中斷規則中各類被保險人之分佈-2.....	28
表 13、易中斷規則中各類被保險人之分佈-3.....	28
表 14、易中斷規則中各類被保險人之分佈-4.....	28
表 15、易中斷規則中各類被保險人之分佈-5.....	29
表 16、易中斷規則中各類被保險人之分佈-6.....	29
表 17、易中斷規則中各類被保險人之分佈-7.....	29
表 18、易中斷規則中各類被保險人之分佈-8.....	30
表 19、易中斷規則中各類被保險人之分佈-9.....	30
表 20、易中斷規則中各類被保險人之分佈-10.....	30
表 21、易中斷規則中各類被保險人之分佈-11.....	31
表 22、易中斷規則中各類被保險人之分佈-12.....	31
表 23、不易中斷規則中各類被保險人之分佈-1.....	31
表 24、不易中斷規則中各類被保險人之分佈-2.....	32
表 25、不易中斷規則中各類被保險人之分佈-3.....	32
表 26、執行環境.....	40
表 27、原始與目的資料表格之對應.....	42
表 28、資料轉換封裝說明.....	43
表 27、分析參數.....	44
表 29、承保中斷待預測檔之格式.....	50
表 30、一類繳費異常待預測檔案之格式.....	51
表 31、六類繳費異常待預測檔之格式.....	52