

## 活用巨量資料

臺北駐大阪經濟文化辦事處派駐人員

自網路登場後，累積在世界中之電子資料爆炸性持續增加。即所謂之「BIG DATA」，從電車延遲等與生活息息相關之預測、未知物質之探索等，蘊藏各式各樣值得研究之可能性。而其活用，取決於資訊之公開與人才培育。

2014 年舉辦了一場競賽，預測什麼樣的氣象條件可能影響鐵道路線造成停駛或延遲。其預測之基礎即來自於 NTT DOCOMO 基地台等所計測之氣象資訊，提供首都圈約 200 個地點之資料。將其與 JR 東日本 1 年份 10 條線路之運行資料套用比對，建構計算模式，半年後使用這些結果預測，第 1 名可獲 20 萬日圓獎金。由社會人士及學生等共 48 人參加，募集到 683 個作品。命中率第 1 名為 75%，且若扣除因施工等原因造成者則可超過 90%。會影響電車延遲的不只是氣象因素，電車之班次也是原因之一。計算公式設計時的重點為時間帶、假日或平日等。開發者為上班族男性(25 歲)，已參加過數次資料分析競賽琢磨技巧。「最大之魅力為可接觸到平常接觸不到的資訊。且因此獲得成果時會有很大的滿足感」。主辦資料分析之公司 OPT Data Science Lab. 之齊藤秀負責人表示「不只是專家，開放性的資訊可由各種不同人才加入自己的感性來分析，開放式科學在巨量資料 BIG DATA 的活用上相當重要的」。

這樣的競賽結果也被嚐試活用在研究或現實社會裡。美國的資料分析公司 Kaggle 將來自企業或公家機關所委託的資料請各種不同人才進行分析。美國航空總署 NASA 委託透過宇宙圖像來推測暗黑物質的中心位置，醫師團體委託透過入院或定期到醫院者之履歷來預測未來的醫療費支出。

巨量資料未來還會繼續增加。因而當前的問題是資料科學專家之不足。文部省公佈 2020 年的電子資料量為 2000 年的 6,500 倍，預測將產生 25 萬人之人才缺乏。所謂資料科學專家必須是精通統計學，會分析資料並做到可視化，會設定透過資訊解決的課題，能與不同分野之研究或業界相互合作。文部省計畫分為「學徒」、「獨自操作」、「棟樑」、「業界代表」分階段來培養人才。惟美國一年培養 2.5 萬人，中國 1.7 萬人，印度 1.3 萬人，日本卻只有 3,400 人。文部科學省研究振興局之榎本剛參事官表示「能掌控現場且傳授技術給後進之棟樑級人才特別少」。

於 2013 年成立，以培養人才為目標之資料科學專家協會之代表理事草野隆史表示「非常缺乏能獨自完成課題的設定、統計處理、資料運用的人才。聚集擁有各自的擅長分野並組成一個取得平衡的團隊是相當重要的」。

### 一、關於隱私問題

為進一步活用巨量資料，於 2015 年 9 月修正個人情報保護法。只要將企業所持有之個人資料加工至無法鎖定某個個人且無法復原的程度，即可不經本人同意對外提供資料。但在國會審議過程中受到可能侵害隱私的批判。關於要加工到何種程度之標準，接下來將由第三機關之「個人情報保護委員會」制定。

### 二、大學設立專門學院

滋賀大學將於 2017 年度開設全國第一個學習巨量資料分析之「資料科學院」，一學年招收 100 人左右。擬定透過交通卡片或電子錢包之資料預測改善塞車或流行預估等。

### 三、中央之補助

文部省計畫自 2016 年起針對以蒐集巨量資料或人工智能等之分析計畫，10 年期間投資 1 千億日圓。首先將先選出具獨創性構想的 100 個案子，補助研究者 10 億日圓。

資料來源:2015 年 10 月 18 日朝日新聞

