臺灣地區燙傷住院治療型態之研究－－應用階層式集群分析於全民健保資料庫

Treatment Patterns of Hospitalized Burn Injury in Taiwan --- Hierarchical Clustering in National Healthcare Insurance

中文摘要

本研究利用近年來在基因研究及文獻探勘上流行的階層式集群分析，應用於全民健保險研究資料庫，研究台灣地區 1996 年至 2000 年住院燒燙傷病人的治療型態。在醫學資訊學方面，則探討階層式集群的建構過程中，有多處應用專業知識微調的機會，此外並試驗在資料倉儲的線上分析處理(OLAP)上應用的可能性。在各種正規化(Normalization)處理資料以因應問題所需的嘗試後，強調對於極端值及闕漏值多的資料應有特殊考量，提出在適當窗口正規化重整資料的重要性(Windowed_Normalization)。群集有效性的證明可分兩方面－－內在的一致性係對照有無正規化或窗口正規化的三種群集方法的結果的差異；外在的意義則分析各種傷害程度，病人，及醫院特徵在各群集的不同分布，並利用 C4.5 決策樹顯現前述參數與各群集的關係。

本研究利用換藥，清創，植皮三個主軸為群集根據，發現台灣區的住院燙傷的外科相關治療，可分為六大群。高額植皮型及多清創少換藥型較常發生在有燙傷專業服務的醫院，而無燙傷專業服務的醫院則偏向無植皮低花費型及多換藥少清創型。

利用階層式層集找出治療型態的分群，使預測住院總費用的迴歸模型的誤差降低，在同類治療型態中也較容易挑出歧異病例。以 R2 來評估，使用受傷及醫院的基本資料加上治療型態的資料，或加上總住院天數，讓多變量線性模型的 R2 由 0.650 升到 0.702 及 0.876，而迴歸樹模型則由 0.697 升到 0.790 及 0.924。

這套燙傷治療的群集，可應用於醫院管理上比較低風險及高風險病人組成的用途。小群集可分離出同質性的病人群，進行比一般流行病學死亡率或住院時間長短以外更細緻的研究。

英文摘要

My research applys hierarchical agglomerative clustering (HAC), popular recently in phylogenomics studies and text mining, in clinical research in National Healthcare Insurance Research Database (NHIRD), for the treatment patterns of hospitalized burn patients from 1996 to 2000 in Taiwan. In medical informatics, the integration of domain knowledge into clustering algorithms was stressed. Data visualization of hierarchical clusters could be applicable in On-Line Analytical Processing (OLAP) of data warehouse.

By experience on data normalization specific to the problem, the importance of management of extreme values and missing values was stressed. "Windowed normalization" was proposed. The cluster results were validated in two aspects --- the intrinsic consistency was shown by comparison of results of clustering with or without normalization; the extrinsic characteristics of clusters were shown by distribution of parameters among injuries, patients, and hospitals, and were analyzed by decion tree models.

Dressing-change, debridement operations, and skin grafting were finally used as criteria of clustering. Six groups of treatment patterns were noted. More heavy-grafing type and heavy-dressing-change type occurred in hospitals with burn specialties. No-grafting-and-low-fee and more-dressing-change-than-debridement styles were favored by hospitals without burn specialties.

Clustering of treatment patterns improved the fit and decreased the errors of regression models for prediction of total medical expenses, and helped the judgment of outliers. After clustering, R square of multiple linear regression was improved from 0.650 to 0.702 and 0.876, and R square of regression trees were improved from 0.697 to 0.790 and 0.924, by demographic factors and cluster information, or by addition of length of stay as an independent variable.

The patterns of clustering could be used as reference for hospital administration, for comparison of high-risk and low-risk patient components. More delicate studies could be arose by the homeogeneous patient partitions after cluster analysis.