

人工智慧產業實務 —— 刀具磨損分類模型

Team 4

408346054 蘇芮萱

408346119 林文萱

409346025 張嘉真

409346063 林威廷

指導老師：葉承達 副教授

目錄

一、	問題描述.....	2
1.	資料集簡介.....	2
2.	資料集內容.....	3
3.	目標問題.....	3
二、	研究方法.....	4
1.	資料預處理.....	4
2.	資料分析方法.....	4
三、	分析結果與討論.....	6
1.	分析結果.....	6
2.	三種算法各衡量指標比較.....	8
四、	參考文獻.....	9

一、 問題描述

1. 資料集簡介

使用 Kaggle 上的資料集：[CNC Mill Tool Wear](#)。此資料集為透過 CNC milling machine 中提取出來的銑床加工數據。

該資料集中有 18 個不同實驗之數據資料集(experiment.csv)，每個實驗中包含了特定時段內機器移動的座標軌跡等。除此之外，還有一資料集內容是針對這 18 個實驗最後狀態的統整 (train.csv)，其中包含：實驗編號、材料(蠟)、進料速度和夾緊壓力、工具狀況(未磨損和磨損的工具)以及該工具是否通過外觀檢查。本資料集可用於：

- (1) 刀具磨損之監督式分類
- (2) 夾緊壓力不足所導致需要人工檢驗的情況預測

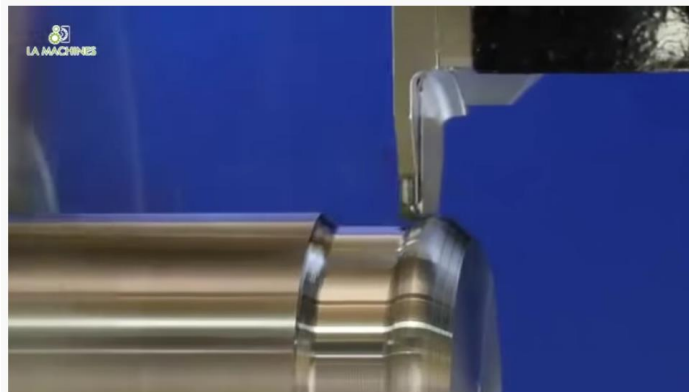


圖 1 Modern High Speed CNC Lathe Machine Working

資料來源：<https://www.youtube.com/watch?v=jF4F8Zr2YO8&t=25s>

2. 資料集內容

(1) train.csv

包含 7 個欄位，其中前四個欄位為 input feature，其餘三個欄位為 output。

表 1 train.csv 的資料欄位

INPUT FEATURE	OUTPUT FEATURE
實驗編號(No)：共有 18 次實驗	刀具狀況(Tool Condition)：刀具是否磨損
材料(Material)：固定為 Wax	加工(Machining Finalized)
進料速度(Feed Rate)：刀具切割的速率	人工檢查(Passed Visual Inspection)
夾緊壓力(Clamp Pressure)	

在此問題中我們使用了刀具狀況(Tool Condition)、進料速度(Feed Rate)、夾緊壓力(Clamp Pressure)這三個特徵。

(2) experienment_1.csv - experienment_18.csv

每個檔案裡包含 48 個欄位，紀錄了某段時間內機器移動的軌跡。其中有 X、Y、Z、S、M 五種軸向的紀錄資料，並包含了運動相關的位置、速度、加速度、負載力大小、驅動器負載電流等資訊。

3. 目標問題

目標函數為 train.csv 中的 Tool Condition(刀具狀況)，用以判斷何時須更換銑床上的刀具，進行刀具磨損之監督式分類，分為 worn(磨損)與 unworn(未磨損)。

表 2 決策樹優缺點

優點	<ul style="list-style-type: none"> • 決策樹易於理解、解釋和提取出規則，且可視化分析 • 可以同時處理標籤型和數值型數據 • 比較適合處理有缺失屬性的樣本 • 能夠處理不相關的特徵 • 測試數據集時，運行速度比較快
缺點	<ul style="list-style-type: none"> • 容易發生過擬合 • 容易忽略數據集中屬性的相互關聯 • 對於那些各類別樣本數量不一致的數據，在決策樹中，進行屬性劃分時，不同的判定準則會帶來不同的屬性選擇傾向

(2) 支援向量機 (SVM)

SVM 是一種基於統計學習理論基礎的機器學習模型，針對小樣本、非線性、高維度與局部最小點等問題具有相對的優勢。簡單說就是找到一個決策邊界 (decision boundary) 讓兩類之間的邊界 (margins) 最大化，使其可以完美區隔開來。

表 3 支援向量機優缺點

優點	<ul style="list-style-type: none"> • 可以解決高維問題，即大型特徵空間 • 能夠處理非線性特徵的相互作用 • 無局部極小值問題 • 無需依賴整個數據 • 泛化能力比較強
缺點	<ul style="list-style-type: none"> • 當觀測樣本很多時，效率並不是很高 • 用 SVM 解決多分類問題存在困難 • 對於核函數的高維映射解釋力不強，尤其是徑向基函數 • 對缺失資料敏感，對引數和核函式的選擇敏感

(3) 邏輯斯回歸 (Logistic Regression)

邏輯斯回歸為平滑的曲線，是解決工業規模問題最流行的算法，核心思想是利用現有數據對分類邊界建立回歸方程，以此進行分類。回歸可以理解為最佳擬合，是一種選擇最優分類的算法。

表 4 邏輯斯回歸優缺點

優點	<ul style="list-style-type: none">• 適合需要得到一個分類概率的場景• 分類時計算量非常小，速度很快，存儲資源低• 可以結合 L2 正則化解決多重共線性問題• 廣泛的應用於工業問題上
缺點	<ul style="list-style-type: none">• 當特徵空間很大時，邏輯回歸的性能不是很好。• 不能很好地處理大量多類特徵或變量• 對於非線性特徵，需要進行轉換• 只能處理兩分類問題(在該基礎上衍生出來的 softmax 可以用於多分類)，且必須線性可分• 容易欠擬合(Underfitting)

三、 分析結果與討論

1. 分析結果

(1) 決策樹

使用基尼係數(Gini)作為資訊量的超參數，另外因此資料集數量較少，樹最高(max_depth)設為 2 以避免 Overfitting。

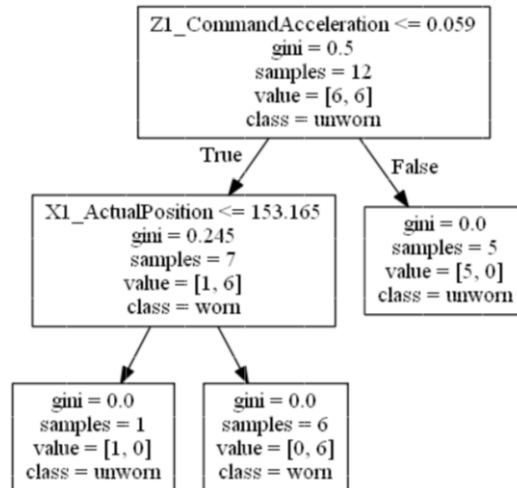


圖 4 決策樹視覺化結果

表 5 決策樹混淆矩陣

混淆矩陣		預測值	
		Unworn (0)	Worn (1)
真實值	Unworn (0)	2 (TN)	0 (FP)
	Worn (1)	3 (FN)	1 (TP)

(2) 支援向量機

其中 SVM 的超參數包括：decision_function_shape、kernel、C、gamma 等，在手動嘗試了不同的 Kernel function 後，選擇了使用預設的 kernel RBF。

表 6 SVM 中不同 Kernel 的 Accuracy 表現

Kernel	Linear	Polynomial	RBF	Sigmoid
Accuracy	0.5	0.3333333	~0.8333333	~0.6666666

表 7 SVM 混淆矩陣

混淆矩陣		預測值	
		Unworn (0)	Worn (1)
真實值	Unworn (0)	1 (TN)	1 (FP)
	Worn (1)	0 (FN)	4 (TP)

(3) 邏輯斯回歸

表 8 邏輯斯回歸混淆矩陣

混淆矩陣		預測值	
		Unworn (0)	Worn (1)
真實值	Unworn (0)	1 (TN)	1 (FP)
	Worn (1)	1 (FN)	3 (TP)

2. 三種算法各衡量指標比較

表 9 三種算法在準確率、精確率、召回率與 F1-score 的比較

	Decision Tree	SVM	Logistic Regression
Accuracy	0.5	0.83	0.67
Precision-P $\left(\frac{TP}{TP+FP}\right)$	1	0.8	0.25
Recall-R $\left(\frac{TP}{TP+FN}\right)$	0.25	1	0.75
F1-score $\left(\frac{2PR}{P+R}\right)$	0.4	0.888	0.375

就準確度(Accuracy)而言，以 **SVM** 的結果最好。召回率(Recall)是實際上有磨損的被預測出有磨損的比例，精確率(Precision)則是預測為有磨損實際上有磨損的比例。但是在刀具磨損的分類問題中，已經被磨損的刀具若是仍被判定為未磨損，可能導致銑床機器受損，而提早停止使用「可能耗損的」刀具也會導致刀片的浪費（陳健立，2012）。因此在這個問題中，我們更應該參考召回率(Recall)的結果，此時可以發現也是 **SVM** 比較符合期待。

但若是同樣看重召回率與精確率，可以使用 F1-score 進行衡量。此時則可以發現，**SVM** 的表現最好。我們未來或許能透過調整兩個種算法的超參數以達該算法的最佳預測結果，或者是增加資料量讓預測結果更具參考價值。

四、 參考文獻

陳健立，2012。以電腦視覺畫面鑑定車削刀片等級之技術研究。朝陽科技大學工業工程與管理系碩士論文。

知乎，2018。常用機器學習常用算法優點及缺點總結。取自

<https://zhuanlan.zhihu.com/p/36928215>。搜尋日期：2021 年 6 月 3 日。

Chao, CM., Yu, YW., Cheng, BW. et al. 2014 Construction the Model on the Breast Cancer Survival Analysis Use Support Vector Machine, Logistic Regression and Decision Tree. *Journal of Medical Systems*, 38, 106.

Daniel, F.H., & Bernd M. 2019. Tool wear monitoring of a retrofitted CNC milling machine using artificial neural networks, *Manufacturing Letters*, 19, 1-4.

Data School. 2014. Simple guide to confusion matrix terminology.

<https://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/>

Access June 3, 2021.