

大數據 之 醫療運用與人文反省

The Medical Application and Humanistic Reflection in
Big Data

蔡甫昌 主編



元照出版提供 請勿公開散布

黃瀚萱、陳信希、陳豈嶼、陳建煒、陳育群、李偉強
盧子彬、李建樂、張耀尹、謝嘉珊、邱鈺喬、賴亮全
蔡孟勳、莊曜宇、顏如娟、許明暉、陳聰富、蔡甫昌
蔡玫芬、林子倫 合著

元照出版



醫療大數據及其 應用

黃瀚萱

臺灣大學資訊工程學系

陳信希

臺灣大學資訊工程學系



元照出版提供 請勿公開散布。

本文曾刊載於《台灣醫學》2016年20卷6期，頁589-594。

壹、前言

大數據 (big data) 是近年新興的名詞，意指數量太多、產生或更新速度太快、內容太複雜，而難以用人工處理的大規模資料。隨著科技進展、資訊裝置普及以及網路使用的盛行，每分每秒都出現各式各樣的資料、「大數據」一詞也涉及處理這些資料所需的科技以及應用¹。

常見的醫療衛生資料包含醫學文獻、電子病歷、臨床紀錄、檢驗結果等。從資料格式來說，可概分為結構化 (structured)、半結構化 (semi-structured) 以及非結構化 (unstructured) 三種²。其中，結構化資料包含體溫、血糖、血壓等檢驗指數，以及診斷代碼、會計帳目等有意義的數量 (numerical) 或類別 (categorical) 特徵，電腦可以直接分析。非結構化資料或半結構化資料則像是電子病歷上的文字敘述，以及X光片、電腦斷層攝影、核磁共振造影等醫學影像。對於非結構化或半結構化資料，影像辨識、自然語言處理、資訊擷取的技術扮演了重要的角色，可讓電腦從原始資料中找出有意義的內容，轉成結構化資料再做進一步分析。

- 1 Andreu-Perez J, Poon CCY, Merrifield RD, et al: Big data for health. *IEEE J Biomed Health Inform* 2015;19:1193-208.
- 2 Feldman B, Martin EM, Skotnes T: Big data in healthcare hype and hope. Dr. Bonnie 360°. 2012. <http://www.west-info.eu/files/big-data-in-healthcare.pdf/> Accessed August 4, 2016.



在醫療衛生（healthcare）領域，目前已有許多醫學、護理、公共衛生等方面的數據，能透過各種管道取得，亦有醫學資訊學（health informatics）和生物資訊學（bioinformatics）等學門，研究技術運用資料解決問題、創造新的應用。例如個人化照護、臨床運作、藥物開發病預防、癒後管理、藥物基因組學（pharmacogenomics）、轉譯基因組學（translational genomics）等議題，都有大數據發揮價值的空間。尤其在高齡化社會，大數據及其技術可以讓醫療資源的使用更有效率，強化預防醫學與早期介入的效果，促進人類個體與群體的健康。

本文擬從資料型態以及應用方式的面向，來介紹大數據於醫療衛生的發展近況。首先分析大數據於醫療領域的挑戰性因素，接著描述常見的醫療資料，包含電子病歷、生物特徵、醫學影像、社群媒體，並逐項介紹現有的應用。

貳、大數據的挑戰

一般來說，大數據的處理程序包含取得、儲存、分享、分析、檢索、決策等步驟。在處理大數據的過程中，有所謂的四種技術性考量：容量（volume）、時效性（velocity）、多樣性（variety）、準確性（veracity），合稱4V³。除此之外，亦有學者提出價值（value）、效度

3 Feldman B, Martin EM, Skotnes T: Big data in healthcare hype and

(validity)、揮發性 (volatility)、變異性 (variability) 等其他考量。

一、容量

大數據基本的特徵就在於其龐大的資料量。美國的健康照護系統在2011年時，資料容量已達150 exabytes (10^{18})，且隨著社群媒體、穿戴裝置、影片串流、感測器 (sensors) 的發展，文字、影像等多種媒體的資料，仍在持續地快速成長中。龐大的容量意味著儲存設備與計算複雜度的挑戰。



元照出版提供 請勿公開散布

二、時效性

每天都有大量的新資料不斷湧現，除了有效率地儲存、運用這些新資料，反應速度也是大數據技術的另一發展重點。例如可以從社群網站的留言內容中，發掘流行病的傳播趨勢，或是穿戴式裝置可以隨時監測使用者的心跳、心壓等資訊，一旦偵測到狀況有所異常，便能立即警示使用者緊急處置，預防更嚴重的傷害。然而，這樣的應用必須建立在資訊系統的即時性上，如果處理資料的流程

hope. Dr. Bonnie 360°. 2012. <http://www.west-info.eu/files/big-data-in-healthcare.pdf/> Accessed August 4, 2016.

Fang R, Pouyanfar S, Yang Y, et al: Computational health informatics in the big data age: a survey. *ACM Comput Surv* 2016;49:12.



有太多延遲，從接收到回應的間隔（latency）太長，則無法在黃金時間介入危機。

三、多樣性

醫療衛生資料有各種來源，包含病歷、學術論文、基因、醫學影像、社群媒體、感測器等，有些是文字資料，有些是影像資料。如前所述，半結構化與非結構化資料須先轉為結構化資料，而不同的資料類型有不同的背景技術，例如文字資料涉及自然語言處理，影像資料則與電腦視覺、影像辨識有關。



元照出版提供 請勿公開散布

四、準確性

大數據從不同來源蒐集而得，無法完全掌控品質，往往包含雜訊與取樣偏差。病歷資料中常常夾帶錯字、縮寫、甚至曖昧難解的片段，例如常見的縮寫DM一字，可能是糖尿病（diabetes mellitus），也可能是多巴胺（dopamine），亦可能是舒張期雜音（diastolic murmur）。手機或穿戴式裝置內的感測器所測定的數值（心跳、血壓、活動狀況等），也常有不準確的情況。各種雜訊造成了大數據處理的阻礙。

雲端運算（cloud computing）以及機器學習（machine learning）這兩個熱門的資訊領域，可以對應以上四種挑

戰性因素。雲端運算的概念由分散式系統延伸而來，發展出一系列的機制，讓許多部同質或異質的電腦系統協同運作，儲存、計算大量資料。比起單機計算，雲端計算的技術可以用更低的成本、更快的速度來處理更多的資料。而機器學習的技術，則是讓電腦從大量的過往範例中訓練出決策模型，從而對未知的新資料進行預測，解決各式各樣的問題。

醫療衛生資料涉及敏感度極高的個人資訊，所以在取得、儲存、分享這些資料時，須考慮隱私的議題⁴。取得資料的過程必須明確告知個體，取得知情同意。而資料的儲存與傳輸則要注重安全性，除了不能讓資料外洩，也要保護資料的正確與完整。

提供 請勿公開散布

參、常見的醫療資料與應用

一、電子病歷

電子病歷（electronic health records, EHRs）記錄患者每次就診的摘要資訊，包含就診時間、科別、主訴、病史、治療經過等部分。過去病歷是由醫師手寫記錄，現在漸漸電子化，部分醫師改為在電子病歷系統上輸入，而手

4 Kieseberg P, Schantl J, Fruhwirt P, et al: Witnesses for the doctor in the loop. In: Guo Y, Friston K, Aldo F et al, eds. Brain Informatics and Health, 1st ed. Heidelberg, Springer, 2015:369-78.

國家圖書館出版品預行編目資料

大數據之醫療運用與人文反省 / 黃瀚萱等合著；
蔡甫昌主編。-- 初版。-- 臺北市：元照，
2018. 04

面；公分

ISBN 978-957-8607-02-6(精裝)

1.醫療資訊學 2.資料探勘 3.醫事法規 4.論述
分析

410.29

106022864

大數據之醫療運用與人文反省

元照出版提供 請勿公開散 5L035HA

2018年4月 初版第1刷

主 編 蔡甫昌
作 者 黃瀚萱、陳信希、陳豈嶼、陳建煒、陳育群、
李偉強、盧子彬、李建樂、張耀尹、謝嘉珊、
邱鈺喬、賴亮全、蔡孟勳、莊曜宇、顏如娟、
許明暉、陳聰富、蔡甫昌、蔡玫芬、林子倫
出 版 者 元照出版有限公司
100臺北市館前路18號5樓
網 址 www.angle.com.tw
定 價 新臺幣350元
專 線 (02)2375-6688
傳 真 (02)2331-8496
郵政劃撥 19246890元照出版有限公司

Copyright ©2018 by Angle Publishing Co., Ltd.

登記證號：局版臺業字第1531號

ISBN 978-957-8607-02-6