

線上健康類新聞之分析與預測-巨量資料 架構

On-Line Health News Analysis and Prediction-A Framework of Big Data

企業管理學報 第 113 期(106 年 06 月) 頁 1-29

DOI: 10.3966/102596272017060113001

吳家豪*

Jia-Hao Wu

馬麗菁**

Li-Ching Ma

* 國立聯合大學資訊管理學系碩士
Master, Department of Information Management, National United University.

** 國立聯合大學資訊管理學系教授 (聯絡作者)
Professor, Department of Information Management, National United University.
(correspondence author)

摘要

近年來台灣受到食安風暴及環境污染等事件影響，健康議題常受到大眾關注。因此，本研究針對線上健康類新聞的標題與內文，利用文字探勘技術，支援向量機、順序邏輯迴歸與決策樹三種預測方法，進行按讚數多寡的分析與預測，找出預測正確率最高的組合，以預測新文章的按讚數多寡。此外，為了因應未來可能面臨的巨量新聞資料，本研究進一步建置巨量資料平台，進行資料預測與分析。研究結果顯示，按讚數較多的標題常包含疫苗防治與慢性病等議題的相關詞彙；包含警語類與飲食類的按讚數較少；在預測方法上，以支援向量機搭配50個縮減概念維度組合的預測正確率最高。本研究結果可供業者參考，業者可篩選較受歡迎的健康新聞，吸引更多讀者上線閱讀與回應。

關鍵字:文字探勘、商業智慧、巨量資料、健康新聞、預測

ABSTRACT

People in Taiwan have been alerted by the problems of food safety for the past few years; therefore, people have paid more attention to health news. This study aims to find the critical terms in the on-line health news and predict votes for the "Like" of the news based on text mining and business intelligence algorithms. In addition, in order to deal with the possible big data from on-line news, this study proposes a framework of big data by parallel processing on multiple data nodes. The results show that the support vector machine with 50 concept dimensions has the best prediction accuracy. When the amount of data becomes huge, the performance of distributed computing structure will improve significantly. The proposed approach can help managers of on-line news to choose or invest more popular health news thus to attract more potential readers. The proposed structure and analytic results can also provide a better understanding of big data for the future studies.

Keywords: Text mining, Business intelligence, Big data, Health news, Prediction.