應用醫學文獻分類於小兒輸血實證醫學之研究

The Application of Automatic Categorization to Medical Literature in Pediatrics Transfusion for Evidence-Based Medicine

中文摘要

小兒輸血醫學是跨學科領域,臨床醫師在面對輸血決策時所需面對的各種不確定性,往往需要查檢大量的醫學文獻,而醫學文獻即是透過許多嚴謹的臨床實驗所撰寫而成的非結構性資料。本研究在探討運用文獻探勘技術進行文獻自動分類之研究,建立以 MeSH 為基礎佐以臨床醫師經驗所建立的「使用者導向」分類架構。透過 Clever Craft 系統的輔助,進行有關小兒輸血的醫學文獻分類,並以視覺化的知識網絡圖呈現文獻相關概念,期盼臨床醫師能透過該小兒輸血文獻探勘系統,在面對病患時能提供快速且精確查檢資料與相關概念分析圖,藉由一些相關文獻的佐證以作出對輸血安全及血品適應症最佳處置。

系統架構中有兩個重要模組:(1)為自動分類的訓練模組(2)為階層式知識分類模組,皆以貝氏(Bayesian)定理為主要方法結構;以其概念將文件向量化對所使用的詞庫進行比對找出字詞關聯性,利用系統中的文件自動分類技術,經臨床專家指導式學習所產生的分類規則,準確進行文獻的分類,提供使用者能在眾多資料中精確得到所需文獻。

本研究採用較高標準,所以並未將所有資料庫非相關文獻納入分母(樣本母數)計算,而是將經由關鍵詞檢索後所製成的資料庫中擷取測試樣本及訓練樣本,經過類目選擇、關鍵詞篩選定詞、樣本數篩選,最後選擇最佳訓練模組為 319 篇訓練文獻,對該文件探勘系統的自動分類功能重新訓練,並以不同相似度對此 100 篇測試文獻進行評估,發現當相似度為 0.7 時,得到平均精確度 65.33%,回收率 35.05%,並以 Kappa 值評估自動分類系統與臨床主治醫師判斷結果具有一致性。透過與 PubMed 比較其檢索功能發現,透過本文件探勘系統可大大降低檢索時的不確定性。

最後,建議目前台灣實證醫學教育,除了醫學教育人員外,應當結合醫學工程、圖書資訊人員,提供良好的自動分類、系統評讀文獻系統以及文件探勘系統,對於非結構性資料與文獻加以彙整,提供臨床醫學更方便、實用的實證醫學步驟。

英文摘要

The pediatrics transfusion is a cross-domain knowledge. When physicians face to the decision of transfusion, they always must to search many medical literatures. Medical literature is a non-structure data as a result by rigorous experiment. This research is focused on an automatic categorization technique of text mining. To build a classification is based on "drawn to user" by MeSH and physicians experience. Using a text mining system-Clever Craft to catalog the medical literature. Then the

system generated the knowledge networks to show the relationships between concepts. We hope this system has a high-speed and accurate search with an analysis of concepts that help doctors searching when they face on patients. Using an evidence of medical literature to have the best decision.

There are two models in the system. One is the training model for automatic categorization. The other is the hierarchies-supervised learning model that is based on Bayesian method. It is used the document vector to find the association tuning between the term to term. By the technique of automatic categorization, the classification rule is produced by supervised learning. The rule helps the automatic categorization to be precision that enabled user to find the literatures from the knowledgebase.

The assessment of this research is a high level criterion. The denominator is not included the non-relational literatures from the database. The information extraction is searched the keyword "newborn and fetal and pediatrics and transfusion" from database. Then we extract the testing samples and training samples from the information collection. Through the categories selection, keyword set, sample selection, we try to find the best set that is 319 literatures to be training samples. Then training the automation categorization of this system again. We selection the same testing samples are 100 literatures. We evaluate the different similarity with these test samples. When the similarity is 0.7, the precision value is 65.33%. The recall value is 35.05%. By Kappa test, this research result is coherency between the automatic categorization and experts apprehension. The search interface in this system compares with PubMed, this text mining system can reduce the chanciness.

At last, I want to propose something with the evidence-based medicine education in Taiwan. We should combine the expert by medical educationist and medical engineer and librarian. It provided fine automatic categorization and evidence literatures and text mining system. The information collection can automatic extract the unstructured data and literatures. To provided a simple and fine system in EBM procedure.