**應用機器學習建立紡織設備零件加工之良率檢測模型**

邱守燦、洪鈺欣

國立雲林科技大學 工業工程與管理研究所

通訊作者: [hungyh@yuntech.edu.tw](mailto:hungyh@yuntech.edu.tw)

**摘要**

紡織設備製造是國內主要的成衣產業之一，以研發與製造導向為主，為了把關紡織設備的產品品質，在作業員加工紡織設備零件的作業時，針對零件的良率進行檢測。本研究採用某紡織設備製造商提供的零件規格與作業人員紀錄資料，以500筆工單數量，交由20位作業員從事生產，同時逐一記錄各工單完成後的零件規格，透過真實提供的資料集進行探索性資料分析，繪製各個作業員工單完成的數量與零件規格的統計圖表，找出零件規格的離群值視為不良品，經過統計方法計算後，得知產品不良率為4%和允收機率為95.61%。得知以上資訊後，本研究利用機器學習的分類演算法建立紡織設備零件加工作業的良率檢測模型，包括:決策樹、隨機森林、支持向量機、極限梯度提升，評估模型的分類績效，找出影響加工良率的主要資訊，使最終結果能協助廠商未來建置產品良率監測系統的參考依據。

**關鍵字：**紡織設備零件加工、良率檢測、探索性資料分析、分類演算法

1. **緒論**

**1.1 研究背景**

成衣產業為新興國家倚重的產業，國內的成衣產業企業已多數外移，轉往東南亞與印度等勞力密集的市場進行投資，但紡織設備製造仍是國內成衣產業轉型當中，具有舉足輕重的地位，國內許多紡織設備出口至歐美、日本等先進國家，在先進國家紡織設備視為精密技術的新創產品，以本研究的廠商為例，廠商主要的產品包刮:電腦式紡織設備、電子式紡織設備等，紡織設備導入自動化和資料庫系統，讓紡織設備在業界的傳統機械式紡織設備製造觀念中，轉型為高度研發與精密技術的結合；近年來，紡織設備的需求逐年增加，原先在新興國家，紡織設備以家庭式代工的企業為主要客戶，但在先進國家，除了在文創與休閒的客戶需求之外，同時也成為跨行業以服務導向的產品，持續產品創新與客戶建立品質保證。

此研究案例的廠商提供了500筆工單的紡織設備零件加工資料集，而負責這些零件製造的作業人員與零件規格(長、寬、高)皆已被詳細記錄於資料集當中，於是本研究以零件規格的資料與作業人員當作研究對象，且廠商已多年沒有進行零件加工的良率檢測，廠商僅依照經驗法則評斷，因此進行此廠商在零件加工的品質相關研究。

**1.2 研究動機與目的**

為了得知紡織設備零件製造商品質管理的情況，本研究透過廠商提供的資料集建置零件加工的良率檢測模型，先利用敘述統計學與統計圖表等方法執行探索性資料分析，觀察各要因的數據，並從統計圖表呈現要因當中重要的資訊將其視覺化，同時得知不良品的比率為4%，進一步判別出作業人員的生產情形。

得知以上的資訊之後，本研究根據隨機抽樣的理論，界定品質管理的要求:隨機抽樣20個零件，若有2個以下的不良品，即可允收這批貨，從而利用二項分配數學式求出允收機率為95.61%，當作廠商進行品質改善的參考指標，最後利用機器學習方法中的分類演算法建立零件加工良率檢測模型，由分類報告的績效，探討人工智慧技術結合紡織設備製造產業進行品質管理與良率檢測的應用，並彙整分析資訊，協助廠商未來建立零件加工良率監控的演算法參考依據。

**2. 文獻探討**

**2.1 決策樹(Decision Tree)**

決策樹(Decision Tree)是一種應用於分類與預測問題的機器學習演算法(Song & Ying, 2015)，從大量的資料當中取得關鍵資訊，並在各個因素與決策當中找出關聯性，依照隨機性的決策規則與模型訓練深度向下分支葉節點，呈現出樹狀的結構，即可獲得分類結果。決策樹可應用在工業與傳統產業上，以塑膠產業為例，塑膠射出成型的製程良率監控透過決策樹演算法，更有效率找出影響塑膠射出成型的不良之可能因素(林宜慧, 2019)；再以當今國內的半導體產業為例，可經由半導體生產設備的資訊找出製程可能產生的缺陷，並使用決策樹找出缺陷與製程良率的關聯性(劉雅方, 2019)。

**2.2 隨機森林(Random Forest)**

隨機森林(Random Forest)屬於集成學習的演算法之一，可應用於分類跟迴歸的問題，原理是透過隨機分配的方式合併所有決策並整合為眾多子集合，且樣本間互相獨立，根據集成學習的概念，訓練方式為裝袋法(bagging)，三分之二的資料裝袋，稱為袋內樣本並進行訓練，剩餘樣本被稱為袋外樣本，當作驗證集執行交叉驗證來評估準確率(Biau & Scornet, 2016)。隨機森林可以用於分析品質的檢測，透過交叉驗證與模型的學習曲線找出影響磨耗特徵的重要因素(張瑞哲, 2019)；機台的故障診斷也可透過隨機森林分類找出重要的因素，S. Aravinth與V. Sugumaran兩位學者利用統計特徵掘取找出因子再利用隨機森林進行空壓機故障診斷(Aravinth & Sugumaran, 2018)、PLC機台的故障診斷可應用隨機森林進行檢測，皆經過準確率與分類績效進一步評估模型的可行性(Liu, Zhang, Wu, & Fu, 2020)。

**2.3 支持向量機(Support Vector Machine,SVM)**

支持向量機(SVM)適合解決小量樣本且維度較高的資料。支持向量機可分為線性支持向量機與非線性支持向量機(黃崇豪,2015)，本研究的支持向量機為線性支持向量機，線性支持向量機的核心觀念為:各個分類之間的差距需達到最大值，在進行訓練的資料中，若有一個超平面可將資料分為兩類，將超平面做區分，並要求樣本點符合線性支持向量機的定義。在實務上，工業生產製程報表的資料複雜度高，屬於高維度資料，即可搭配降維技術與支持向量機組成混和模型進行分類，來觀察影響良率的因素(Jing & Hou, 2015)；另一方面，支持向量機可進行製程監控與機台故障診斷，建立預警與良率分類模型同時發揮支持向量機分類與預測的功能(Yin & Hou, 2016)；支持向量機亦可結合特徵工程，將資料維度提升再進行特徵萃取，隨後透過支持向量機進行分類執行機台診斷(謝宜耿, 2019)。

* 1. **極限梯度提升(****eXtreme Gradient Boosting,XGBoost)**

極限梯度提升(XGBoost)是利用集成學習(ensemble learning)找到最佳解的監督式學習演算法，集成學習原先是由多個較弱的學習器結合而成，集成學習的運作模式分為Bagging和Boosting，經由極限梯度提升演算法有效地建構，得以讓集成學習的演算法運算能力更加出色，透過並行的方式解決分類和迴歸問題(Chen & Guestrin,2016)。極限梯度提升可應用在產品品質的檢測上，從影響鋁合金鑄造的特徵挖掘出主要因素，再透過多個機器學習模型進行評估，得知極限梯度提升的模型在品質檢測上有著良好的績效(Yan, Chen, Wang, & Liu, 2020)；極限梯度提升若應用在故障診斷上也能找出重要的因素，經由極限梯度提升模型的分類成果找出影響機台故障特徵的重要程度，以利後續故障診斷系統建立的參考依據(Duan, Li, Qiang, & Zhou, 2021)；利用極限梯度提升處理高維度的資料可以利用降維技術輔助建立混和模型，透過降維讓模型的訓練更有效率，實現液壓系統的故障診斷(Lei et al., 2019)。

**3. 研究方法**

本研究使用廠商提供共計500筆零件規格的資料表進行分析，起初執行探索性資料分析觀察資料，註記成品標籤，並計算不良率和允收機率，而後進行資料前置處理，最後運用分類演算法建立良率檢測模型。

**3.1 探索性資料分析**

利用敘述統計學與統計圖表觀察資料，在這分析過程中，主要觀察各規格(長、寬、高)的統計資訊，包括各個作業員製作零件的數量(如圖1所示)，並繪製規格的箱型圖找出離群值、直方圖觀察離群值的分佈情形(如圖2、圖3所示)，探索性資料分析主要目的用於對資料的初步認識，以利後續分析。

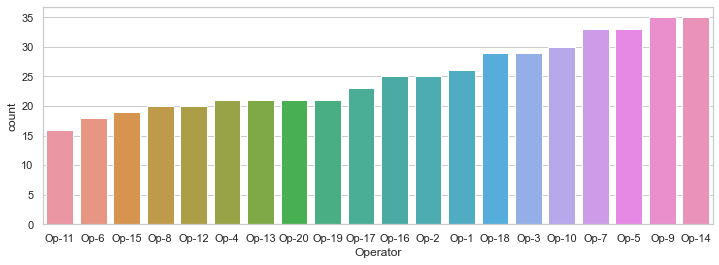


圖1. 各個作業員零件完成數量

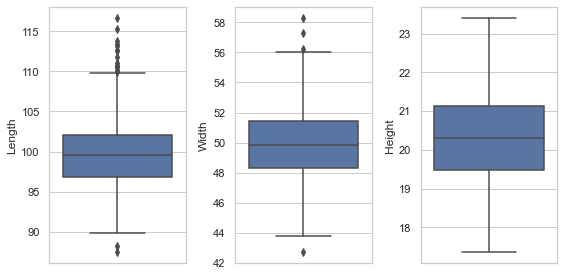


圖2. 各零件規格的箱型圖

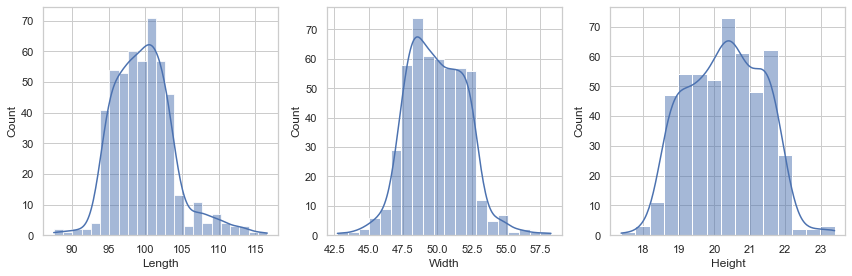


圖3. 各零件規格的直方圖

**3.2 零件品質註記**

在統計圖表的箱型圖中觀察，若有一數值與四分位距(IQR)的範圍相差1.5倍時，該數值為離群值，並將離群值視同缺陷，對應的工單註記為不良品，透過註記良品與不良品得知良品和不良品個數，得知生產不良品的作業人員和不良品的總產量為20件，經過逐一對應建立新的欄位為零件品質註記(quality)，並利用群組長條圖呈現每位作業員的工件數量和不良品生產數量，如圖4所示。

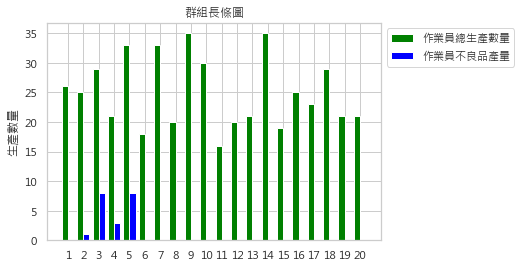


圖4. 作業員生產情形的群組長條圖

**3.3 良率和允收機率計算**

從探索性資料分析的結果，不良品數量為20件，其餘為480件良品，不良率為4%，接著根據隨機抽樣的理論求出允收機率，假定隨機抽樣20個零件，若有2個或更少的不良品，則允收這批貨物，此假設需透過二項分配進行探討，二項分配的母數:試驗次數(n)、成功機率(p)，而參數k代表抽樣的樣本數，在此案例中，試驗次數為20，成功機率即不良率4%，首先經由二項分配的累積機率函數，公式如(1)所示，求出所有組合的拒絕收貨機率，再利用可靠度函數的定義，100%減去累積機率，得知允收機率為95.61%。

(1)

**3.4 資料前置處理**

完成以上分析後，即可建立機器學習模型進行分類，在進行建模之前需先完成資料前置處理，包括:遺漏值檢查、資料不平衡處理、正規化。遺漏值清理確認資料當中是否有遺漏值，接著利用SMOTE方法進行零件品質註記欄位的資料不平衡問題改善，平衡資料後使用最大最小正規化方法，將零件規格的數值正規化為[0,1]的區間當中當作訓練集。

**3.5 建立分類演算法模型**

將經過正規化的訓練集分別帶入決策樹(Decision Tree)、隨機森林(Random Forest)、支持向量機(SVM)、極限梯度提升(XGBoost)分類模型訓練，模型建立時，針對各模型重要的超參數進行調整使分類演算法得以優化，重要的超參數如下表1所示，最後測試集帶入訓練完成的模型進行分類。

表1. 分類演算法與超參數

|  |  |
| --- | --- |
| **演算法** | **超參數** |
| Decision Tree | 剪枝參數(max\_depth=5,min\_samples\_split=5,min\_samples\_leaf=4) |
| Random Forest | Bagging參數(n\_estimators=90) |
| SVM | 懲罰係數(C=0.1)、內核係數(gamma=auto) |
| XGBoost | Booster參數(max\_depth=5,subsample=0.8) |

**4. 研究結果**

**4.1 資料集說明**

本研究使用500筆紡織設備零件加工資料進行分析，總共有6個欄位當作特徵，分別為:工號(item\_no)、零件長度(length)、零件寬度(width)、零件高度(height)、作業人員編碼(operator)、零件品質註記(quality)，工號為各工單的號碼；零件長度、零件寬度、零件高度等三個特徵是紡織設備零件經過焊接與組裝後進行測量的數值資料；作業人員編碼是公司的作業人員，為了整理與蒐集資料採取編碼處理；零件品質註記是透過探索性資料分析繪製箱型圖得知，判別出該工號的零件為不良品或良品。

**4.2 資料前置處理結果**

利用演算法分析之前，將原始資料進行前置處理，進行遺漏值檢查時，確認該份資料並無遺漏值，但發現不良品僅有20件，使零件品質註記欄位產生資料不平衡的問題，本研究利用SMOTE合成樣本解決零件品質註記欄位的資料不平衡問題，最後透過最大最小正規化將零件規格的數值正規化為[0,1]的區間，並當作演算法的訓練資料。

**4.3 分類報告**

將訓練資料帶入經過超參數調整的分類演算法進行訓練，接著將測試資料帶入訓練完成的模型進行分類，本研究的零件良率檢測是二元分類問題，透過機器學習的分類報告觀察分類問題的重要績效，包括:F1-Score、Precision、Recall。Recall的定義是在真實值當中有幾個是預測正確的，而Precision的定義為所有預測為真實值中，確定是真實值的比例，F1-score則是兩者的調和平均數，是一個概略指標來觀察分類績效。從分類報告得知，決策樹的模型應用在本研究廠商提供的資料集進行零件良率檢測可得到較佳的分類績效，如表2所示。

表3. 各模型的分類報告

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Decision Tree | Random Forest | SVM | XGBoost |
| Precision | 97.92% | 96.47% | 91.95% | 93.98% |
| Recall | 96.58% | 56.55% | 55.17% | 53.79% |
| F1-Score | 97.24% | 71.30% | 68.97% | 68.42% |

**4.4 研究結論**

為了評估客戶對於該廠商貨品的接受度，利用隨機抽樣理論與二項分配求出允收機率為95.61%。完成良率分析後，進行資料前置處理，確認資料集無任何遺漏值，利用SMOTE解決零件品質註記欄位的資料不平衡問題，再透過最大最小正規化將零件規格的數值正規化，以利加速模型的訓練並完成資料前置處理。接著調整各個分類演算法模型的超參數，完成調整超參數後進行模型訓練與測試。根據分類報告的結果，得知使用決策樹分類模型在該廠商的紡織設備零件加工良率檢測上可得到較佳的分類績效。在未來的研究當中，應可採取自動化超參數搜尋方法找出其他分類模型的最佳參數，透過優化不同演算法的超參數，讓分類的結果更加完善，或是利用其他演算法進行實驗，進一步協助廠商開發紡織設備零件良率檢測系統的演算法。

**參考文獻**

林宜慧，2019，*應用雙屬性節點分枝決策樹於塑膠射出成型製程績效之研究*，朝陽科技大學工業工程與管理系碩士論文。

張瑞哲，2019，*以資料探勘應用於銑削刀具磨耗之研究*，國立雲林科技大學工業工程與管理系碩士論文。

劉雅方，2019，*運用決策樹預測產品不良因素之研究-以半導體封裝廠為例*，中原大學工業與系統工程研究所。

謝宜耿，2019，*支持向量機暨特徵工程應用於中空滾珠導螺桿螺帽預壓螺桿預拉油冷系統及負載之診斷*，國立彰化師範大學機電工程學系碩士論文。

黃崇豪，2015，*使用支持向量機於二變量製程之非隨機同步管制圖樣式辨識*，國立雲林科技大學工業工程與管理系碩士論文

Aravinth, S., & Sugumaran, V., 2018, Air compressor fault diagnosis through statistical feature extraction and random forest classifier. Progress in Industrial Ecology, *an International Journal*, 12(1-2), 192-205.

Duan, R., Li, Y., Qiang, B., & Zhou, L., 2021, A Feature Selection-Based XGBoost Model for Fault Prediction. *Paper presented at the 2021 17th International Conference on Computational Intelligence and Security (CIS)*.

Jing, C., & Hou, J., 2015, SVM and PCA based fault classification approaches for complicated industrial process. *Neurocomputing*, 167, 636-642.

Lei, Y., Jiang, W., Jiang, A., Zhu, Y., Niu, H., & Zhang, S., 2019, Fault diagnosis method for hydraulic directional valves integrating PCA and XGBoost. *Processes*, 7(9), 589.

Liu, P., Zhang, Y., Wu, H., & Fu, T., 2020, Optimization of edge-PLC-based fault diagnosis with random forest in industrial Internet of Things. *IEEE Internet of Things Journal*, 7(10), 9664-9674.

Song, Y.-Y., & Ying, L., 2015, Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry*, 27(2), 130.

Yan, S., Chen, D., Wang, S., & Liu, S., 2020, Quality prediction method for aluminum alloy ingot based on XGBoost. *Paper presented at the 2020 Chinese Control And Decision Conference (CCDC)*.

Yin, Z., & Hou, J., 2016, Recent advances on SVM based fault diagnosis and process monitoring in complicated industrial processes. *Neurocomputing*, 174, 643-650.

Biau, G., & Scornet, E., 2016, *A random forest guided tour*. Test, 25(2), 197-227.

Chen, T., & Guestrin, C., 2016, Xgboost: A scalable tree boosting system. *In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).