



應用機器學習演算法於剖面資料之監控

指導教授：鄭春生 教授 學生：郭品誌 黃昱翔 羅健庭

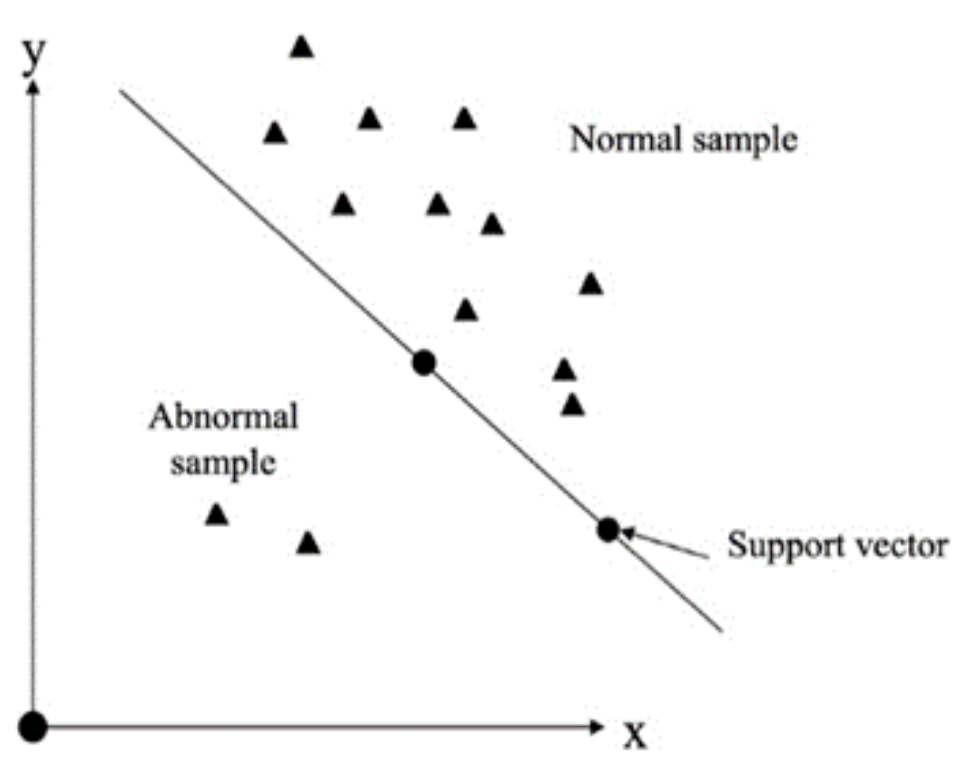
研究動機與目的

比較	監控對象	數據分析	應用範圍	目標
傳統 SPC	單一變數	管制圖	生產線上的單一產品或過程	保持單一變數的穩定性
剖面監控	多個變數	多變量變異數分析	多變數系統	維持整個剖面或多變數系統的品質

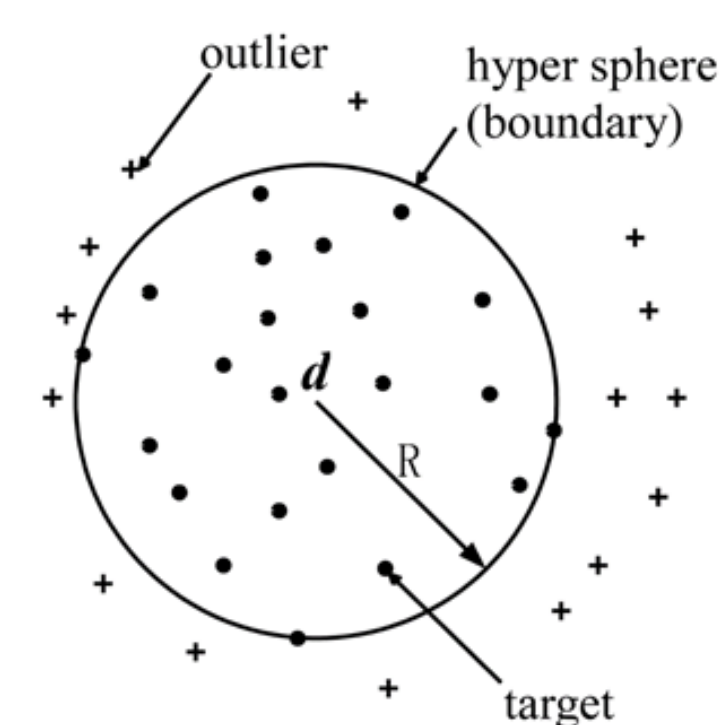
在網路快速發展下，我們獲取資料的速度變得比以往更快且更多資料，剖面 (profile) 是指透過一個或多個獨立自變數之間的函數關係來表達產品的品質特徵，傳統管制圖與剖面監控不同之處在於，前者是用來判斷製程分布參數是否改變，後者是判斷函數關係是否改變，表為傳統管制圖與剖面監控比較。

剖面監控 (profile monitoring) 分為非線性剖面 (nonlinear profile) 監控和線性剖面監控，是一種用於監測、測量和記錄特定區域或系統的性能、狀態或數據的方法，在品質管理中常用於異常檢測，剖面監控可用於檢測產品或過程中的異常情況。

研究方法



傳統 SVM 和 OCSVM 之間的一個重要區別。傳統的 SVM 通常用於二元分類問題，目標是找到一個超平面，使得不同類別的資料點與這個超平面之間的邊距 (margin) 最大化。相較之下，OCSVM 則是在非監督式情況下工作，其目標是尋找一個超平面，能夠將資料集中的正常類別資料與原點 (origin) 分開，同時盡可能靠近這些資料點。這個超平面被認為是正常資料所在的區域的邊界，並且它與原點的距離越遠越好，以確保它能有效地識別異常值，概念如圖 2 所示在建立分類器時，通常使用 RBF 核函數，在資料集密集區域周圍，擬合一個非線性邊界，將其餘資料點作為異常值進行分離。



深度支援向量數據描述 (deep support vector data description, Deep SVDD) 是一種半監督式學習的演算法，為一種單類別的分類模型，用於執行異常檢測的階段 II 任務。SVDD 的想法與 OCSVM 有關，它是建立一個超球體 (hypersphere) 來將資料分開，而非 CSVM 所使用的超平面 (hyperplane)。SVDD 試圖產生一個高維度空間的超球面，以最小的半徑將訓練集中的資料包起來。新來的待測資料映射到高維度空間後，如果落在這個超球面內，則認為它屬於 target class，否則就認為它是一個 outlier

真陽性 (true positive, TP)：表示正確的預測出陽性。

真陰性 (true negative, TN)：表示正確地預測出陰性。

偽陽性 (false positive, FP)：實際為陰性卻被誤判為陽性，又稱為型 I 錯誤 (type I error)。

偽陰性 (false negative, FN)：實際為陽性卻被誤判為陰性，又稱為型 II 錯誤 (type II error)。

		Predicted	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

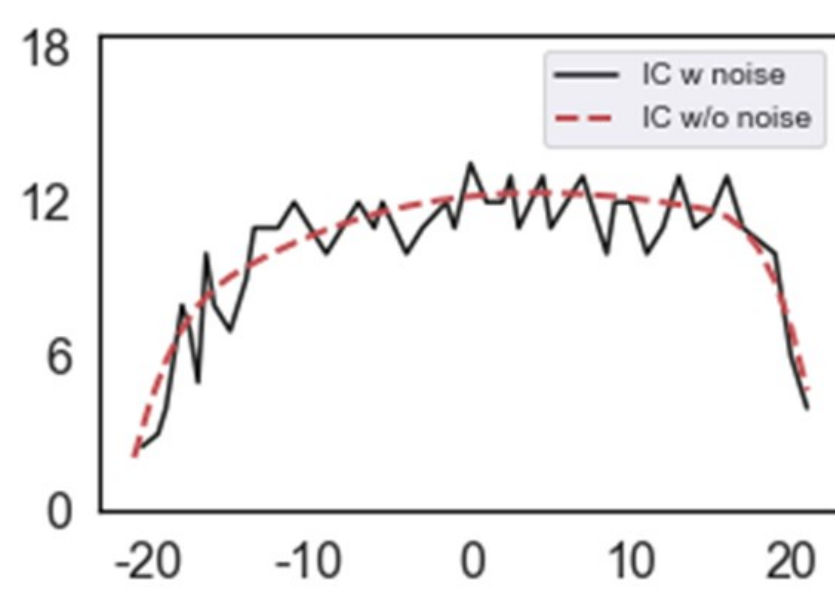
$$F_1 \text{ score} = \frac{2TP}{2TP + FP + FN}$$

$$\text{recall} = \frac{TP}{TP + FN}$$

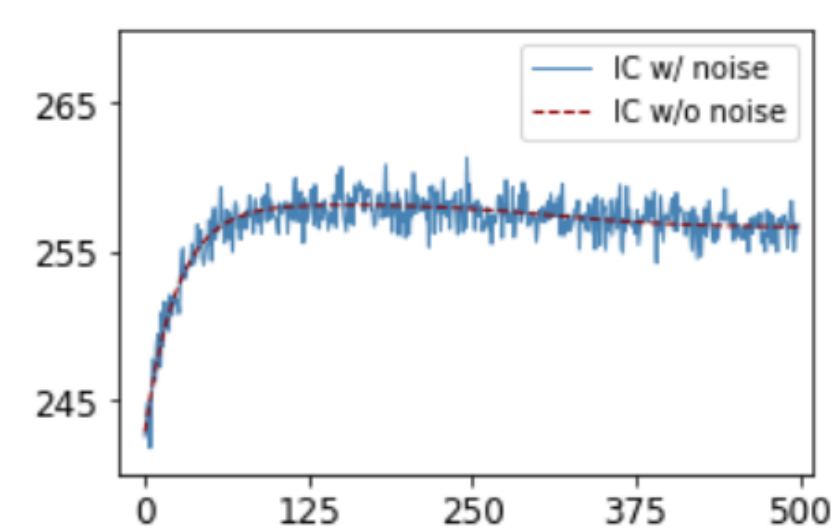
$$F_2 \text{ score} = \frac{5TP}{5TP + FP + 4FN}$$

$$\text{precision} = \frac{TP}{TP + FP}$$

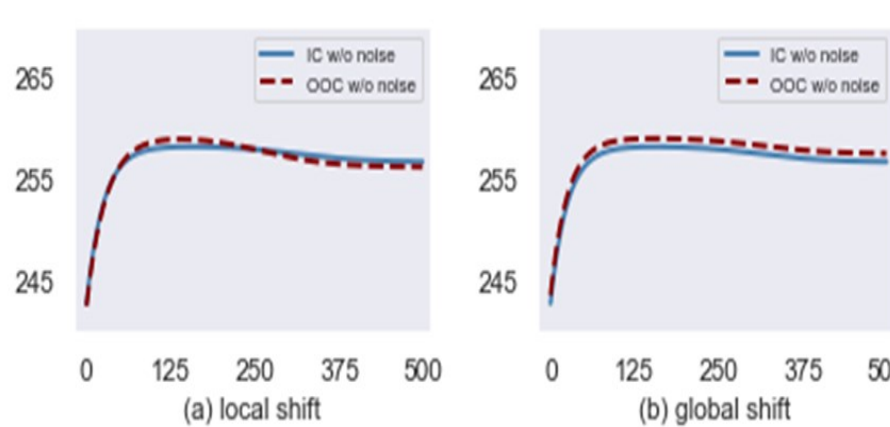
研究結果



在本研究中，我們採用非線性剖面作為數據集。第一種非線性剖面是參考 John and Agarwal (2019) 之研究其剖面函數如公式所示，並對此函數進行微調，增加了隨機雜訊，其平均數為 0，標準差為 1。圖 5 中，紅色虛線為作者提出的擬和樣條模型之示意圖。



Jensen et al. (2016) 所使用的參數設定，其中 $\epsilon \in [1, 1500]$ 亦即一個剖面包含 500 個資料點 (相當於樣本資料個數 $n=500$)。



- 局部偏移：本研究調整 5 種參數 θ_i 作為異常剖面，詳細的參數說明如下： $\theta_1=256.1$ 、 $\theta_2=0.065$ 、 $\theta_3=0.035$ 、 $\theta_4=259.2$ 、 $\theta_5=262.9$ 。
- 整體偏移：本研究將剖面整體偏移量設為 0.8 此相當於在公式 (13) 中加入常數 0.8。

表 1. $\nu=0.05$ 下的數據結果表						
		F1 分數		F2 分數		
訓練樣本	包含異常剖面	包含異常剖面	包含異常剖面	包含異常剖面	包含異常剖面	包含異常剖面
演算法 γ	1/43	1/129	1/215	1/43	1/129	1/215
OCSVM	0.73285	0.7245	0.72566	0.65848	0.63317	0.63374

表 2. $\nu=0.04$ 下的數據結果表						
		F1 分數		F2 分數		
訓練樣本	包含異常剖面	包含異常剖面	包含異常剖面	包含異常剖面	包含異常剖面	包含異常剖面
演算法 γ	1/43	1/129	1/215	1/43	1/129	1/215
OCSVM	0.7474	0.7403	0.7415	0.6774	0.6529	0.653

表 3. $\nu=0.015$ 下的數據結果表						
		F1 分數		F2 分數		
訓練樣本	包含異常剖面	包含異常剖面	包含異常剖面	包含異常剖面	包含異常剖面	包含異常剖面
演算法 γ	1/43	1/129	1/215	1/43	1/129	1/215
OCSVM	0.8055	0.7974	0.798	0.7594	0.7294	0.7288

表 4. $\nu=0.005$ 下的數據結果表						
		F1 分數		F2 分數		
訓練樣本	包含異常剖面	包含異常剖面	包含異常剖面	包含異常剖面	包含異常剖面	包含異常剖面
演算法 γ	1/43	1/129	1/215	1/43	1/129	1/215
OCSVM	0.8555	0.8563	0.8569	0.8362	0.813	0.8119

表 5. $\nu=0.000$ 下的數據結果表						
		F1 分數		F2 分數		
訓練樣本	包含異常剖面	包含異常剖面	包含異常剖面	包含異常剖面	包含異常剖面	包含異常剖面
演算法 γ	1/43	1/129	1/215	1/43	1/129	1/215
OCSVM	0.882	0.8795	0.8773	0.8795	0.8483	0.8433

表 1. $\nu=0.05$ 下的數據結果表						
		Precision		Recall		
訓練樣本	包含異常剖面	包含異常剖面	包含異常剖面	包含異常剖面	包含異常剖面	包含異常剖面
演算法 γ	1/43	1/129	1/215	1/43	1/129	1/215
OCSVM	0.61675	0.5841	0.5844	0.9028	0.9538	0.957

表 2. $\nu=0.04$ 下的數據結果表						
		Precision		Recall		
訓練樣本	包含異常剖面	包含異常剖面	包含異常剖面	包含異常剖面	包含異常剖面	包含異常剖面
演算法 γ	1/43	1/129	1/215	1/43	1/129	1/215
OCSVM	0.6376	0.6052	0.6049	0.9027	0.9532	0.9578

表 3. $\nu=0.015$ 下的數據結果表						
		Precision		Recall		
訓練樣本	包含異常剖面	包含異常剖面	包含異常剖面	包含異常剖面	包含異常剖面	包含異常剖面
演算法 γ	1/43	1/129	1/215	1/43	1/129	1/215
OCSVM	0.7314	0.6901	0.6889	0.8964	0.9443	0.9479

表 4. $\nu=0.005$ 下的數據結果表						
		Precision		Recall		
訓練樣本	包含異常剖面	包含異常剖面	包含異常剖面	包含異常剖面	包含異常剖面	包含異常剖面
演算法 γ	1/43	1/129	1/215	1/43	1/129	1/215
OCSVM	0.8239	0.7865	0.7844	0.8896	0.9398	0.9441

表 5. $\nu=0.000$ 下的數據結果表						
		Precision		Recall		
訓練樣本	包含異常剖面	包含異常剖面	包含異常剖面	包含異常剖面	包含異常剖面	包含異常剖面
演算法 γ	1/43	1/129	1/215	1/43	1/129	1/215
OCSVM	0.8778	0.8287	0.8221	0.8863	0.9369	0.9485

結論

1. 本研究利用 scikit-learn 程式庫之 One Class SVM 函數進行兩項實驗，

第一項為改變 ν (nu)，從上述表格可以看出在固定 γ 下， ν (nu) 為 0 時 F1 分數及 F2 分數最高。

2. 第二項調整參數值為 γ ，從上述表格可以看出在固定 ν (nu) 下， γ 為 1/43 時，F1 分數及 F2 分數最高

3. 由此可知往後做實驗可以依照此 ν (nu) 及 γ 操作半監督式學習之剖面監控相關實驗，可以獲得最大的 F1 分數及 F2 分數，後續研究方向可以在上述方向上進行。

4. 此外，未來研究可以探討訓練資料集中，資料組數不同對於績效之影響，研究結果將有助於提升智慧型製程管制之技術水準。