

工業大數據之異常偵測分析挑戰：以紙業為例

白浩廷

工業技術研究院 綠能與環境研究所

htpai@itri.org.tw

摘要

異常偵測應用包括防治金融詐欺、改善工業設備營運等。以紙業為例，造紙流程涉及上千個感測器，形成一個高維度且巨量的資料集。德國 PTS 指出每次斷紙事件將造成至少 6,000 歐元(相當於 210,000 新臺幣)損失，斷紙頻率約每日發生 6 至 9 次。此外，斷紙造成的能耗虛耗約占總生產量的 2%至 7%，不僅是經濟上的虧損還會浪費大量資源。本研究探討異常偵測技術與其應用在斷紙分析之發展狀況，我們以非線性支持向量機分類方法(Non-linear SVM, N-SVM)分析國外斷紙資料集，並從分析結果論述工業數據之異常偵測挑戰、異常因素探索以及資料品質的重要性。

關鍵詞：大數據分析、機器學習、斷紙

Abstract

Anomaly detection technology has been widely applied to varied areas, e.g., fraud detection for credit cards, fault detection in safety critical systems, and so on. In paper industry, the paper-making process involves in thousands of sensors, which forms high-dimensional large amounts of data. In analyzing such big data, the state-of-the-art methods probably suffer from computation and distortion problems. According to PTS, a paper break costs around 210,000 NT\$, and it occurred 6 to 9 times per day. In addition, paper breaks cause 2–7% of the total production loss. In this paper, we survey the taxonomy of anomaly detection methods and their applications in analyzing paper breaks. Moreover, we adopt the non-linear SVM method (N-SVM) to analyze the paper breaks dataset. Finally, we discuss the findings and illustrate the importance of anomaly exploration and data quality.

Keywords: Big data analytics, machine learning, paper breaks