

醫學診斷決策支援系統

李友專 劉立 邱文達*
台北醫學院 醫學資訊中心
萬芳醫芳 院長室*

前言

「專家系統」(Expert System)一直是電腦科學中的人工智慧(Artificial Intelligence)研究者的熱門話題。顧名思義，專家系統便是具有如同人類專家一般，能對特定領域的問題做判斷、解釋及認知的一組電腦程式。但由於此特定領域可大可小，且對「認知」的定義亦有不同解釋，故可有小如某些所謂「汽車辨識專家系統」只能依照汽車外型等幾項特徵辨認十種車；亦有大如某些實驗階段的超大型醫學專家系統，可依據十二萬個不同的醫學表徵(Manifestations)分辨八千個疾病。儘管專家系統的定義未盡明確，但基本上，當此系統所能處理的問題其複雜性及對專業知識的需求相近或高於人類專家，且其表現(performance)亦相近或超過人類專家時，我們便可稱之為專家系統。但醫療診斷專家系統之研究人員逐漸把研究目標從建造超越醫師診斷能力之系統轉移至以協助醫師達成正確診斷為目的之診斷決策支援系統後，Medical Expert System 之名稱也逐漸被 Medical Decision Support System 所取代。

與醫學相關的決策支援系統可以依其功能分為診斷、治療或監測(Monitor)之用。本文將以診斷決策支援系統為主，對於其應用領域及應用範圍做一簡單之介紹，並對此類系統的構建機制，如知識表現法(Knowledge Representation)及知識工程(Knowledge Engineering)等作一初步之介紹。

診斷決策支援系統之緣起

大體而言，醫學診斷主要是以細心觀察、配合各種檢驗檢查所得之資料，再以豐富的醫學知識利用所謂的假說-演繹法(Hypothetico-deductive process)之心智推理行為而成，長久以來，被認為是醫學中最深奧難學的部份。一位醫師由醫學生到住院醫師訓練的十數年間，可說有一大部份是花在學習如何作出最正確精準之醫學診斷上。

但近代醫學進步一日千里，醫師們縱使窮盡畢生之力，亦未必能盡覽所有的醫學新知。由於知識的爆炸，許多資訊處理工具因應而生，其中與醫學診斷最相關的便是利用人工智慧(Artificial Intelligence)研究領域中的專家系統(Expert System)技術所建造的醫學診斷決策支援系統(Medical Diagnostic Decision Support System-MDDS)。

最早的 MDDS 可追溯自 1960 年代 Warner 之先天性心臟病診斷系統、1970 年代 de Dombal 之急性腹痛鑑別診斷系統及 Shortliffe 之 MYCIN-腦膜炎(meningitis)的診斷系統。其中 de Dombal 之系統被設計於急診做急性腹痛之診斷(de Dombal et al, 1972)，在一次 304 個 case 的 Study 中，此專家系統成功地診斷出所有真正有 Acute Appendicitis 之 case，但將 non-specific abdominal pain 誤診為 Acute Appendicitis 之情形則有六例；另一方面急診的專科醫師卻誤診了六例真正 Acute Appendicitis 的病人，並且將二十例以上的 non-specific

Title: Medical Diagnostic Decision Support System

Authors: Yu-Chuan Li, Li Liu, Wen-Ta Chiu

Center for Biomedical Informatics, Taipei Medical College

Office of Superintendent, WanFang Hospital*

Key Words: diagnostic support, decision support, expert system, artificial intelligence

abdominal pain 病人誤診為 Acute Appendicitis 而送進開刀房。此系統後來發展出可在 IBM-PC 上執行的版本而在英國的急診室被廣泛地使用。以腦膜炎為知識範圍的 MYCIN 則可經由一連串의 交談式資料輸入而協助診斷腦膜炎之病原菌並提出用以治療此病之抗生素種類及劑量。MYCIN 也曾被證明具有與傳染病科專家一樣好的診斷及處方能力(以其所能辨認之疾病為限)。但由於此系統當時僅能在價值數百萬美元的大型電腦上執行，它從未真正被應用在臨床診斷上。

繼 MYCIN 之後，仍有許多診斷專家系統陸陸續續在各研究機構被發表，如針對 Glaucoma(青光眼)的診斷及治療而設計的 CASNET 及麻省理工學院的 PIP 等。然而這些知名的診斷專家系統雖多半已被證明為在某方面相當或優於人類專家，但由於其知識範圍均限於狹小的醫學領域(如青光眼之診斷)，且多為研究階段的程式，許多硬體及使用者介面的限制使得其廣泛應用的可能性大大降低。一般使用者只能望文興歎，卻無機會親自操作使用，幸運的是，隨著電腦軟/硬體技術的日新月異，在八〇年代末期，終於出現了數套具有龐大知識庫與親善的使用者介面(User-friendly Interface)，並可在 PC 或 Macintosh 等個人電腦上執行的醫學診斷專家系統。

新一代醫學診斷決策支援系統

在這些使用方便且價格便宜的新一代 MDD S 中，以猶他大學的 Iliad，匹茲堡大學的 QMR(Quick Medical Reference)及麻省綜合醫院(Massachusetts General Hospital)與哈佛大學合作的 DxPlan 為最受世人矚目者。其中 QMR 及 DxPlan 是由原本在大型電腦上累積了十多年經驗的診斷專家系統轉化而來，Iliad 則是以大型醫院資訊系統中的決策模組(decision modules)為雛型，在 Macintosh 上發展出來的。據估計，光是為了蒐集及轉化醫學知識，研究人員已花費了約 50,000 man-hour 在 Iliad 上，至今每天仍有四~六位各科的專科醫師在猶他

大學的知識工程室中不斷地改進及擴張其知識庫。

Iliad QMR 及 DxPlan 所涵蓋的知識領域包括了內科各次專科的多數疾病及臨床表徵(manifestation)。在使用者輸入已知的臨床表徵如病史，身體檢查(Physical examination)，實驗室檢驗及放射科或病理檢查之後，系統會根據這些資料再反問用者相關的問題，以期得到完整而精確的資訊。經過這樣一段交談式的資料輸入過程後，電腦會根據這些資料，用知識庫中的醫學知識加以分析推論，並回饋予使用者推論之結果，使用者則可依此結果而輸入或修改資料以得到更精確的診斷，或在病情尚未明朗的情況下，詢問電腦下一步應作何種檢驗或測試，方可以最少之花費得到最多的資訊(如 rule in 或 rule out 某些 diagnosis)。

這些診斷專家系統最大的功能便是在這使用者輸入資料與電腦回饋出結果的過程中，使醫師(預設的使用者)能對此病人的病情有全面性的考慮與了解，並能在不浪費醫療資源之情況下，清楚地計畫下一步的診斷或醫療策略。同時這些系統對年輕的醫師亦有教導臨床診斷技術的效果。美國已有三十個以上的醫學院用 Iliad 作學臨床診斷課程的教學工具。

在診斷的正確性上，許多醫學文獻一再證明這些系統至少具有一般內科住院醫師相當的診斷能力，在 Annals of Internal Medicine 的一篇以特別困難的臨床病例之初步診斷為實驗樣本的文章裏，QMR 以 85% 的正確診斷率超越了臨床醫師的 60% 的正確診斷率。這些應歸功於電腦龐大的記憶量及不偏差(unbiased)的推理能力。但在診斷專家系統的研究發展上，仍有許多極具挑戰性的問題有待克服；如多重診斷，時間性的病情變化等，仍是這些系統較弱的部分，此外，如果所期待的診斷或疾病不在此系統的知識庫內，那電腦便完全無能為力了。

用於醫學診斷之知識表現法

自從西元 1961 年第一個醫學診斷決策支

援系統被公開以來，computer scientists 及人工智慧之研究者紛紛提出各種新的機制及演算法以便製造更有智慧的系統。這些用來轉換知識成爲電腦可處理的法則/數據的方法，通稱爲知識表現法(Knowledge Representation-KR)，用在 MDDS 上之 KR 可分爲以下五類：

一、法則式(Rule-based Approach)：以 IF...THEN 之句子建造所有的知識法則，下例來自 MYCIN 系統中千餘條法則中的一條：IF the site of culture is blood AND the morphology of the organism is rod AND the gram stain is gram(-) AND the patient is a compromised host, THEN there is suggestive evidence(0.6) that the organism is *Pseudomonas aeruginosa*。

二、機率式(Probabilistic Inference)：此種知識表現法利用機率來表現知識中的不確定性，並利用機率學及數學方程式來計算推論的結果，其中最常爲 MDDS 之研究者採用之方式爲貝氏定理(Bayesian formulation)，前面所提 de Dombal et al 之急性腹痛診斷系統便是使用了所謂的 Simple Bayes(單純式貝氏定理)，另有 multi-membership Bayes(多元式貝氏定理)，及近幾年方被提出之 Bayesian Networks(貝氏推論網路)，則是結合機率學及圖學的機率性知識表現法，在醫學上之應用頗有潛力。

三、統計式(Statistical Modeling)：利用各種統計技巧來達成推理之目標，常用的方式爲 Logistic Regression、Linear Discriminant Function 及 Recursive partitioning 等。

四、類神經網路(Artificial Neural Networks)：Artificial Neural Networks(ANN)是一種利用軟體方式模擬一組神經元改變彼此間的連結比重(weighting)以適應/學習一群資料中隱含的 pattern 之技術。此種方式特別適合由大型之病例資料庫中衍生診斷之推論模型。

五、其它：除了上述四大類，其它如 Fussy Sets Theory, Belief function 也偶有應用在 MDDS 上的例子，其中 Fussy Sets Theory 更因其被利用在機械之自動控制(如：地下鐵、洗衣機、機車等)而聲名大噪。有些系統則不仰賴任何教學或資訊理論而自行發展評分系統來決定

某一症狀對某一疾病之重要性，此類系統則稱作 Ad hoc Scoring System，如上述之 QMR 系統便是一例。

以上各種 Knowledge representation 各有其特長，亦各有其缺點，其中 Rule-based approach 較適合規則明確之知識，一旦 rule 確立，建造其資料庫便相當容易，但其主要缺點在於(1)不易表達推論中之不確定性。(2)若 rule 太多則不易維護，新舊 rule 可能含有相反或部份衝突之邏輯。

Probabilistic Inference 則可完整地表達推論中的不確定性(uncertainty)，並藉由機率理論靈活地 manipulate 這些 uncertainty。有鑑於醫學診斷中不確定因子甚多，機率性之推論不失爲 MDDS 之優良選擇。但此法之缺點在於知識工程(Knowledge Engineering-KE)的過程中需要蒐集(或請專家估計)大量的機率性資料，相當費時費力，但目前已有不少之技巧可用來降低此種困擾，將在文後詳述。

Statistical modeling 及 Artificial Neural Networks 皆屬於利用大量已知之病例資料來訓練(train)一可變之非線性數學模型，其好處在於 KE 過程的單純化，基本上只要有足夠之樣本供訓練之用，則電腦可自行建構概論之模型。但另一方面，若訓練之樣本不足或資料缺損則難以造出可信賴之模型；又如果所提供之樣本有偏差，則所造出之模型亦爲偏頗失準之模型。美國國防部曾訓練 Artificial Neural Networks 辨認藏在草叢中的戰車之照片，結果 ANN 將每一張陰天之照片皆判別爲有戰車，實驗小組百思不得其解，最後發現在提供給 ANN 作訓練的照片之中，每一張有戰車之照片皆爲陰天所拍攝，故 ANN 所學習到之 pattern 實爲陰天之景色而非戰車之影像，此乃偏差之樣本造成 ANN 失誤的一有趣範例。

醫學診斷決策支援系統 所遭遇之困難

醫學診斷決策支援系統雖然由於資訊技術的進步而漸漸地有更多的使用者可以接觸到，

但過去的系統卻一直無法在臨床上廣泛地被應用，其主因有二：一、冗長的資料輸入過程：欲達到正確之診斷，使用者多半需要輸入許多的臨床相關資料。依據過去的研究，對一個內科診斷系統而言，輸入一個 case 平均約需 15~30 分鐘，在繁忙的臨床事務中，醫師們幾乎不可能有如此長之時間可以作資料輸入的動作，是以醫師們對早期的診斷專家系統之使用興趣不高。現在及未來之趨勢則傾向於將 MDDS 整合於醫院資訊系統之中，如此便可由醫院資訊系統中取出所需之病患資料而省下大量的輸入時間，但前提是在此醫院資訊系統中已存有完整的病患資訊，也就是 Computerized Patient Record 必須先實現。二、取代醫師診斷的偏差目標：早期之系統由於人們對人工智慧技術期待過高，傾向於以取代或超越醫師之診斷能力為目標，是以困難重重。醫師們之排斥抗拒，亦使此類系統在臨床上之推展徒勞無功。事實上雖然此類系統之診斷正確性已達一定水準，但仍不能取代醫師之診斷。部分的原因來自病人可能同時有多重診斷，時間性的病情變化等。且此系統之正確性評估最終仍來自醫學專家，而醫學診斷上常有專家意見不一致之狀況，如此高的 Inter-observer variability 使得此類系統的評估困難度甚高。但 Medical Expert System 之觀念在 1990 年代已有重大之轉變，醫療診斷專家系統之研究人員不再以建造超越醫師診斷能力之系統為目標，而改以協助醫師達成正確且完整之診斷為目的之診斷決策支援系統為研究目標。Medical Expert System 之名稱也逐漸被 Medical Decision Support System 所取代。

未來之展望

綜觀這三十多年來診斷專家系統的演變，我們不難看出研究的重心已經逐漸的由找出更好更準確的診斷機制演算法(Algorithm)轉移到如何將此類系統整合在真正的臨床環境中，且利用此類系統的長處來彌補人類醫師的不足。記得在幾次演講後曾有數位醫師向我提出這樣

的問題：「若是診斷專家系統繼續如此演化下去，將來醫師要何去何從？」相信許多醫師也有著類似的疑問。就此問題我除了強調 MDDS 的研究者早已不將取代醫師做為研究的目標，個人更以為：一位好的醫師除了下立正確的診斷之外，仍需具備優秀的 Physical Examination 技巧，作 Study 的能力及了解並治療患者心理與身體病痛的胸懷。這絕不是任何診斷輔助工具或電腦程式所能取代的。相反的，醫師們應試著去了解並利用電腦的特性來補足人類的不完美性，讓臨床醫學也像許多其他領域一樣，藉著電腦的輔助而更上層樓。

推薦讀物

1. Shortliffe EH, Perreault EL, Fagan LM, eds: Wiederhold G (assoc eds) Medical informatics: computer applications in health care. Addison-Wesley, Reading, Massachusetts, 1990.
2. Warner HR, Toronto AF, Veasey LG, Stephenson RA: Mathematical approach to medical diagnosis. JAMA 1961;177:75-81.
3. Ledley RS, Lusted LB: Reasoning foundations of medical diagnosis. Sci 1959;130:9-21.
4. de Dombal FT, Hartley JR, Sleeman DH: A computer-assisted system for learning clinical diagnosis. Lancet 1969;1:145-8.
5. Warner HR, Toronto AF, Veasey LG. Experience with Bayes' theorem for computer diagnosis of congenital heart disease. Ann NY Acad Sci 1964;115:558-67.
6. Dombal FT, Leaper DJ, Horrocks JC, Staniland JR, Clamp SE: Human and computer-based diagnosis of lower gastrointestinal disorders in 468 patients. Br J Surg 1973;60:915.
7. Perry CA: Knowledge bases in medicine: a review. Bull Med Libr Assoc 1990(3):271-82.
8. Miller RA, Pople HEJ, Myers JD: Internist-I: An experimental computer-based diagnostic consultation for general internal medicine. N Engl J Med 1982;307:468-76.

9. Barnett GO, Cimino JJ, Hupp JA, Hoffer EP: DXplan: An evolving diagnostic decision-support system. *JAMA* 1987;258:67-74.
10. Bankowitz RA, McNeil MA, Challinor SM, Miller RA: Effect of a computer-assisted general medicine diagnostic consultation service on house staff diagnostic strategy. *Meth Inf Med* 1989;28:352-6.
11. Bankowitz RA, McNeil MA, Challinor SM, Parker RC, Kapoor WN, Miller RA: A computer-assisted medical diagnostic consultation service-implementation and prospective evaluation of a prototype. *Ann Int Med* 1989; 110:824-32.
12. Li YC, Haug PJ: Evaluation of the quality of a probabilistic diagnostic system using different inferencing strategies. *Proceedings of the 17th Symposium on Computer Applications in Medical Care, Washington DC: 1993;471-7.*
13. McDonald CJ: Protocol-based computer reminders, the quality of care and the non-perfectability of man. *New Engl J Med* 1976; (24):1351-5.