以消化道內視鏡基礎標準術語第二版〔MST2〕名詞爲範本的自動內視鏡報告分類系統

Development and Evaluation Of A Gastroendoscopic Automatic Classification Report System Based on Gastroendoscopic Minimal Standard Terminology Version 2

中文摘要

本研究以資訊擷取及自然語言處理的技術,將非結構化自由型態消化道內視鏡報告,以消化道內視鏡基礎標準術語(MST2)爲本體,經過分析整理成「可延伸性標示語言」(Extensible Markup Language,簡稱 XML)的消化道內視鏡報告,提供查詢,跨平台交換及醫學統計。未來也可與消化道內視鏡影像儲存系統結合,合成完整消化道內視鏡電子資料庫系統,提供醫學照顧與治療決策更精確的資訊。研究方法:首先,評估國內消化道內視鏡檢查報告的結構性符合 MST2 架構的程度,依據 MST2 架構將消化道內視鏡檢查報告依檢查的部位、檢查的發現及診斷分三個部分,每一個部份有三個評估子項,每一子項依據與 MST2 符合程度給分,由完全符合給與 3 分至完全不符合給與 0 分,由一位消化系專科醫師對每篇文章評分,每篇報告最高總分共 27 分。其次,進行名詞剖析,以 MST2 同義字引擎將已剖析名詞進行同義字轉換成 MST2 標準名詞,處理文章切割則將文章依據四個文章中的關鍵詞 Esophagus、Stomach、Duodenum 及 Impression 將文章分爲以此四個字爲首的四個區段。計算每一個標準名詞在各區段文章中的序位。將轉換後擷取出的標準名詞以 XML 進行文字索引及標示,最後形成以 MST2 爲索引典的結構式自動分類內視鏡報告。預期經過處理後的消化道內視鏡報告其文件分類可以更爲有效。評估方法將以公元 2001 年由某市立醫院 5 位消化系專科醫師紀錄的 1266 篇消化道內視鏡報告爲訓練資料庫,以 2005 年某市立醫學中心 1699 篇檢查報告爲測試資料庫,其中由六位消化系專科醫師撰寫的消化道內視鏡報告 489 篇及上消化道治療性內視鏡檢查報告 30 篇。以"回現率"與"精確率"二因素來衡量評估檢索效能。程式語言使用 JAVA。轉換後檢查報告以 XML 格式儲存。

結果:選取 1235 篇訓練組資料庫進行 MST2 結構式相符評分分析,發現只有 33 篇(2.7%)報告檢查內容能完全符合標準名詞規範,但是大部分(60%)的內視鏡報告 9 個子項中仍有 4 到 6 項符合 MST2 的要求。將訓練資料庫 1266 篇文字進行同義字分析後,設計十種 MST2 同義字轉換形式,發現轉換後可增加訓練資料庫標準名詞的擷取數達 2.56 倍以上。訓練資料庫與測試資料庫皆能以超過 95%的精確率與回現率來將文章正確的切割成爲以 esophagus、stomach、duodenum 與 impression 爲段首的四個區段。若以 stomach、ulcer 及 body 三個標準名詞來評估文章檢索功能,將訓練資料庫調教至精確率爲 76.9%、回現率 100%時,在測試

資料庫精確率爲 68.75%、回現率 100%。若能以"esophagus/varices/grade II/lower third"固定序位的次序模式擷取，文件檢索效能在訓練資料庫會由精確率:80.7%、回現率: 74.2%變爲精確率:93.6%(44/47)、回現率: 70.9%(44/62)。

結果與討論：對非使用英文爲母語的醫師，無法很精確使用英文準名詞來描述消化道內視鏡檢查結果，只有 60%的檢查報告內容能部份符合 MST2 標準名詞規範，很少(2.7%)能完全符合，因此這類報告在資訊擷取或文章分析時，將無法得到正確結果。但是消化道內視鏡檢查內容皆具有一固定格式，因此可依據關鍵詞來進行文章分割，有超過 95%的精確率與回現率。不過當有不同類型報告混合進行分析時，如將超音波或大腸鏡檢查報告一起分析時，精確率會下降至 60%以下。在醫學報告中進行名詞檢索，若先經過同義字轉換成標準名詞後，有非常高的回現率，但是精確率只有 70%，原因在於醫學專有名詞較統一且經過同義字轉換，容易擷取，但是檢索的名詞間是否在文章中有相關，則需經過自然語言處理來協助提高精確率。MST2 標準名詞在報告內容中會有固定的出現序位，將序位模式列入分析時，可增進檢索精確率。同樣發現在一些檢查爲正常的報告，醫師會以固定報告範本輸入，如檢索 ulcer 時，可發現 ulcer 出現於文章中第 18 個位置時，內容爲 no ulcer 的可能有 95%。但是當擷取名詞時，有多個從屬名詞出現時，如 bleeding、erosion 與 ulcer 都被檢索出時，需藉助三個名詞彼此間在文章中序位的向量分析，才能正確分類。自由型態非結構式消化道內視鏡檢查報告在自然語言處理後，是可以轉換成標準名詞結構式報告，且可進行檢索與分類。如此，可以達到更有效的分類，本研究發現轉換後的消化道內視鏡檢查報告分類，食道區的主要病變爲 Erosion 及 Ulcer 且大都是單一病灶出現；在胃部主要病灶爲多發的 erosion，或是單一的 superficial ulcer 且沒有 Bleeding；在 12 指腸以單一的 cratered ulcer 爲主。不過分類仍會因爲錯別字，自創簡寫及負面敘述，造成錯誤，需有更強大的字典庫及同義字庫與人工智慧來協助分辨，文字序位向量及文章模式或許可作爲人工智慧分析的條件。

結論：在分析不是以英文爲母語的文章撰寫者而言，不論在文字、文法及描述上有相當大的認知差距，再加上有自創縮寫及包括格式及打字錯誤的書寫錯誤，造成自然語言處理的困難，若能用標準名詞及架構、導入概念空間(Concept Space)及知識主體(Domain Knowledge)的協助，資訊擷取及文件將可更精確。本研究嘗試用世界消化內視鏡醫學會頒布的消化道內視鏡基礎標準術語第二版(MST 2)名詞作爲同義字轉換及結構式資料庫轉換依據，以可延伸性標示語言〈XML〉作爲轉換後標準語言格式，提供一種可行方式將非結構化自由型態消化道內視鏡報告以自然語言處理轉換成標準名詞結構且自動分類的消化道內視鏡報告，提高資訊擷取與文件分析的效能。

英文摘要

In this research we used information retrieval (IR) technique and natural language processing (NLP) to transform nonstructural free style gastroendoscopic reports into

structured formatted with the extensive markup language (XML) based on the minimal standard terminology version 2 (MST2) ontology. These structured reports can be used for classification, query, sharing and medical decision making. These reports and their counterparts, the endoscopy images, can be integrated as an intelligent endoscopic report system to support more accurate and precise information for decision making on care and treatment.

Study design: The terms proposed in for describing location of findings, terms, and diagnosis were used to match the ones in the gastroendoscopy reports. The MST2 terms are classified into three categories. The match score was evaluated based on the fitness of the terms described in each subcategory with MST2 from complete match scored 3 to no match scored 0.The total scores are 27. Each report was processed by the techniques including synonym transformation, segmentation, ordering, indexing, marking up and classification. Synonym transformation technique is used to map the local terms to the corresponding MST2 terms. A report is segmented into four paragraphs by four leading words: Esophagus, Stomach, Duodenum and Impression. We marked up the standard MST2 terms and calculated each standard MST2 term's position in each segmented paragraph. Finally, we transformed the reports into structured gastroendoscopic reports. We evaluated the power of information retrieval and text analysis by using the precision rate and recall rate. We selected 1266 gastroendoscopy reports recorded by five gastroenterologists in 2001 in a municipal hospital as the training set and other 1699 medical examination reports, in which 489 gastroendoscopy reports and 30 therapeutic endoscopy reports recorded by 6 gastroenterologists, from a municipal medical center as the testing set. The programs were written by JAVA language and the reports were stored as XML files.

Results: There were only 33 (2.7%) reports that matched the MST2's standard terms and ontology in 1235 reports from the training set, but 60% of the reports could score from 4 to 6 in each category. We analyzed 1266 reports from the training set and developed ten synonym templates based on MST2 standard terms. Using the synonym templates for extraction of MST2 terms from the reports, we could increase the number of terms up to 2.56 folds compare to those without the templates. There were higher than 95% precision rate and recall rate to segment those reports from both the training set and testing set into 4 segments. We evaluated information retrieval function with stomach, ulcer and body. The training set tuned the precision rate as 76.9% and the recall rate as 100%. The results of the testing set were precision rate 68.75% and recall rate 100%. We set the extraction model in fixed orders as "esophagus/varices/grade ii/lower third". We modified retrieval function from precision rate 80%, recall rate 74.2% to precision rate 93.6% and recall rate 70.9% .

Discussion: It is discovered that nonnative English doctors could not precisely use

English standard terms to prescribe findings in the gastroendoscopy examinations. Hardly to extract accurate information from these reports. But, there were still some fixed formats in gastroendoscopic reports may segment it into designed paragraphs with high (>95%) precision rate and recall rate. If the data sets mixed with various reports from colonoscopy or sonography examinations, the precision rate will decrease lower than 60%. We could retrieval terms with high recall rate after synonym transformation to be standard terms but lower precision rate (70%) in medical examination reports. The reasons were that the terms we retrieved may not being relevant. We need natural language processes to parse the terms to get higher precision rate. There were fixed position orders in some gastroendoscopic reports. If we input some terms' orders then we could improve our retrieval precision rate. We found that physicians recorded normal reports with fixed template. There were 144 reports recorded ulcer in the position 18 of stomach paragraph and 95% reports associated with no ulcer. When we retrieved more than one terms with relevant such as bleeding, ulcer and erosion, we need vector analysis with position among them and to get more accurate precision rate and classification rate.. We can process free style nonstructural gastroendoscopic reports by natural language to be a structured standard reports. Also errors may cause by erratum, personal abbreviations and negative descriptions. We need more powerful medical dictionary and synonyms map to parse the terms written by nonnative English doctors. Word's position vectors and report's models could be the factors for artificial intelligence. We analyzed automatic classified structured reports according to NLP algorithm. The main lesions in Esophagus segments were a single erosion or ulcer. The main lesions in Stomach segments were multiple erosions or a single superficial ulcer without bleeding. The main lesion in Duodenum segments was a single cratered ulcer.

Conclusion: It was difficult to analyze the medical reports recorded by those nonnative English doctors due to the large discrepancy in terms, grammar and prescriptions, There were personal abbreviations and wrong spellings including formatted or typed errors that made natural language processing more difficult. What if we could use standard terminology concept space and domain knowledge that we could extract information more accurately. This research selected World Organization for Digestive Endoscopy (OMED) published Minimal Standard Terminology Version 2 for synonyms and structured transformation's templates and transformed the gastroendoscopic reports to XML reports. These transformations could improve information retrieval and text analysis accuracy. The standard terms synonym transformation and markup with XML reports could be a useful nature language process to analyze the nonstructural free style gastroendoscopic reports.