

Machine Learning Model for CNC Machining Failure Prediction - an Analyst and
Comparison of Two Models Trained by the Same Data Set
以同一資料集訓練的兩個CNC 加工故障預測機器學習模型的分析與比較

摘要

本篇報告探討了使用機器學習模型進行CNC加工故障預測的重要性和應用。本報告和新是以同一CNC加工資料集為出發點，分析CNC加工故障數據的特徵與難題，並比較二不同模型在解決同一份資料中所提出的CNC加工故障預測所提出的思路。報告介紹了兩個不同的模型，分別是模型一和模型二。模型一使用無監督式學習配合多變量統計過程控制以預測CNC加工過程中的故障。模型二則引入了重播的概念使模型能夠持續學習。報告還討論了CNC加工故障預測的挑戰，包括特徵漂移和資料擷取窗長度的設計。這兩個問題分別可由模型二與模型一解決。模型二提出每當遇到新數據時重播20%的過去數據可以保持模型效能。模型一則是將故障的檢測和定位皆納入模型的設計當中。最後，報告強調了將模型一與模型二合併以作為CNC加工故障預測機器學習模型的可能性。

前言

CNC(Computer Numerical Control)加工是一種自動化製造過程，通過控制車床、銑床等機床的運動和工具的操作來加工工件至所需外型。當CNC加工發生故障時，該故障發生期間所製造的零件會無法滿足設計規範，必須重工或丟棄，造成原料與時間甚至於加工機台本身的耗損。這些耗損會給工廠帶來額外的成本。CNC加工故障原因可能受到許多參數的影響，例如刀具磨損、顫振、刀具夾緊不當、切屑堵塞、幾何形狀、材料、冷卻液的使用和許多其他物理條件。CNC加工中的故障預測是提高製造效率、減少停機時間和提高產品品質的關鍵步驟。CNC加工故障預測機器學習模型的應用可以幫助製造業提高生產效率，減少停機時間和成本，並提高生產品質。此外，它還可以提高工人的安全性，因為它可以幫助識別潛在的危險情況，並減少意外事故的風險。透過實施先進的監控和預測性維護技術，CNC擁有者可以最大限度地減少意外機器故障並優化生產流程。

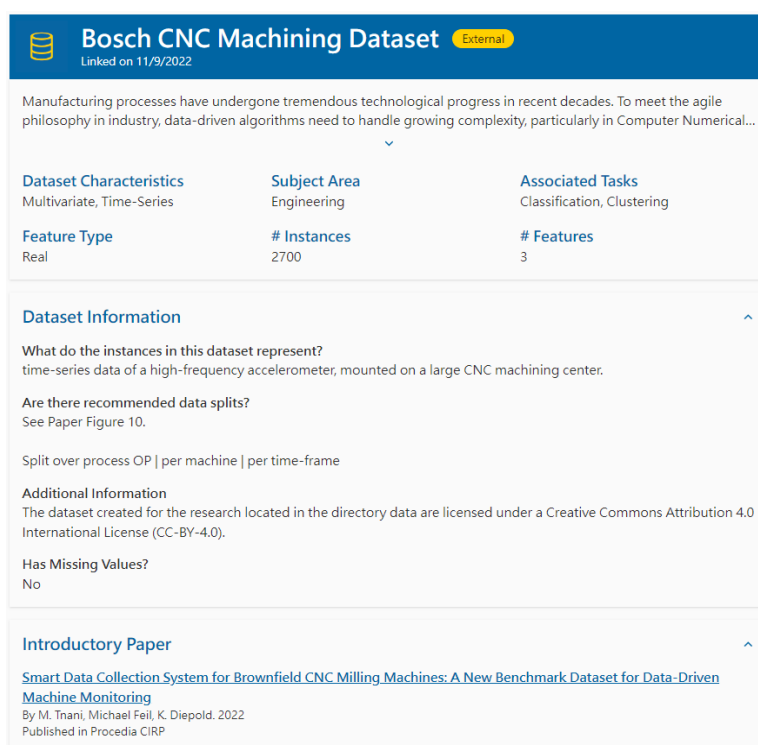
為了提高CNC加工的效率 and 可靠性，可以使用的其中一種預測性維護技術乃是使用機器學習模型來預測可能出現的故障，進行預防性維護，從而減少生產中斷和成本。傳統上，數控工廠的故障預測是透過個人經驗和重複試錯的方法來完成的。有時一些工廠甚至沒有故障預測行動，只是簡單地使用刀具直到其在CNC機床上斷裂，並透過加工後的檢查流程控來制零件品質。其背後的原因是因為可能影響故障的參數過多使實務上難以快速簡便的以量化方式預測一加工手續的好壞。CNC加工中故障預測可能的解決方案是使用數據驅動的加工學習方法。機器學習模型可以為工廠管理提供資料支援的決策，以優化設備運作和生產流程並在故障發生之前預測問題。另外在機器學習模型訓練完成後，其能夠以極快的速度發出預測，使加工者能夠集時監測CNC加工過程，快速偵測異常情況並觸發警報，以便操作人員可以立即採取行動。而理想的機器學習

模型可以不斷學習和適應新數據，以應對新的故障模式和變化的工作環境，從而提高預測準確性並應用於多種狀況。

CNC 加工故障預測機器學習模型的建立與訓練離不開實際CNC加工數據的支持。這些數據可能包含但不限於機床運動、震動數據、工件尺寸和質量、刀具的狀態和性能數據等。這些數據可以通過傳感器和監控系統等方式來獲取。本篇報告將分析一公開的「博世數控加工資料集」(UCI Machine Learning Repository)並比較兩個以此資料集為基礎所訓練的CNC 加工故障預測機器學習模型。此資料集為解決數控加工中的故障預測問題提供了全面的數據。資料範圍廣泛且不限於少數機器和工具，這表明這些資料訓練的最終模型可以具有更好的可擴展性。據資料集所有者稱，該資料集的目標是「透過提供一個包含阻礙製造環境中機器學習演算法可靠性的主要挑戰的資料集，來增強機器學習在現實應用中的可擴展性 (Tnani et al.)」。可以基於此選定的資料集來訓練用於預先預測 CNC 加工故障的機器學習模型。

資料集簡介與初步分析

本報告使用了 UCI 機器學習儲存庫中的「Bosch CNC Machining 資料集」(UCI Machine Learning Repository)。圖1是UCI Machine Learning Repository網頁上此資料集的介紹截圖。



Bosch CNC Machining Dataset External
Linked on 11/9/2022

Manufacturing processes have undergone tremendous technological progress in recent decades. To meet the agile philosophy in industry, data-driven algorithms need to handle growing complexity, particularly in Computer Numerical...

Dataset Characteristics	Subject Area	Associated Tasks
Multivariate, Time-Series	Engineering	Classification, Clustering
Feature Type	# Instances	# Features
Real	2700	3

Dataset Information

What do the instances in this dataset represent?
time-series data of a high-frequency accelerometer, mounted on a large CNC machining center.

Are there recommended data splits?
See Paper Figure 10.

Split over process OP | per machine | per time-frame

Additional Information
The dataset created for the research located in the directory data are licensed under a Creative Commons Attribution 4.0 International License (CC-BY-4.0).

Has Missing Values?
No

Introductory Paper

[Smart Data Collection System for Brownfield CNC Milling Machines: A New Benchmark Dataset for Data-Driven Machine Monitoring](#)
By M. Tnani, Michael Feil, K. Diepold. 2022
Published in Procedia CIRP

圖 1: UCI 機器學習儲存庫網站「Bosch CNC Machining Dataset」的螢幕截圖 (Tnani)

資料儲存在 Github 上，可以透過連結「github.com/boschresearch/CNC_Machining」進行存取 (Tnani)。該數據集包含來自 3 台不同 CNC 機器的真實生產環境的振動數據，這些機器在特定時間段內對鋁進行 14 種不同的工具操作。振動的方向在 x、y 和 z 軸上。考量未來的可擴充性與資料取得的難易度，原始數據乃是由安裝在每台 CNC 工具機主軸外殼後端的 Boch CISS 感測器收集。每組振動數據根據該組數據是在機器正常運作時還是在異常情況下記錄，都標有OK或NOK。以上為該資料集內收集原始資料的方式。收集到的原始資料會經過挑選後錄入資料集內。工業資料集中的一個常見挑戰

是標記為OK的數量遠多於標記為NOK的數量。為提供一個典型的數據集，資料集作者從不同時間段中選擇了合理數量的正常過程，以減少類別不平衡。

該資料集作者點出了幾項分析此資料時發現可能對預測模型造成影響的挑戰。首先是機器和時間之間的特徵漂移。除了製程故障之外，還會發生一些只有在機器維護後才能偵測到的條件異常會造成NOK。這些異常主要是由部件磨損、液壓問題、不正確的設定等引起的。然而，在達到關鍵階段產生NOK之前，此種損耗會隨著時間的推移而出現輕微的惡化/變化，從而導致振動數據中出現額外的噪音。這會導致 OK 類別在不同時間範圍之間發生漂移。除了老化漂移之外，機器和機器部件的狀況之間的差異也增加了實際應用中的挑戰。這造成了模型泛化的困難。在此之上，為了實現快速處理和非侵入式解決方案，時間序列訊號通常以固定長度加窗擷取分段。對NOK資料進行分段的缺點是小段的標籤可能與完整的過程不對應。為了解決這個問題需要定義一個合理的訊息窗長度。

模型一使用方法論述

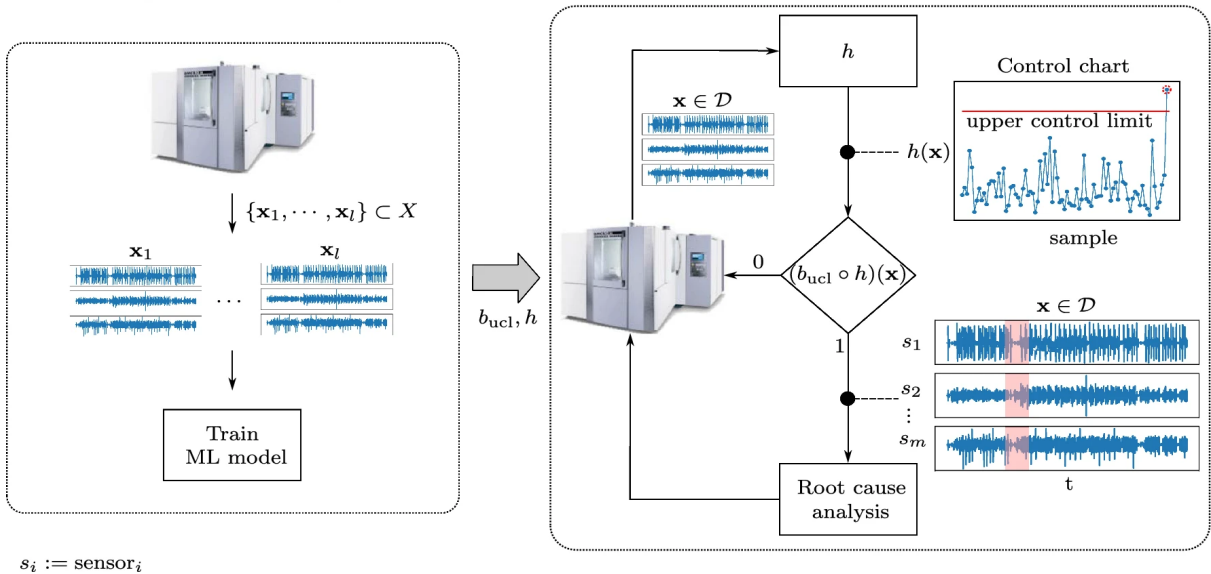
模型一的作者Biegel et al. 提出CNC 加工故障預測模型可用圖2表示。首先以已知之資料集訓練一ML模型a模擬出函數 h 與 b_{ucl} ，再將模型a投入於實際情形b中。當一未知訊號 x 進入時，模型會記錄該 $h(x)$ 值，如果檢測到異常過程條件，即

$$(b_{ucl} \circ h)(x) = 1$$

，模型會在時間序列中定位異常過程條件，以支持機器操作員進行根本原因分析。

(a) Train ML model on normal process data
(phase I monitoring)

(b) Detection and localization of anomalous process condition
(phase II monitoring)



$s_i := \text{sensor}_i$

圖 2: 模型一提出之CNC 加工故障預測模型概念

模型一為Biegel et al. 提出的自監督式學習SSMSPC模型 (Biegel et al.)。自監督學習是一種無監督式學習(unsupervised learning)的形式，它通過解決一個有監督的前提任務來從無標籤的數據中學習有效的標示。基於自監督學習的異常檢測方法可以直接從前提任務中得出異常分數，或者在下游的異常檢測任務中使用學習到的標示。因此，定義合適的前提任務是自監督學習方法的一個重要組成部分。SSMSPC是一種基於自監督學習的多變量統計過程控制(MSPC)的新方法，旨在檢測和定位離散製造過程中的異常過程行為。SSMSPC的動機是利用無監督表示學習的潛力，將自監督學習納入一般的統計過程控制(SPC)框架中，以開發一種全面的方法，用於在運行過程中作為過程監

測工具。SSMSPC提出了一種前提任務，稱為位置+變換預測，其中目標是分類給定時間序列輸入的增強類型和位置。增強是通過在一個隨機選擇的窗口中應用一個預定義的增強函數來實現的。增強後的時間序列數據通過連續小波轉換(CWT)來計算其標量圖表示，然後作為一個多通道圖像輸入到一個LeNet型編碼器網絡中，該網絡有兩個預測頭，分別用於預測增強類型和位置。在前提任務完成後，SSMSPC在學習到的表示上構建一個單類分類器，使用Hotelling's T-squared distribution T^2 統計量作為異常分數。控制限是基於核密度估計(KDE)來擬合的。此外，SSMSPC還提出了一種對傳統控制圖視圖訊號擷取窗的擴展，該視圖結合了元數據和學習到的表示，以將過程數據分段為各個過程步驟，並突出顯示異常的時間步驟，從而幫助機器操作員進行根本原因分析。模型一的架構可用圖3表示。

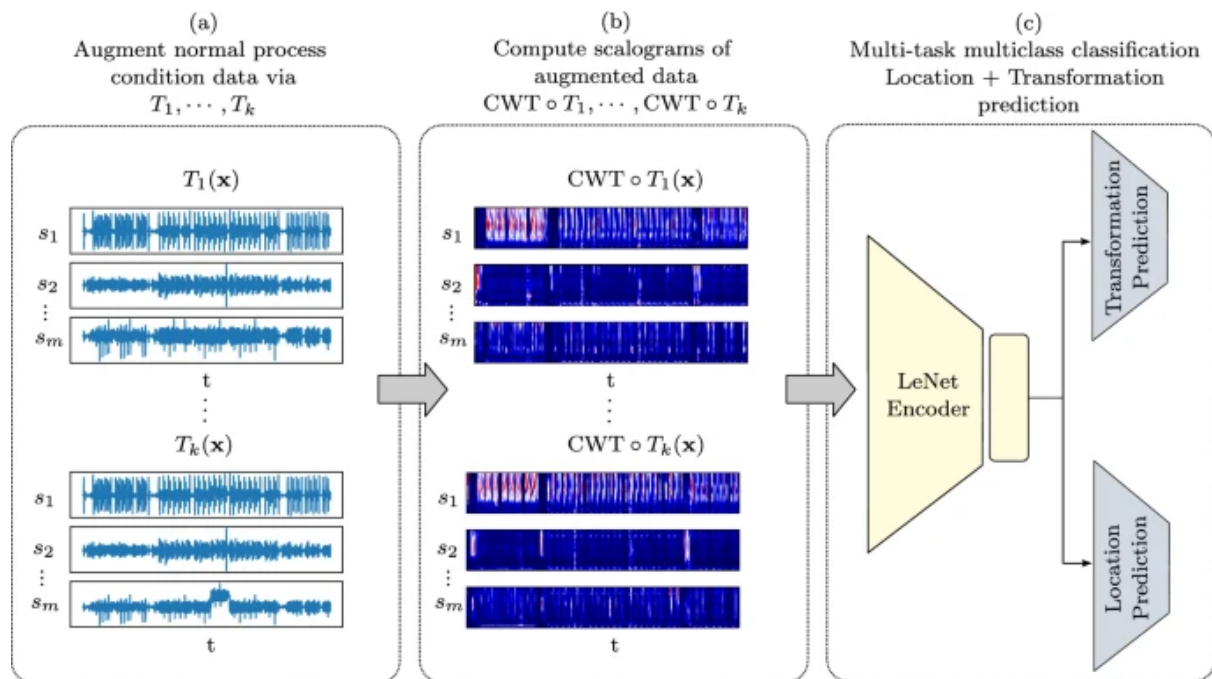


圖3. 模型一組成圖

模型二使用方法論述

模型二為Sen et al. 提出的重播驅動持續學習模型(Replay-Driven Continual Learning) (Sen et al.)。作者提出了一種重播驅動的持續學習管道，該管道架構為先訓練出一基本模型並隨後根據新數據對其進行更新與重新學習，從而最大限度地減少對過去的災難性遺忘，指新經驗覆蓋先前的經驗。模型二所提出之CNC 加工故障預測模型概念如圖4所式。

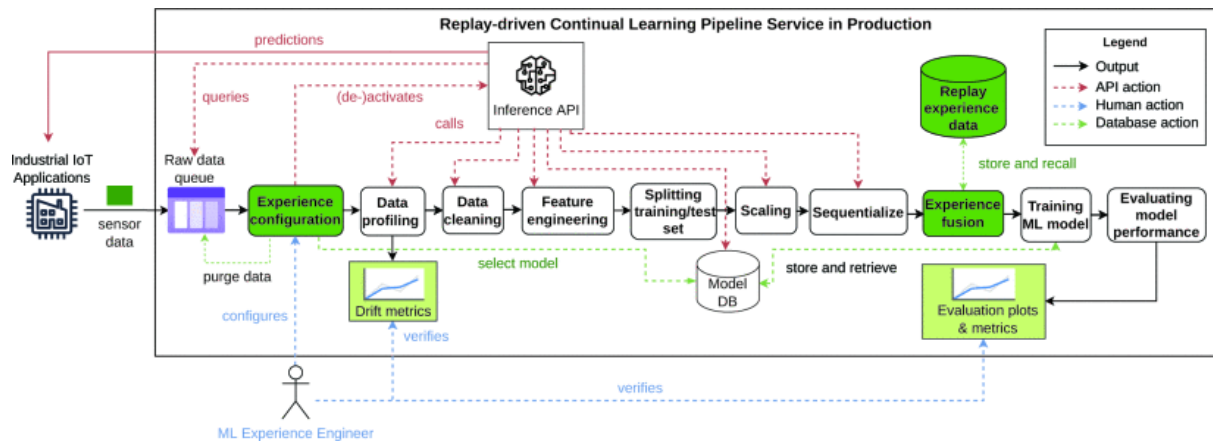


圖4.模型二提出之CNC 加工故障預測模型概念

各種方法(技術)比較

模型一的作者將其模型應用於Bosch CNC Machining 資料集的結果取得了100%的AUROC。首先作者將每台機器的數據根據加工步驟(OP)合併，從而就好像數據來自一台假想的機器一樣。這樣做的原因是為擴大可用數據量。在下一步中，作者將每個 OP 中的數據裁剪為完全相同的長度。這是通過切斷相應 OP 類中超過最小 OP 長度的所有數據點來完成的。經模型預測後，SSMSPC能夠識別異常過程模式，從而為機器操作員找到異常過程條件的根本原因提供有效支援。

模型二的作者將其模型應用於Bosch CNC Machining 資料集的結果發現基本模型在根據新數據進行評估時會隨著時間的推移導致表現不斷惡化，需要不斷學習來更新模型以避免模型效能下降。這符合Bosch CNC Machining 資料集所分析的老化漂移情況。作者所提出的模型引入了重播的概念來重新訓練其模型以解決老化漂移。論文表示20%的重播，即重新訓練時新資料佔總資料80%，足以在持續學習中保持良好的表現，從而最大限度地減少災難性遺忘。

由於兩篇論文所提出的模型評估方式不同，模型一與模型二難以量化比較其優劣。同時設計模型一與模型二時的出發點也略有出入。模型二著重在對模型泛化和資料偏移的優化處理。而模型一則專注於提升在預測CNC 加工故障時的精準程度。模型一提到了使用連續小波轉換與訊號的合併等方式來進行資料前處理，但模型二並無詳細提到其資料處理方式和模型細部參數。

結論與心得

Bosch CNC Machining 資料集提出的資料飄移與訊息窗擷取設定分別在模型二與模型一提出了解決方式。模型一以自監督式學習配合多變量統計過程控制來檢測和定位故障的發生。模型二則是將新資料與舊資料按比例抽出混合來重新訓練模型避免預測結果隨時間推移變得越來越不准。但模型二中如何選取新舊資料、如何界定何時該重新訓練模型、重新訓練模型的成本等議題依然有待研究。如果將模型一與模型二的兩篇論文結合起來看，將模型一作為模型二的基本模型並在其之上引入模型二的重播模式似乎為建立CNC 加工故障預測機器學習模型的良好選項。

透過這次作業我了解到了預測CNC 加工故障的難題與障礙，也對如何解決這些問題有了更深一層的認識。在閱讀的過程中，模型一的論文非常詳細的介紹了其模型的架構與細部參數。其中涉及了集合論、統計學、訊號處理等多領域的知識，坦白講我也沒有完全看懂，也難以分析出為何作者要如此處理這些資料和這樣設計模型。相反的，

模型二的論文對工業物聯網 IIoT和持續學習的機器學習流程提供了從頭到尾的概述，但對其論文中實驗所提到的模型卻並未詳細提及。因此也難以對其進行系統性的分析。對我而言，這次作業最大的收穫乃是了解到機器學習的問題是需要多方面的背景知識來完成的，絕非簡單的套模與改參數。對於以後想望機器學習和資料科學走的我而言，目前還缺乏真正了解各種模型原理的知識，同時若想要將機器學習實際應用於工程問題上，本科的專業領域也不能落下。

引用文獻

Biegel, Tobias, et al. "SSMSPC: self-supervised multivariate statistical in-process control in discrete manufacturing processes." *Journal of Intelligent Manufacturing*, 2023, <https://doi.org/10.1007/s10845-023-02156-7>.

Sen, Sagar, et al. "Replay-Driven Continual Learning for the Industrial Internet of Things." *2023 IEEE/ACM 2nd International Conference on AI Engineering–Software Engineering for AI (CAIN)*. IEEE, <https://doi.org/10.1109/CAIN58948.2023.00014>.

Tnani, Mohamed Ali. "boschresearch/CNC_Machining: data set for process monitoring on CNC machines." *GitHub*, 2023, http://github.com/boschresearch/CNC_Machining. Accessed 4 November 2023.

Tnani, Mohamed Ali, et al. "Smart Data Collection System for Brownfield CNC Milling Machines: A New Benchmark Dataset for Data-Driven Machine Monitoring." *55th CIRP Conference on Manufacturing Systems*, 2022, <https://doi.org/10.1016/j.procir.2022.04.022>.

UCI Machine Learning Repository. "Bosch CNC Machining Dataset." 9 March 2019, <https://archive.ics.uci.edu/dataset/752/bosch+cnc+machining+dataset>. Accessed 4 November 2023.