



台灣人工智慧學校

智慧製造與生產線上的資料科學

Data Science in Manufacturing From Predictive to Prescriptive

Chia-Yen Lee, Ph.D. (李家岩 博士)

Institute of Manufacturing Information and Systems (製造資訊與系統研究所)

Dept. of Computer Science and Information Engineering (資訊工程系)

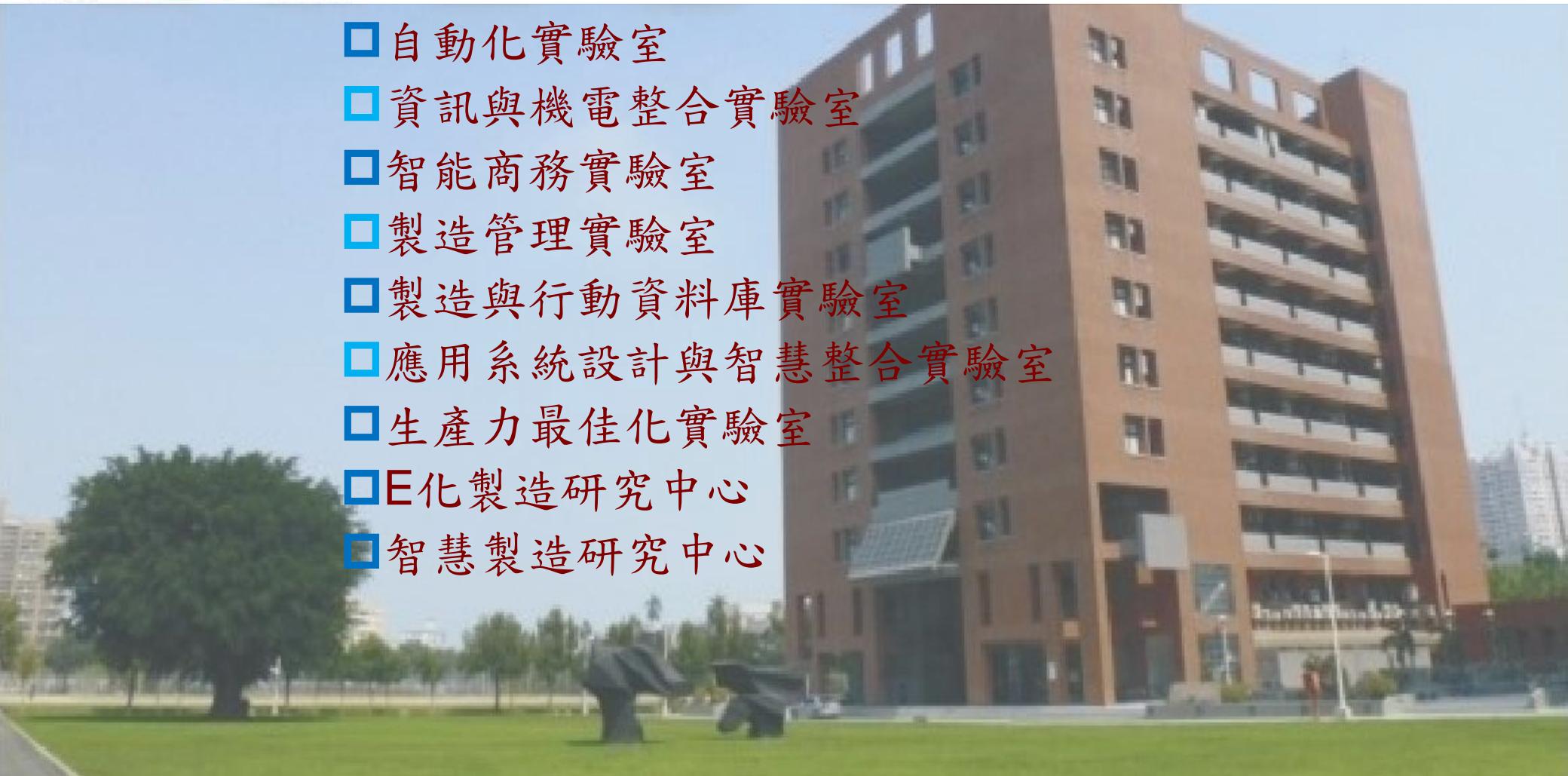
Engineering Management Graduate Program (工程管理碩士在職專班)

National Cheng Kung University (國立成功大學)



成功大學製造資訊與系統研究所(前稱製造工程研究所)，成立於民國八十三年八月，任務在於配合我國高科技產業發展及傳統工業升級之需要，培養具E化製造、製造技術、製造系統、製造知識管理與應用及製造管理之製造資訊與系統整合人才，並以執行產學合作計畫之方式，提昇產業競爭力之製造資訊系統研究。

- 自動化實驗室
- 資訊與機電整合實驗室
- 智能商務實驗室
- 製造管理實驗室
- 製造與行動資料庫實驗室
- 應用系統設計與智慧整合實驗室
- 生產力最佳化實驗室
- E化製造研究中心
- 智慧製造研究中心



□ Education

- Ph.D, 工業與系統工程, Texas A&M University, USA
(Major: Operations Research 作業研究/運籌學)
- M.S., 工業工程與工程管理, 國立清華大學
- B.S. & B.B.A., 應用數學暨資訊管理, 國立政治大學



□ Experience

- 教授兼所長，國立成功大學資訊工程學系暨製造資訊與系統研究所
- 副編輯, Flexible Services and Manufacturing Journal (SCI).
- 台積電工業工程師、陸軍少尉資訊官

□ Award

- 呂鳳章先生紀念獎(2019)
- Micron Teacher Award (2018)
- 李國鼎科技與人文講座研究獎 (2018)
- 科技部吳大猷先生紀念獎 (2017)
- 優秀青年工業工程師獎 (2016)



□ Research Interest

- 製造數據科學、智慧型製造系統、生產力與效率分析、多目標決策

Gartner Hype Cycle for Emerging Technologies, 2019



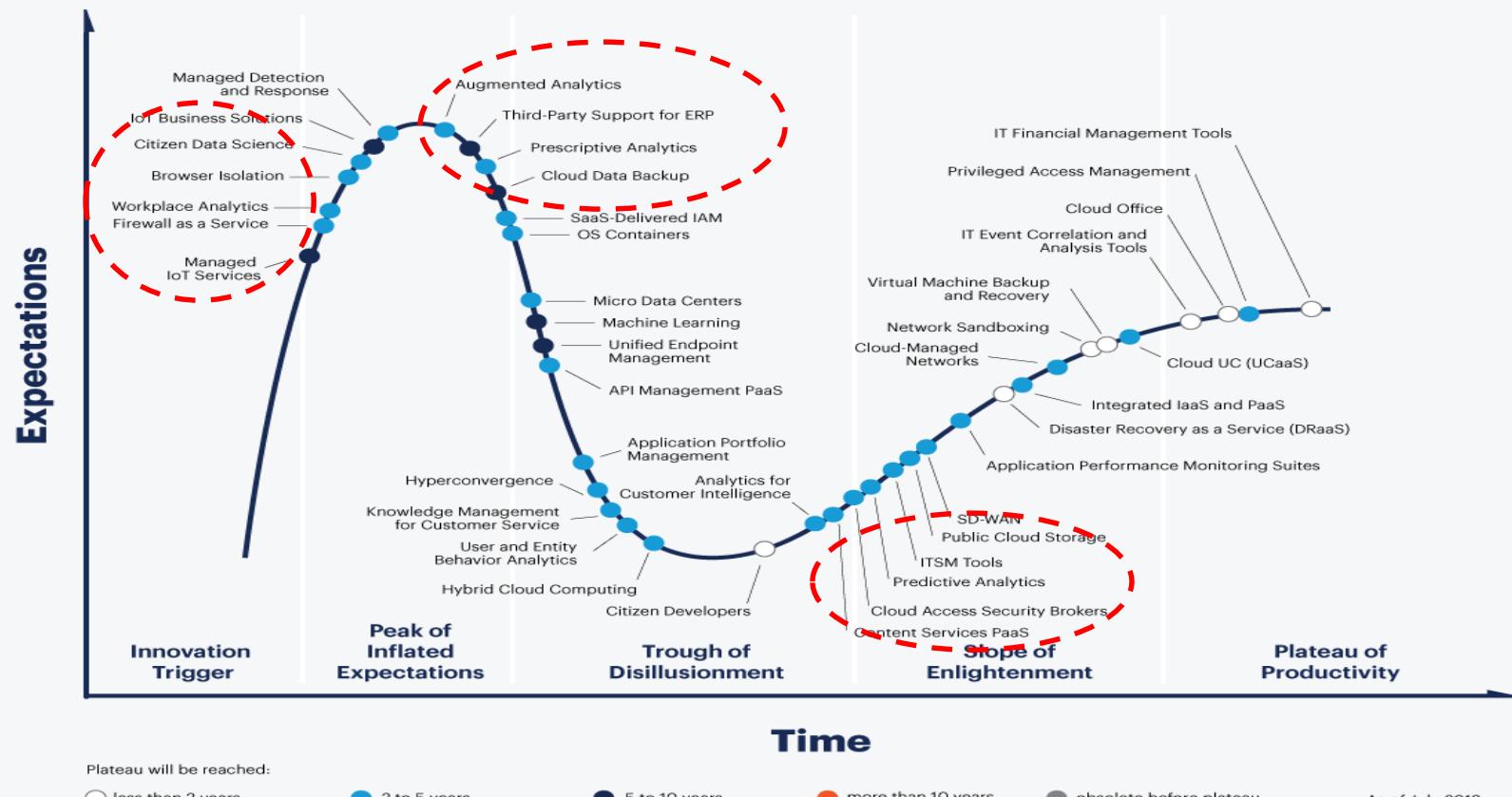
gartner.com/SmarterWithGartner

Source: Gartner
© 2019 Gartner, Inc. and/or its affiliates. All rights reserved.

Gartner

Pereira, D. (2019). Gartner 2019 Hype Cycle for Emerging Technologies. What's in it for AI leaders? <https://towardsdatascience.com/gartner-2019-hype-cycle-for-emerging-technologies-whats-in-it-for-ai-leaders-3d54ad6ffc53>

Gartner Hype Cycle for Midsize Enterprises, 2019



gartner.com/SmarterWithGartner

Source: Gartner
© 2019 Gartner, Inc. and/or its affiliates. All rights reserved.

Gartner®

Panetta, K. (2019). Midsize enterprises can use the Hype Cycle as a guide to investing in emerging technologies.
<https://www.gartner.com/smarterwithgartner/3-major-trends-drive-gartner-hype-cycle-midsize-enterprises-2019/>

1

智慧製造與資料預處理技巧

Intelligent Manufacturing and Data Preprocessing

2

重要參數篩選與精度預測

Empirical Case Studies

3

自動化生產排程與演算法

Automatic Production Scheduling Algorithms

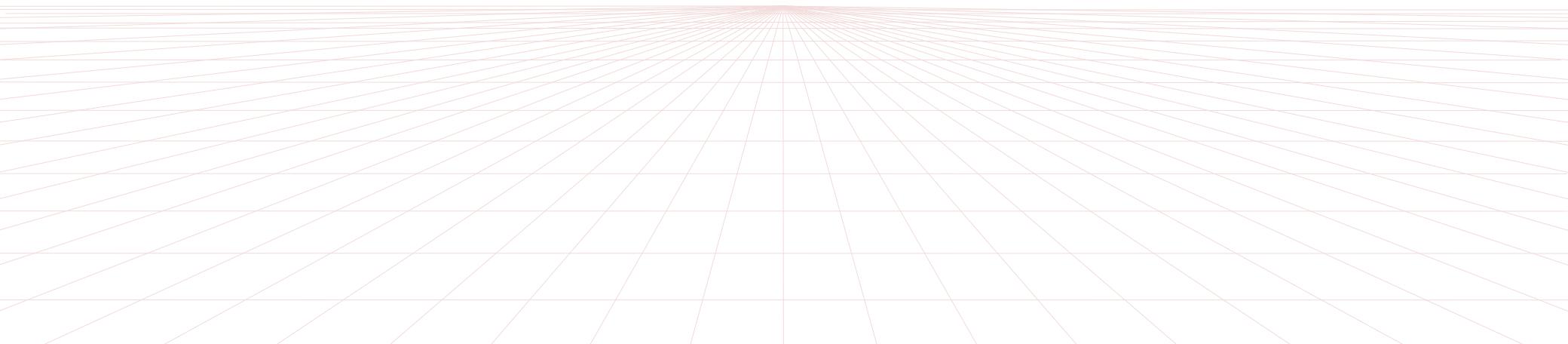
4

從預測性思維到處方性決策

From Predictive to Prescriptive

智慧製造與資料預處理技巧

Intelligent Manufacturing and Data Preprocessing



Machine learning opportunities in manufacturing

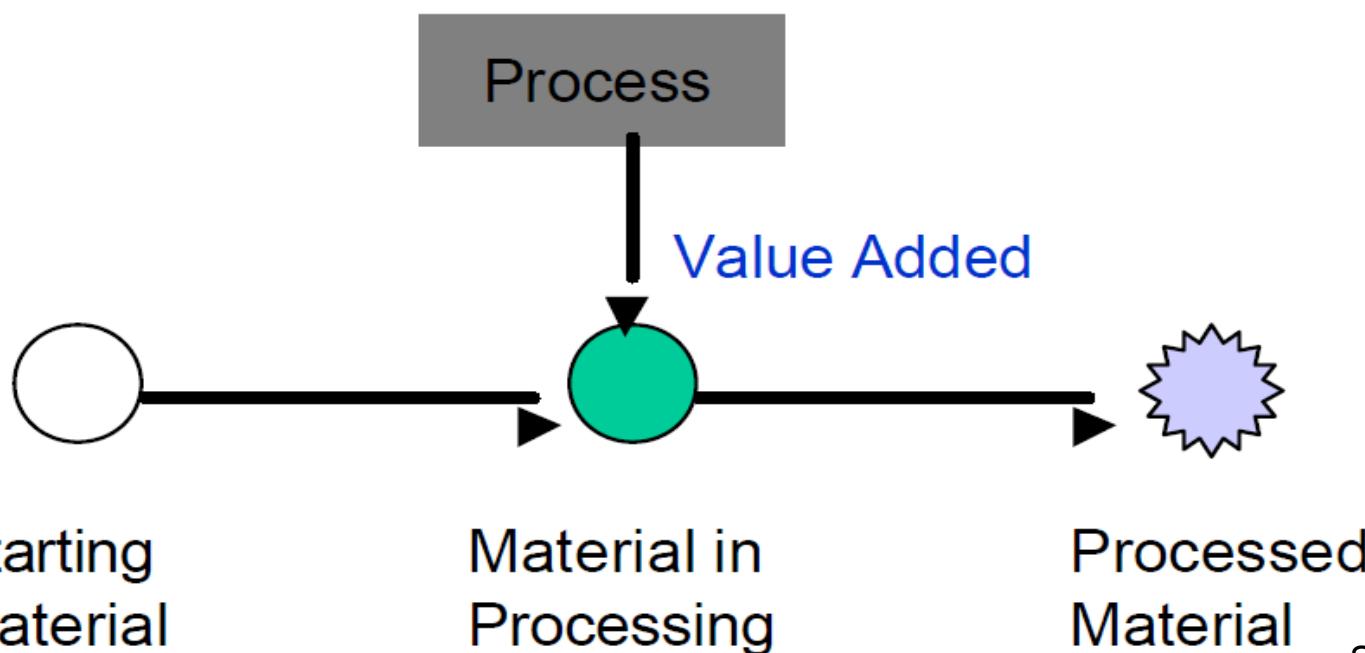
120 Machine Learning business ideas from McKinsey

Highest-ranked use cases, based on survey responses	Use case type	Impact	Data richness
Predict failure and recommend proactive maintenance for production and moving equipment	Predictive maintenance	<div style="width: 100%;">1.3</div>	<div style="width: 100%;">1.0</div>
Optimize complex manufacturing process in real time—determine where to dedicate resources to reduce bottlenecks and cycle time	Operations/logistics optimization (real time)	<div style="width: 100%;">1.1</div>	<div style="width: 100%;">1.0</div>
Predict future demand trends and potential constraints in supply chain	Forecasting	<div style="width: 100%;">0.8</div>	<div style="width: 100%;">0.7</div>
Identify design problems in pre-production to reduce ramp-up time to maximum output (i.e., yield ramp)	Predictive analytics	<div style="width: 100%;">0.6</div>	<div style="width: 100%;">0.3</div>
Identify root causes for low product yield (e.g., tool-/die-specific issues) in manufacturing	Discover new trends/anomalies	<div style="width: 100%;">0.5</div>	<div style="width: 100%;">0.7</div>
Detect defects and quality issues during production using visual and other data	Process unstructured data	<div style="width: 100%;">0.4</div>	<div style="width: 100%;">0.7</div>

- What is “Manufacturing”?

What is “Manufacturing”?

Manus (hand) Factus (make)



Swann (2003)

Manufacturing is the realization (具現化) of product.

What is Intelligent?



Shutterstock (www.shutterstock.com)

DATA



詮釋的觀點

Information



思辨的過程

Knowledge



科學的原理

迷思：資料？資訊？

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	lot	locn1	locn2	locn3	locn4	locn5	locn6	locn7	locn8	locn9	locn10	locn11	locn12
2	3699	2	3	9	4	1	16	5	10	5	2	6	2
3	1427	9	5	8	3	1		2	10	2	1	16	5
4		2	2	7	1	2	7		5		3	8	
5	3553	9	3	11		4	7		3	9	4	8	4
6	3591	9		6		5	13		10	3		17	
7	3190	9	1	2	3	1	4	5	6			11	2
8	716	1	2	9		2	12	1	3	8	4	3	5
9	91	2	2	5		4	8	3	2	5	3	7	1
10	214	1	1	3	3	3	7	4		2	3	10	1
11	3648	8		6	4	3	5	5		6	4	3	2
12	2128	7	3	5	4	3	11	5		3	2	8	5
13	151		3	8	1	2	3	2	7	2	3	10	2
14	800	6	2	2		5	18	1		2	2	17	1
15	224	7	1		3	4	15	4	2		2	4	5
16	588	4	1	2	5	5	5	5	6	7	3	13	4
17	3798	7	4	2	2	3		4	6	3	3	1	5
18	2012	9	3	3	2	4		1	1	3	4	10	1
19	2895		5	8	4	2	9	4	3		2	4	3
20	1740	3	4	2	1		13	1	7	7	3	18	3
21	3951	5	1		3	2	3	3	11	2	1	19	1
22	3111	6	2	2	4	5	16	2	4	7	3	17	5
23	150	7	1	9	3	1	19	4	4	7	2	11	4
24	commonalityx4000												

Advanced Analytics – Intel: SETFI: Manufacturing data: Semiconductor tool fault isolation. Causality Workbench Repository,
<http://www.causality.inf.ethz.ch/repository.php> (2008)

Monetize Data

資料價值驅動



詳情參見：

<http://fredbigdata.blogspot.tw/2013/05/big-data.html>

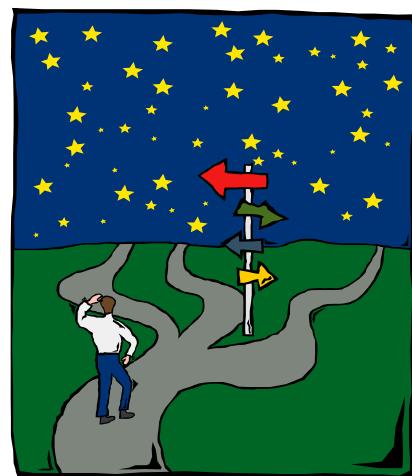
Etu 蔣居裕 (2013)

Problem



權衡的觀點

Decision



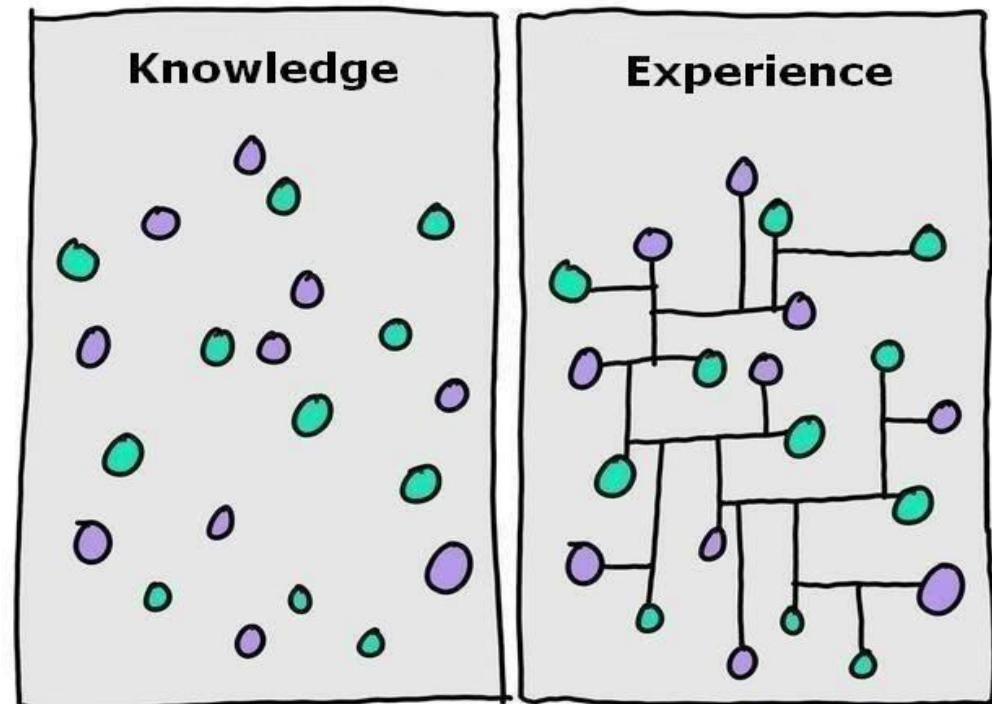
失敗的藝術

Experience



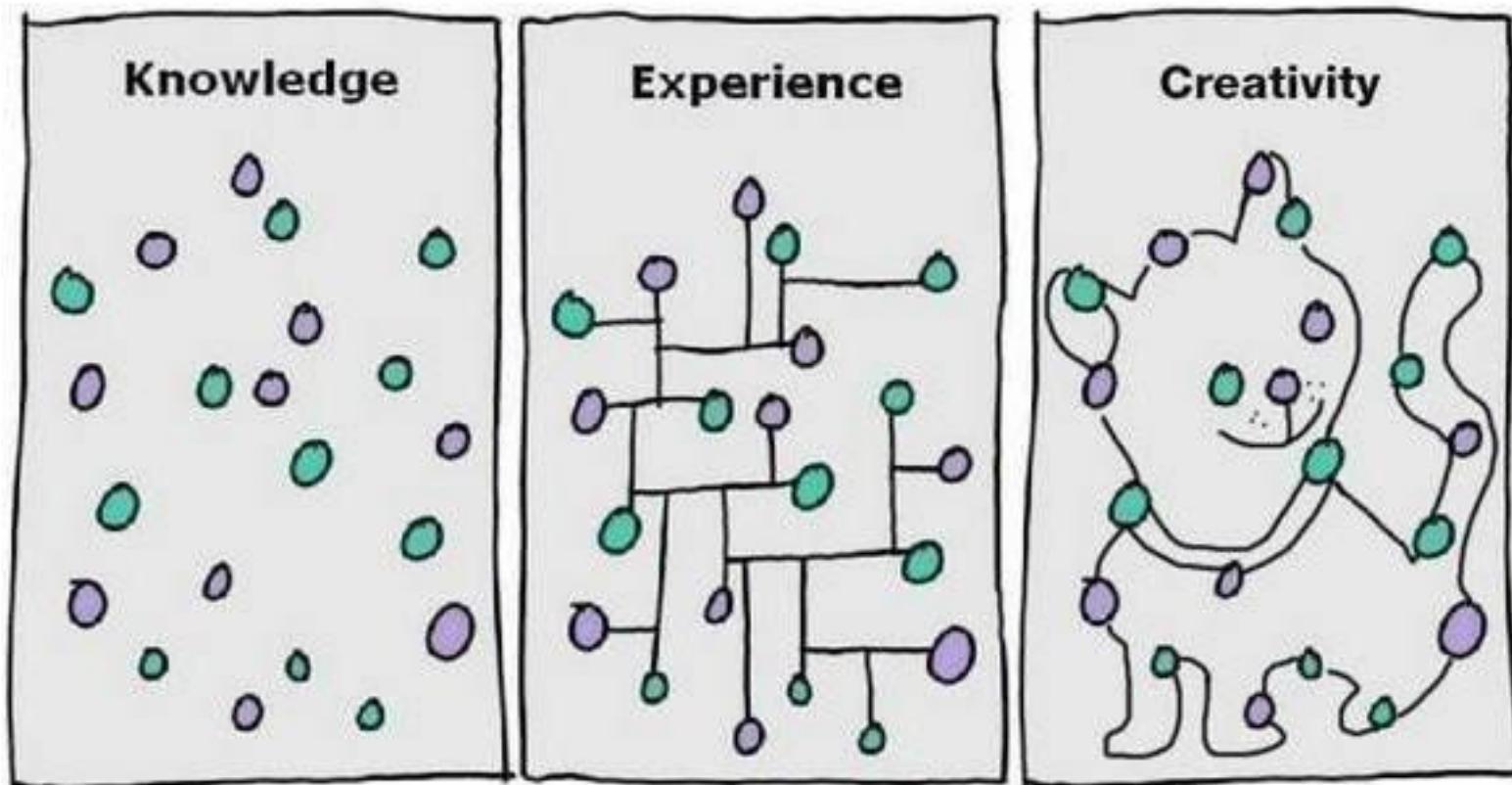
人生的歷練

知識 與 經驗 之 融會貫通 (思辨與討論)



<http://www.herogamingjobs.com/2014/01/07/experience-vs-knowledge/>

- The most important skill in the future will be the ability to “connect the dots” in **your own way!** (Moioli, 2019)



Moioli, Fabio (2019). The most important skill in the future will be the ability to "connect the dots" in your own way!.
<https://www.linkedin.com/pulse/most-important-skill-future-ability-connect-dots-your-fabio-moioli>

Intelligent Factory is a **decision-oriented** system which has the computational intelligence and self-learning ability to optimize the manufacturing process.

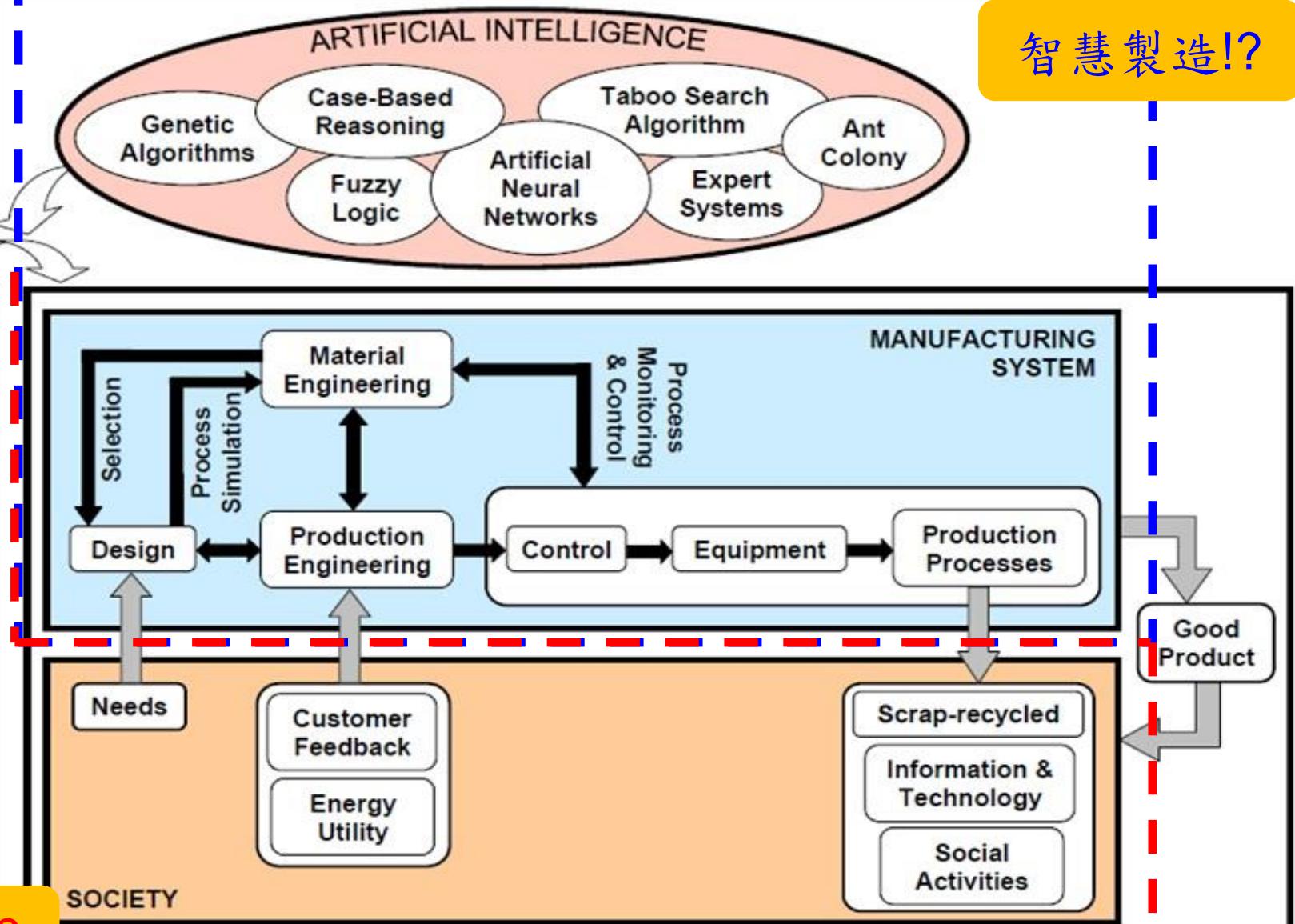
計算智慧 → Based on Data (資料處理與分析)

自我學習 → Real-time Feedback Control (回饋控制)



Adopted and reproduced from
Rao et al., "Integrated distributed intelligence in manufacturing", 1993 and
Kumara S.R.T., "Intelligent Manufacturing Systems: Mapping of AI Tools to
Manufacturing", 2nd Meeting on Applications of AI, 1995.

智慧製造!?



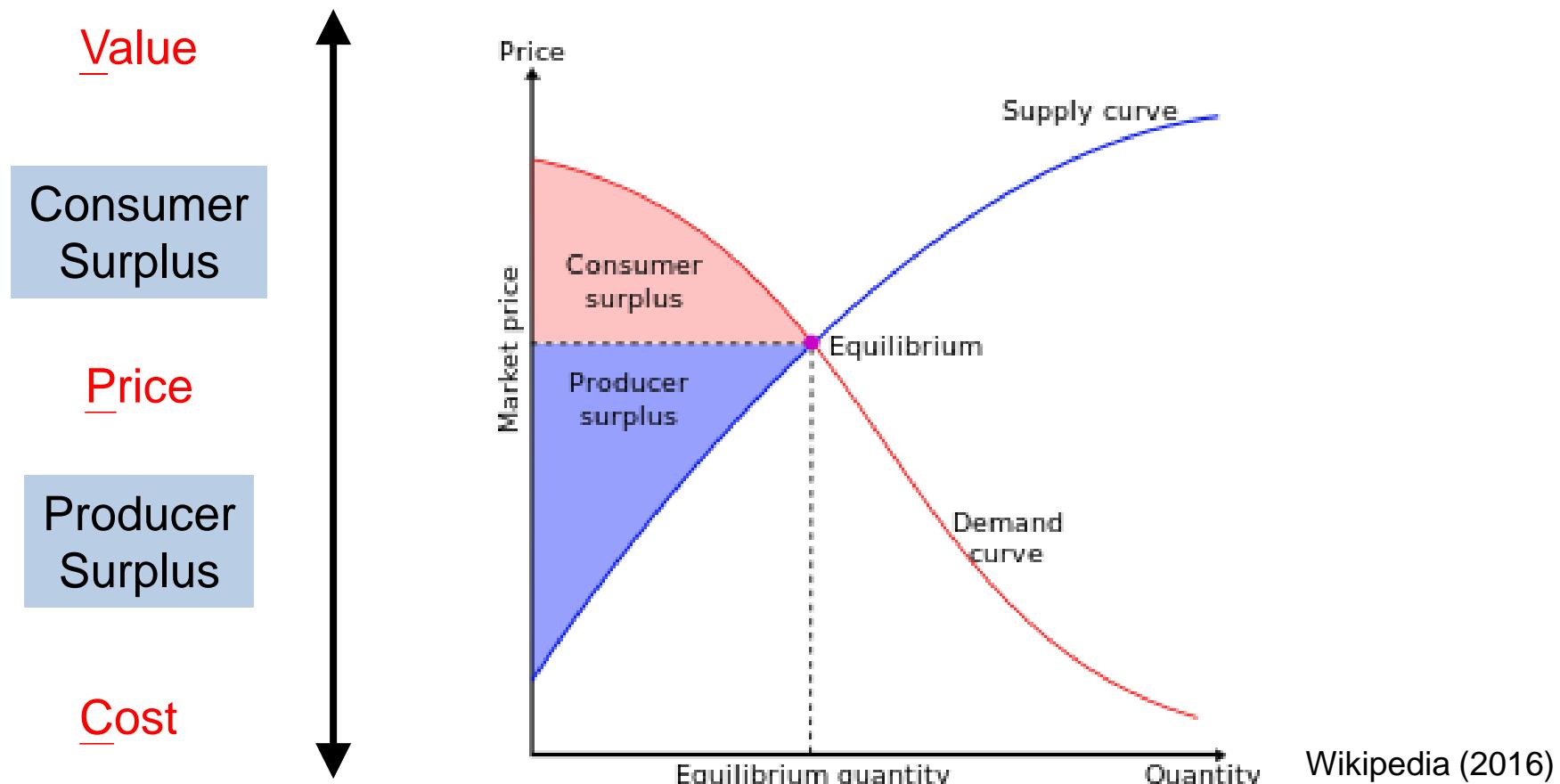
製造智慧!?

Toiga Bozoana (2012)

Manufacturing System (製造系統)

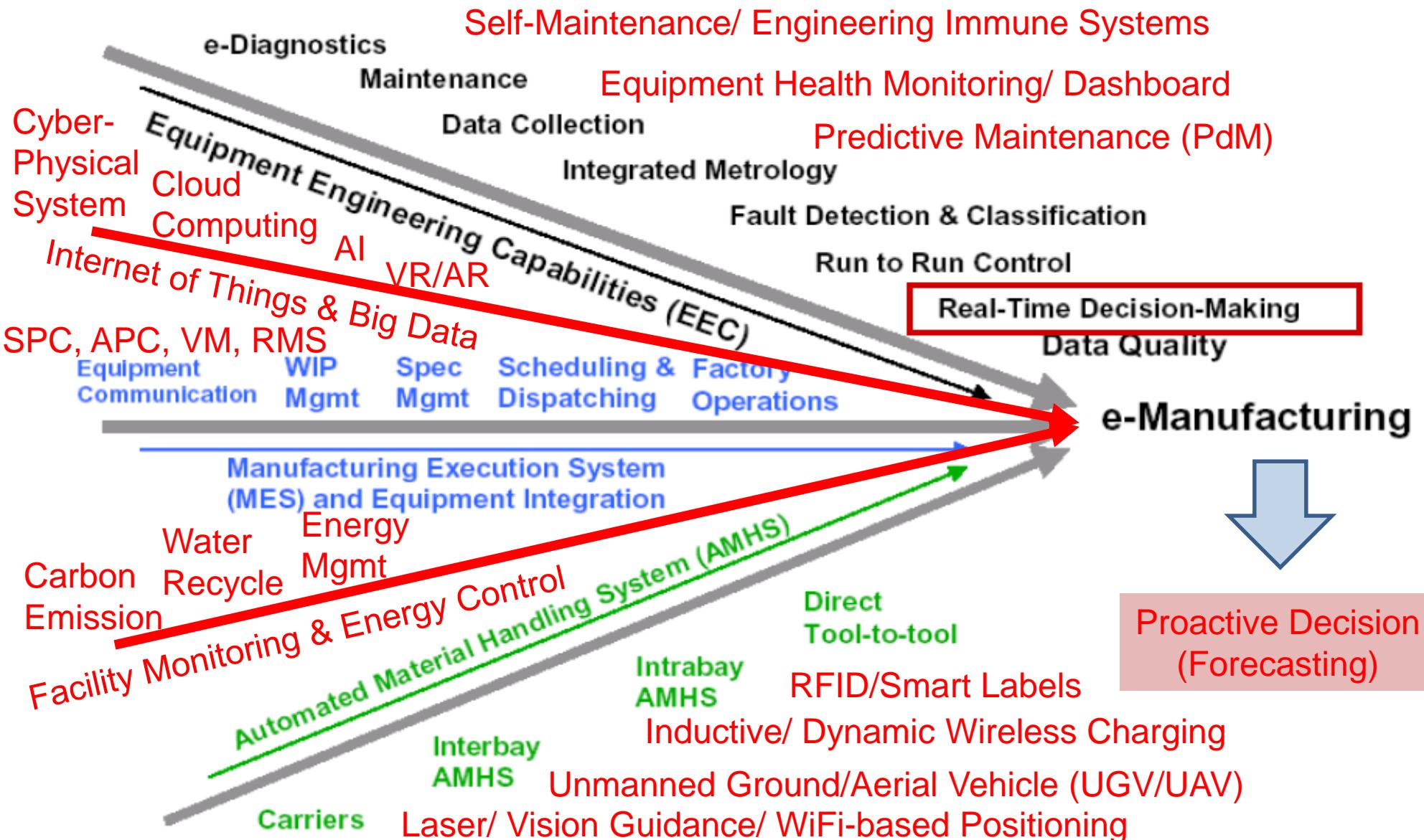
□ Value

- Manufacturing is **value-added** process for the **realization** of the product
- Value – Price – Cost (VPC) framework
- Customer Surplus vs. Producer Surplus



Wikipedia (2016)

製造智慧：決策



Source: International SEMATECH e-Diagnostics and EEC Guidance 2003

智慧製造與生產線上的資料科學

李家岩 (成大資訊系暨製造所) 22

□ Difference between Manufacturing and Service

System	Manufacturing	Service
Input	Material	Labor
Process	Capital-intensive Build SOP	Labor-intensive Building SOP is difficult
Output	Physical product	Intangible Non-separable Non-stable Non-storable Time-concerned
feedback	Quality Control Performance criteria	Difficult to measure quality and performance

□ Now, manufacturing system moves toward **service system**.

□ Manufacturing system vs. Service system

- Main difference is “**Inventory**”!



□ Inventory

- Raw materials, Components, Work-in-process (WIP), Finished goods

Lead Time + Uncertainty = Inventory

□ Inventory Reduction

- Reduce material and production lead time (includes transport)
- Reduce information delay times (**長鞭效應, Bullwhip effect**)
- Improve quality of information (reduce uncertainty)

1. 生產效率

- 提高系統產出量[5, 18, 19]
- 系統產能的穩定
- 精準的在製品控制[5, 10, 20]
- 較高效率的生產[16, 20]
- 減少生產流程時間[10,23]
- 降低每層加工的生產流程時間[18]
- 減少機器閒置
- 提高直接員工生產力[18]
- 提高員工總生產力[18]
- 提高瓶頸機台的效能[18]
- 提高設備使用率[19]
- 平衡生產線
- 縮短定貨的前置時間

2. 生產品質

- 良好的生產品質[8, 16, 21]
- 提高生產線良率[18]
- 提高晶粒良率[18]
- 良好的生產控制[5, 8, 19]
- 避免錯誤操作次數
- 機器的預防保養
- 提高機台穩定度
- 人員操作效率[16]

3. 服務

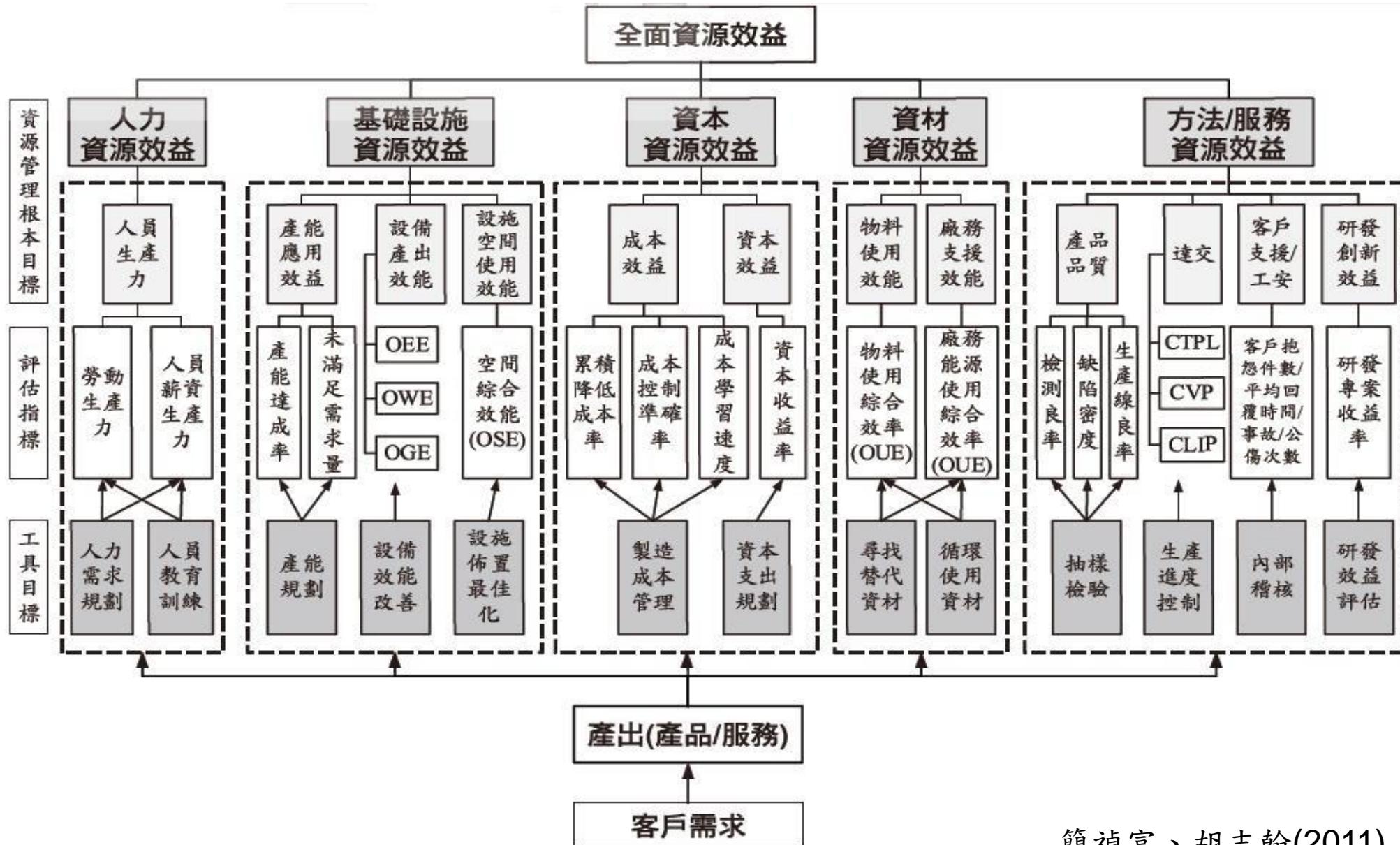
- 生產交期的達成[17, 18]
- 最少化客戶抱怨

4. 成本

- 較低的生產成本[18]
- 降低每片晶圓的平均成本
- 降低晶圓每層加工的平均成本

Chien et al. (2004)

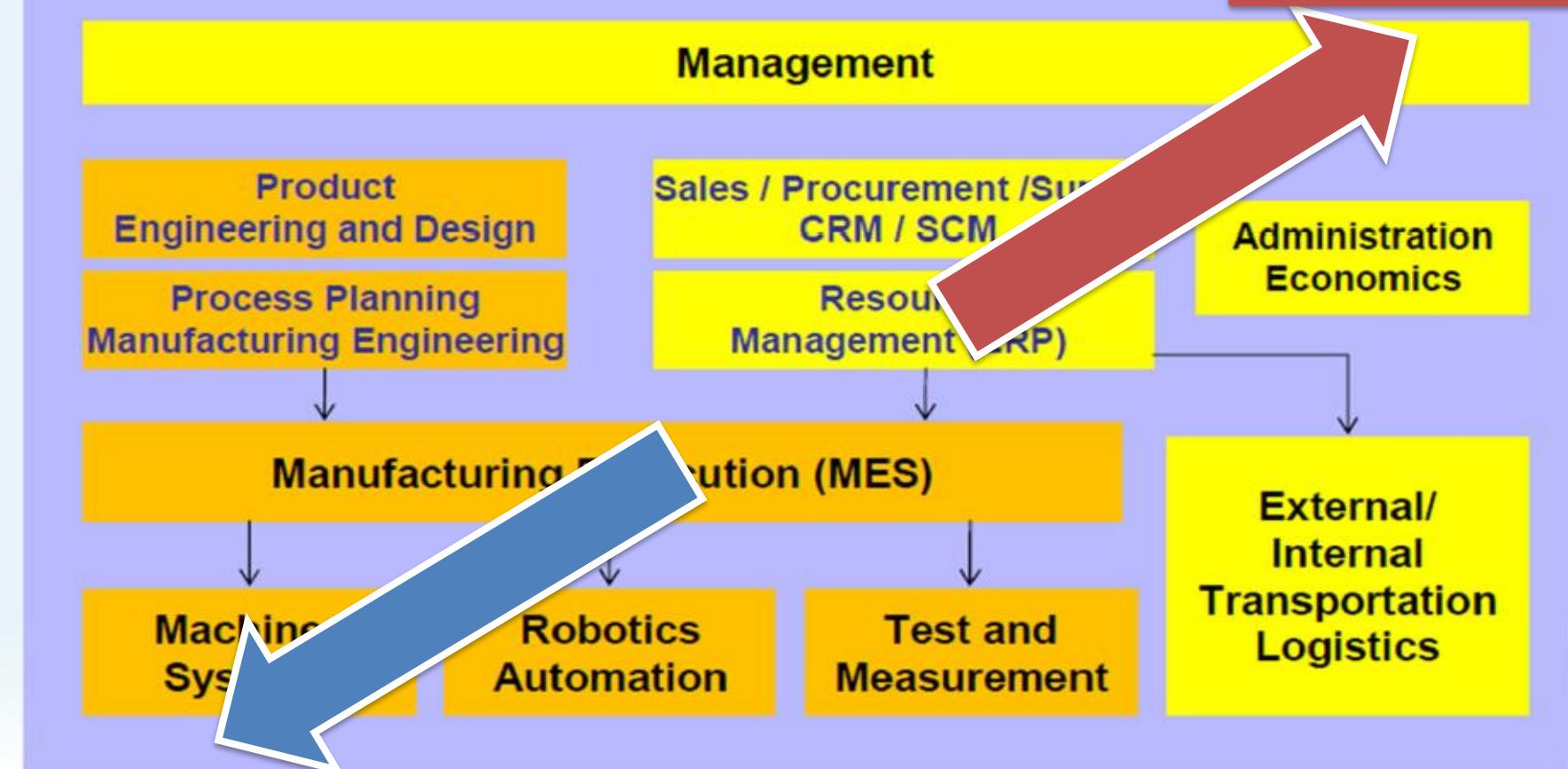
Why is it difficult to manage a manufacturing systems?



簡禎富、胡志翰(2011)

目標的衝突!?

為了營收與價格維持
→ 差異化(少量多樣)



為了效率與品質要求
→ 低成本(大量生產)

process-oriented

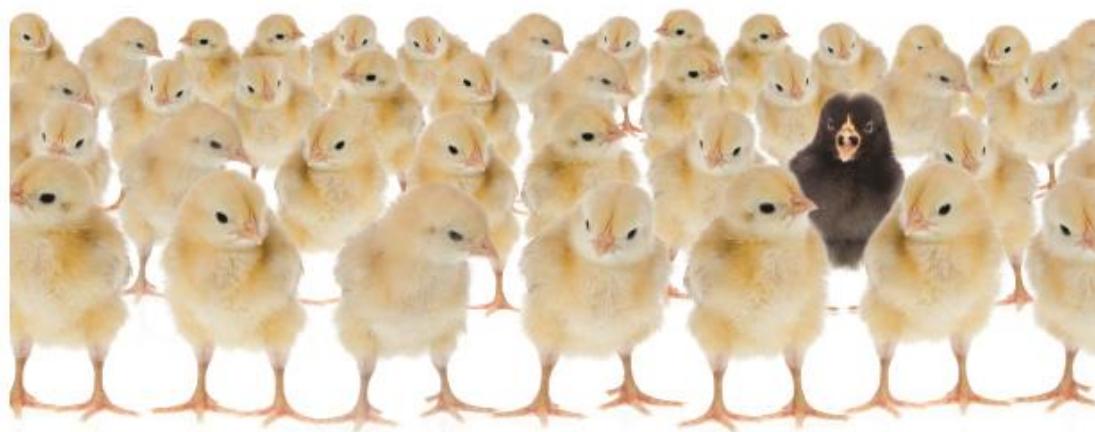
Commercial-oriented

Prof.Dr. Engelbert Westkämper, Fraunhofer IPA Stuttgart, Germany, "Factories of the Future beyond 2013: The role of ICT"
http://cordis.europa.eu/fp7/ict/micro-nanosystems/docs/fof-beyond-2013-workshop/westkaemper-manufuture_en.pdf

- Variability is anything that causes the system to **depart from regular**, predictable behavior.
- Sources of Variability:

setups
machine failures
materials shortages
yield loss
rework
operator unavailability

workspace variation
differential skill levels
engineering change orders
customer orders
product differentiation
material handling



➡ **Variability from Resource!**

現場「變異(variability)」是不好的!!?

但沒有變異，如何”改善”？
就如同考試成績有好有壞，才能對症下藥~

其實...

變異某種程度上也代表了“資訊量”!?

	English	Math
Student_A	80	76
Student_B	80	91
Student_C	80	83
Student_D	80	62
Student_E	80	88

製程控制

- MES
- APC
- FDC
- SPC
- VM
- R2R
- RMS
- Preventive Maintenance
- Predictive Maintenance
- Yield Management
- ...etc



生產管理

- Demand Forecast
- Long-Run Capacity Planning
- Short-Run Capacity Planning
- CapEx
- Cost Structure
- MPS & MRP
- Scheduling & Dispatching
- Inventory Management
- Order Releasing
- ...etc



□ Operation Automation

- Replace Human Efforts
- Benefits
 - Reduce HR Cost
 - Reduce M/O
 - Improve CT and Improve Productivity,

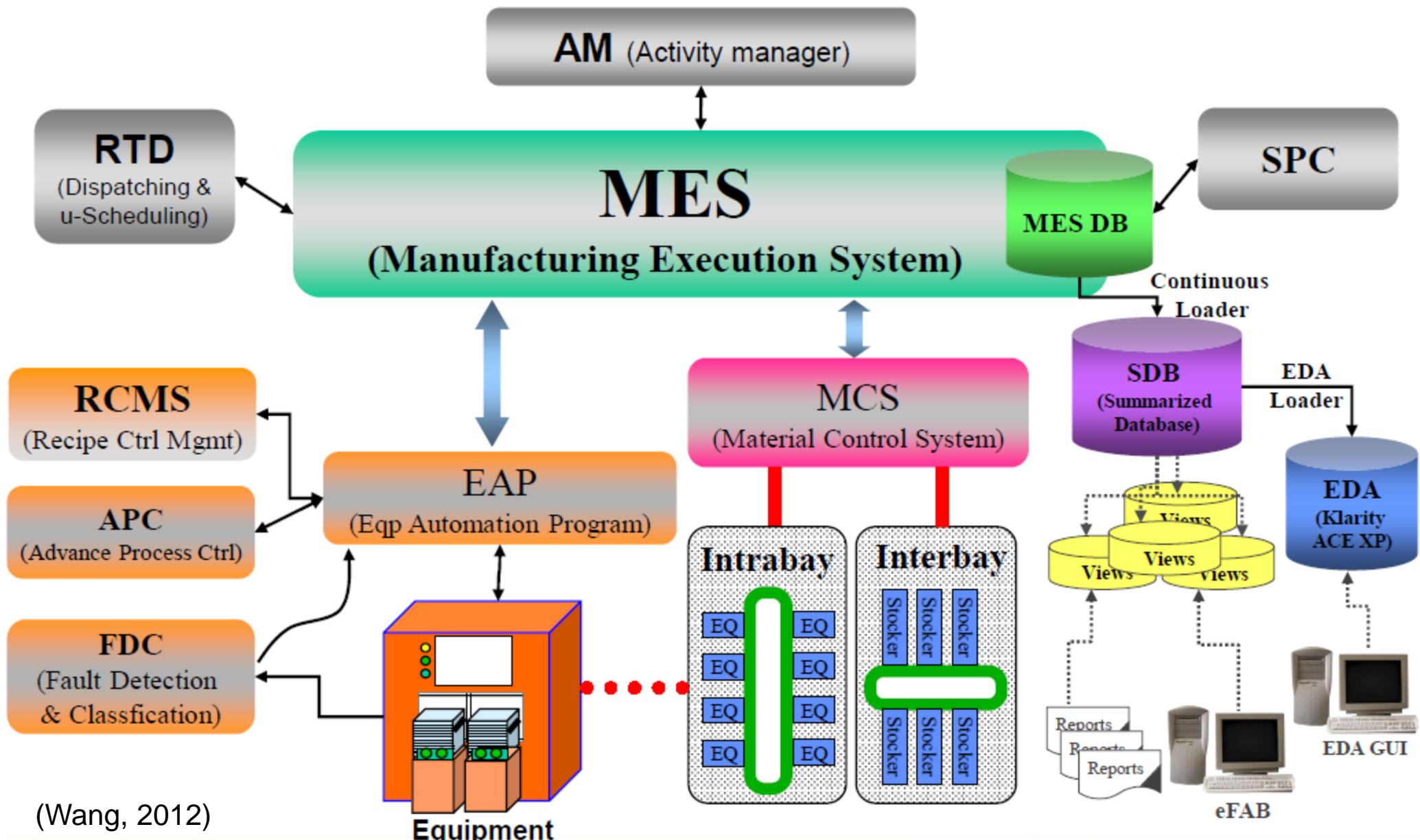


□ Engineering Automation

- Support Engineer for better decision-making or trouble-shooting
- Benefits
 - Sustain EQ variance
 - Improve quality & Yield
 - Support RD
 - Reduce Engineering Cost
 - Improve Time-to-Market



(Wang, 2012)



自動化並不全然意味著大量投資
機器人與設備...

事實上，自動化是製造哲學上的
轉變...

透過消除浪費、降低產線不平衡
達到生產標準化且具有彈性的
一種改善過程

□ Lean (identify non-value-added process and remove it)

- Value
 - Value Stream Mapping (VSM)

- Waste elimination (7 muda, Womack et al., 1990)

- 1. Transportation
 - 2. Inventory
 - 3. Motion
 - 4. Waiting
 - 5. Overproduction
 - 6. Overprocessing
 - 7. Defects

- Continuous flow
 - Line Balancing

- Pull production system



不要把浪費自動化了

AI並非取代人力...
人機共存、相輔相成



組織演化與新制度的建立

MES



APC



AMHS



Big Data
Analytics

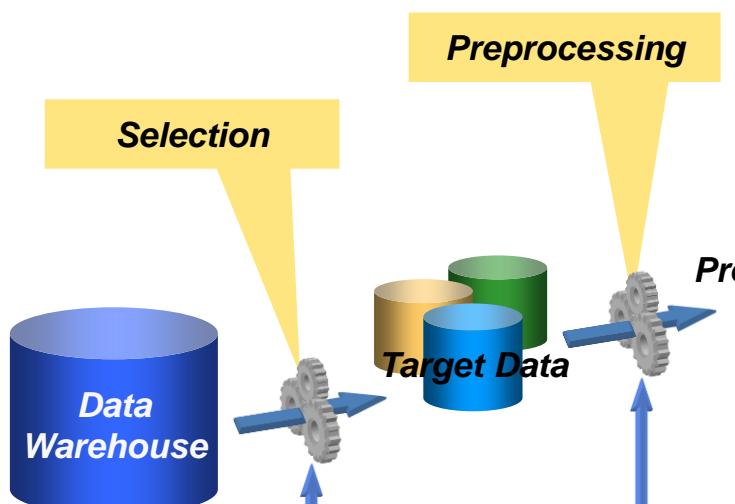
FMEC



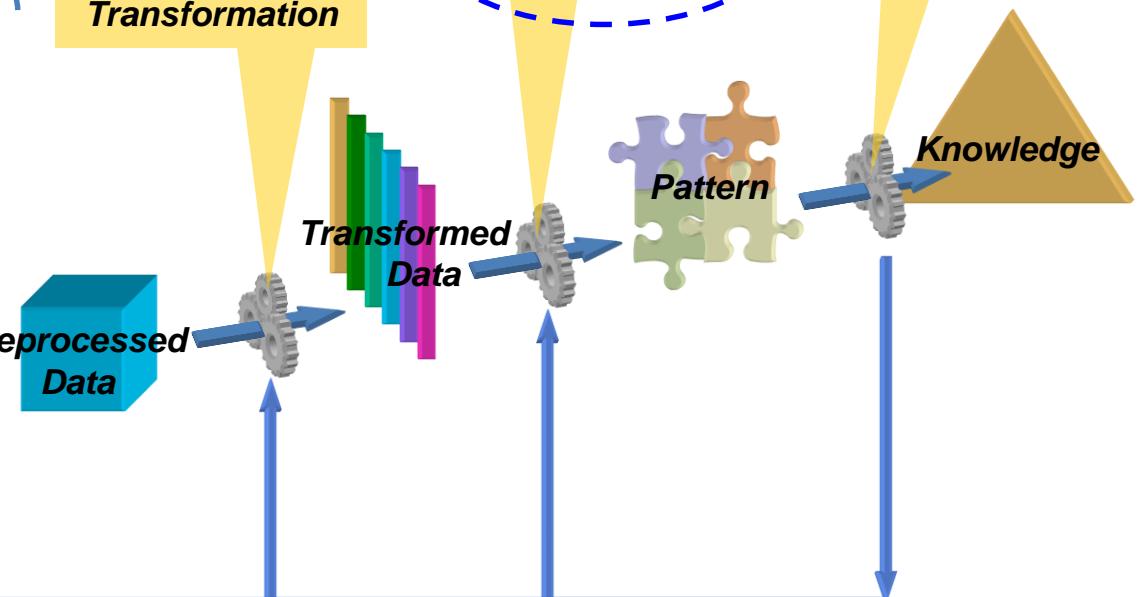
KNOWLEDGE DISCOVERY IN DATABASES (KDD)

- ETL (Extract-Transform-Load)
- Storage and Calculation

IT Infrastructure



Transformation



How to improve a complex manufacturing system?



Pittsburgh Technology Council (2014). <http://www.pgtech.org/media/64942/panouupdated.jpg>

Get to the “Bottom” of Our Problems

(問題本質的探索)

(The Characteristics of Manufacturing)

□ Manufacturing Field (problem clarification)

- Ask “**Why**” in shop floor level (在製造現場問”為什麼”)?
- 一言九鼎(頂)!
- No support (labor/machine/material/management/supplier/inventory...)

□ Manufacturing Dataset (data for analysis...but no data?)

- Ask Why?
- I don't know...no idea...
- Long long time ago...
- No one take over the ...
- None of my business...
- Please ask somebody/window/dept./supplier/business(**competitor**!?)
- I am the newcomer...
- Data is not accurate because...

Characteristics	Data & Management Issues
Batch size (生產批量)	Lot ID decomposition, lot tracing, merge/split
Parallel machine (平行機台)	Missing value, high dimension, multicollinearity
Golden machine (黃金機台)	Utilization, class imbalance → Inference bias
Recipe and parts (處方與零件)	Nominal or categorical variable → too many levels → too many dummy variables → high dimension
Sampling testing (抽樣檢測)	Missing value, multi-response, metrology delay
Engineering or R&D lot (工程與實驗貨)	Outlier, machine contamination, setup capacity loss, small dataset
Maintenance (維修保養)	When? how (大保養 or 小保養)? capacity loss, reliability, typing error, text, choosing "others"
Changeover (換線、換模)	Sequence-dependent setup time, capacity loss
Bottleneck shift (瓶頸站轉移)	Different treatment, WIP transfer, product-mix
Queue time limit (等候時間限制)	Defects, WIP 
Data imbalance (資料不平衡)	Inference bias Inventory = Lead Time + Uncertainty

Lee (2019)

製造數據科學要做到...
看到資料，就能對應到現場的
特性與問題

□ 混批、拆批、併(合)批

- Data Preparation

Lot ID	Sub Lot ID	Final inspection Pass or Fail
Lot001	Lot001	Pass
Lot002	Lot002_1	Pass
Lot002	Lot002_2	Fail
Lot003	Lot003_1_1	Pass
Lot003	Lot003_1_2	Pass
Lot003	Lot003_2	Pass

- Data Merge ([trace back](#) for recipe diagnosis)

Lot ID	WS1_A	WS1_B	WS2_A	WS2_B	...	Sub Lot ID	WS7_A	WS7_B
Lot002					...	Lot002_1		
						Lot002_2		

□ 混批、拆批、併(合)批

- 良率計算 (yield calculation)

Lot ID	Pass or Fail
Lot001	Pass
Lot002	Pass
Lot003	Fail
Lot004	Pass

75%

Lot ID	Pass or Fail	Lot size
Lot001	Pass	20
Lot002	Pass	25
Lot003	Fail	10
Lot004	Pass	15

85.7%

□ Parallel Machine

- Not identical (有機差) → Tool Matching



WS_A_Mach_1



WS_A_Mach_2

□ Data Preparation

- Missing Value

Lot ID	WS_A_Mach_1_Temp	WS_A_Mach_2_Temp
Lot001	820	N/A
Lot002	820	N/A
Lot003	N/A	840
Lot004	N/A	840



Lot ID	WS_A_Temp	WS_A_Mach_Type
Lot001	820	1
Lot002	820	1
Lot003	840	2
Lot004	840	2

□ Golden Machine

- 推論的偏誤 (inference bias)
- Missing Value

Lot ID	WS_A_	WS_A_
	Mach_1_	Mach_2_
	Temp	Temp
Lot001	820	N/A
Lot002	820	N/A
Lot003	820	N/A
Lot004	830	N/A
Lot005	820	N/A
Lot006	820	N/A
Lot007	825	N/A
Lot008	830	120
Lot009	820	130
Lot010	N/A	120



Lot ID	WS_A_	WS_A_	Mach_1	Mach_2
	Mach_1_	_Temp	Mach_1	Mach_2
Lot001	820	Avg.	1	0
Lot002	820	Avg.	1	0
Lot003	820	Avg.	1	0
Lot004	830	Avg.	1	0
Lot005	820	Avg.	1	0
Lot006	820	Av.g.	1	0
Lot007	825	Avg.	1	0
Lot008	830	120	1	1
Lot009	820	130	1	1
Lot010	Avg.	120	0	1

□ Recipe/ Parts- Nominal (名目) or Categorical (類別) Variable

- Transfer to dummy variable (亞變數, 虛擬變數)



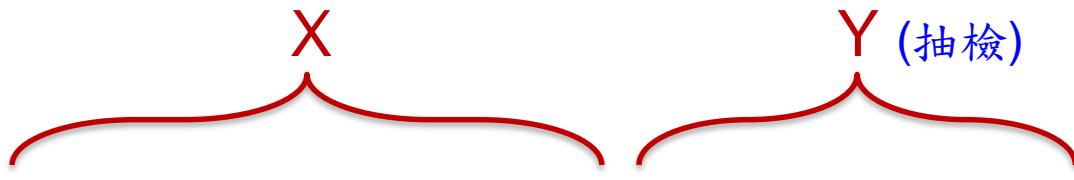
Lot ID	WS_A_	Lot ID	WS_A_	WS_A_	WS_A_	WS_A_
	Mach_1_		Mach_1_	Mach_1_	Mach_1_	Mach_1_
	Parts	PartsA	PartsB	PartsC	PartsD	
Lot001	PartsA	Lot001	1	0	0	0
Lot002	PartsB	Lot002	0	1	0	0
Lot003	PartsA	Lot003	1	0	0	0
Lot004	PartsC	Lot004	0	0	1	0
Lot005	PartsD	Lot005	0	0	0	1
Lot006	PartsE	Lot006	0	0	0	0
Lot007	PartsE	Lot007	0	0	0	0
Lot008	PartsA	Lot008	1	0	0	0
Lot009	PartsC	Lot009	0	0	1	0
Lot010	PartsE	Lot010	0	0	0	0

- Given N levels, the method will generate **N-1 dummy variables**.

□ 某類別變數level過多 (Recipe or Parts數目過多)

- 轉成Dummy Variables會產生許多新變數
 - Issue: Curse of Dimensionality (維度的詛咒)
- 建議方式
 - 將部分level整合 (grouping)
 - eg. 產品 → 產品族
 - eg. tool → tool group
 - 選取特定時間區間的資料進行分析
 - 降低該變數level的數目
 - 將某類別中只有出現一次觀測值的level刪除
 - 沒有再現性!

□ 抽樣檢測



手機外殼 (同產品)			材質	供應商	...	長度檢驗	寬度檢驗
NO.001	鋁	A				13.723	7.235
NO.002	鋁	B				NA	NA
NO.003	鋁	B				13.728	7.237
.							
.							
.							
NO.099	鋁	B				NA	NA
NO.100	鋁	A				13.726	7.236

沒被抽檢

填補遺漏值

沒被抽檢

Skills of missing value imputation:

1. 平均數/第一四分位數/眾數
2. K-Nearest Neighbors
3. Prediction Model

<http://www.theanalysisfactor.com/seven-ways-to-make-up-data-common-methods-to-imputing-missing-data/>

實驗貨工程貨

- 主要是研發產品、或機台測試校正用，資料上有時會呈現Outlier。
- 若針對一般性產品資料分析，需要在分析前先濾掉或移除。
- 一般而言，有特殊的LotID，在收集資料時可先過濾掉，或在預處理中進行。若無給定特殊LotID，那需要在資料中觀察，例如使用特殊機台、特殊recipe，該產品只經過某些特定製程等。

特性 階段	實驗貨工程貨	一般正常貨
資料量	剛起步，較少 ($n < p$)	較多 (大量生產) ($n > p$)
資料數值	實驗設計，參數較分散	很多參數已成為定值
成本	需要反覆試驗，較高	大量生產，較低
良率	較低	較高且穩定
分析方法	最佳化方法、無母數、實驗設計/ 田口方法、LASSO、SVM、 Forward Stepwise...	有母數、GLM、Random Forests, Boosting, Deep Learning...

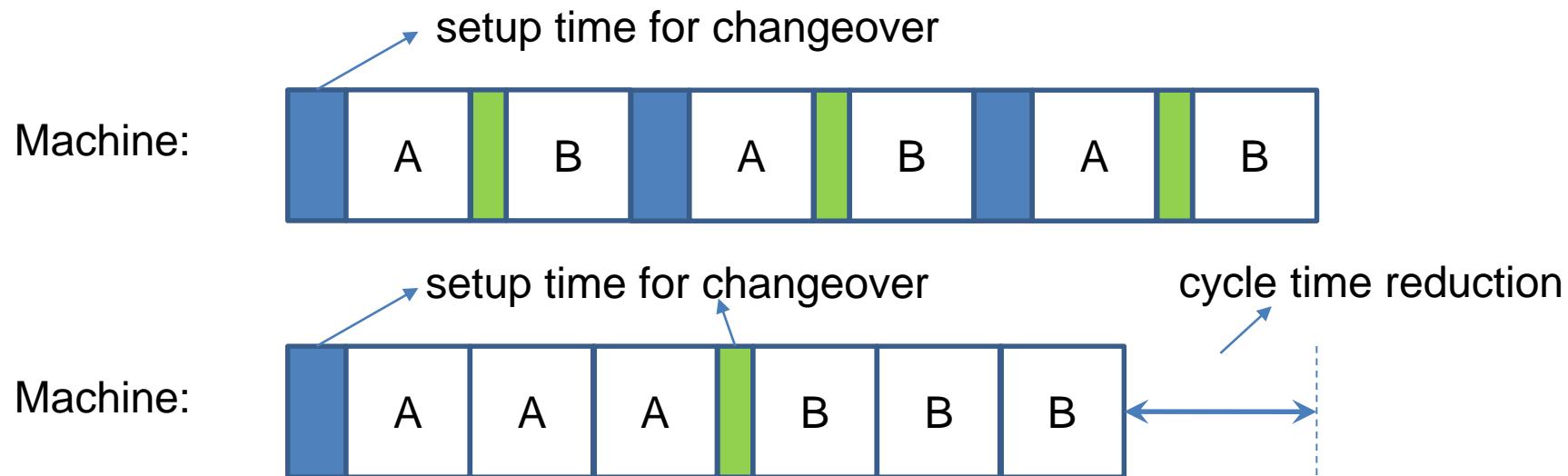
□ 換線、換模

- 當產線要生產不同的產品時，會針對機台進行換線/模的動作。
- 由下表可推測出來，換線/模時間約為1小時，故依據時間間隔可以明顯知道product之間有換模的時間納入。

ID	Product Type	Process Time (hrs)	Start Time
Lot1	A	1	2016/12/26 14:50
Lot2	A	1	2016/12/26 15:50
Lot3	B	2	2016/12/26 17:50
Lot4	B	2	2016/12/26 19:50
Lot5	B	2	2016/12/26 21:50
Lot6	C	1	2016/12/27 00:50
Lot7	C	1	2016/12/27 01:50

□ 換線、換模

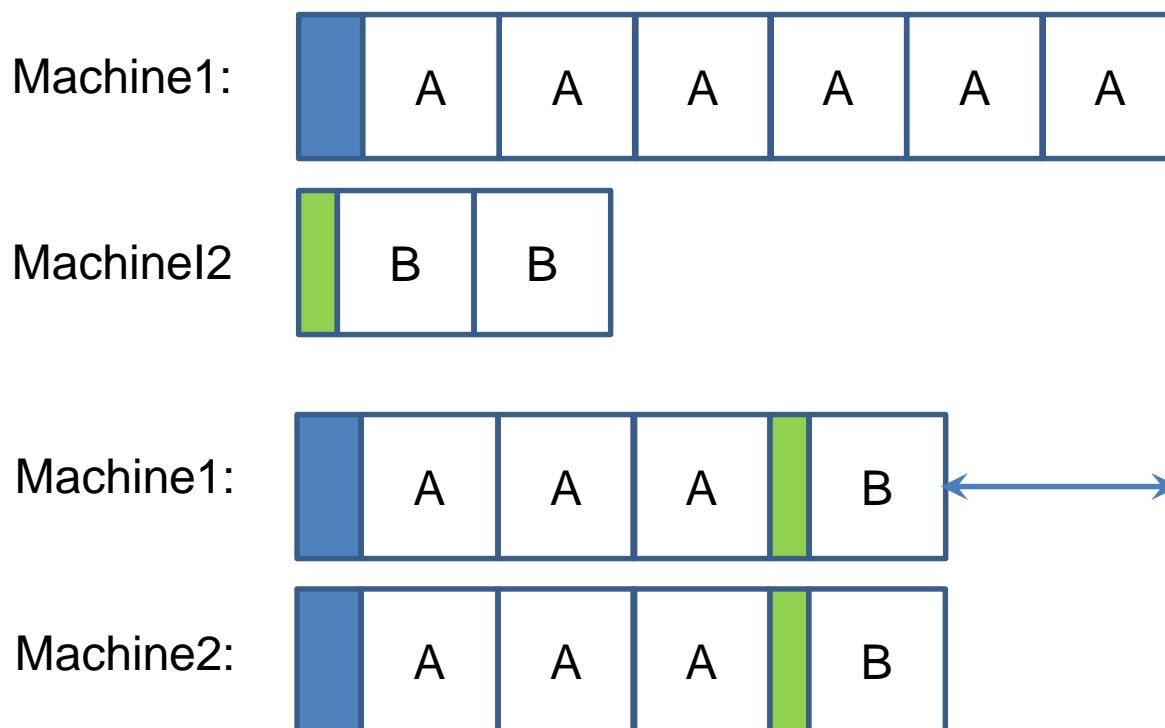
- 當產線要生產不同的產品時，會針對機台進行換線/模的動作。
- 同樣產品類型的儘可能排在一起，以減少換模(線)次數



- 以大批量生產的方式分攤換模(線)的時間
- 大批量生產方式會增加無訂單的庫存，因此必須預估數量，設定"經濟生產批量(Economic Lot Sizing)"來因應

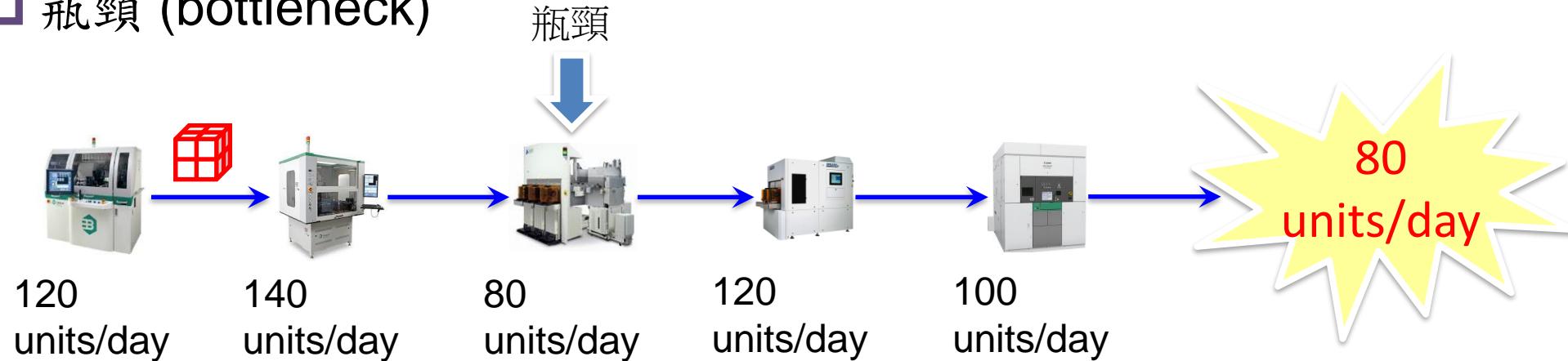
□ 總完工時間(makespan) vs. 換模次數??

- 原則上換模次數愈少，makespan下降，然而...當有大單時...
- 需留意多個平行機台(parallel machine)進行排程時的取捨(tradeoff)



- 資料會反應這件事情!
- 不單只是排程、良率也會因換模狀況而有所改變..

□ 瓶頸 (bottleneck)



- 一般來說，瓶頸機台常是利用率高且週期時間長的機台 → WIP堆積多

□ 內部瓶頸

- 特定機台或工作站的產能限制
- 薪資水準或工作環境無法吸引到優秀員工
- 搬運/運輸/物流形成為生產的瓶頸
- 現場管理團隊能力/生產規劃團隊的排程/規劃
- 管理階層對於系統產能不正確假設/認知

□ 外部瓶頸

楊大和(2016)

- 原物料的供應
- 特定區域的人力供給
(勞工和幹部)
- 公司產品的品牌知名度
- 公司產品的配銷通路

□ 等候時間限制 (Queue Time Limit)

- 由於半導體製程晶圓表面上為化學物質，若長期曝露於一般空氣中，會造成氧化反應而導致缺陷(defect)產生。
- 為了避免在製品於生產線上等待過久而造成製程缺陷，會根據製程與產品特性，在特定製程完成加工之後，規定等候時間的限制 (Queue Time Limit) 以維持產品良率。為了延遲此限制，多於FOUP中填充惰性氣體。
- 等候時間若發生在批次工作站(例如爐管製程, furnace)問題可能更嚴重。對於到達此工作站之晶圓，除了需要等前一批次加工完，還需要另外等候集批(Form Batch)，換言之，需要等待多個批量後(有相同的recipe製程條件)，該工作站才進行作業。此加工型態會造成產品的等候時間過長，甚至超出等候時間限制，而造成不良品產生。
- 通常可計算Qtime當作獨立變數(x)來對良率(y)進行建模，以瞭解Qtime長短如何影響良率的情況。
 - estimated by the difference between check-out of A and check-in of B

□ 等候時間限制 (Queue Time Limit)

- 假設某產品A，規定在每個stage加工後，**等候時間限制為1小時**就需要進入下一站。若發現等候時間異常且超過等候時間限制，須立即通報以免不良品往下游輸送。

ID	Product Type	Stage	Process Time (hrs)	Start Time
Lot1	A	1	1	2016/12/26 14:50
Lot1	A	2	1	2016/12/26 16:00
Lot1	A	3	0.5	2016/12/26 17:20
Lot1	A	4	1	2016/12/26 18:00
Lot1	A	5	0.5	2016/12/26 20:30

等候超過1小時 → Alarm!!

□ 資料合併

- 表單串接 – 注意必須為相同的欄位名稱, i.e., KEY
- Key通常為Lot ID, Machine ID等

Event-based record

Time	SVID 1	SVID 2
2/11 00:06:29			
2/11 00:10:41			
2/11 03:41:09			
.			
.			
2/11 23:11:57			

Periodic-based record

Time	SVID 101	SVID102
2/11 00:00:00			
2/11 01:00:00			
2/11 02:00:00			
.			
.			
2/11 23:00:00			

兩種不同類型的資料紀錄，該如何合併串接呢？Which one could be “Main Table”?

□ Data Merge

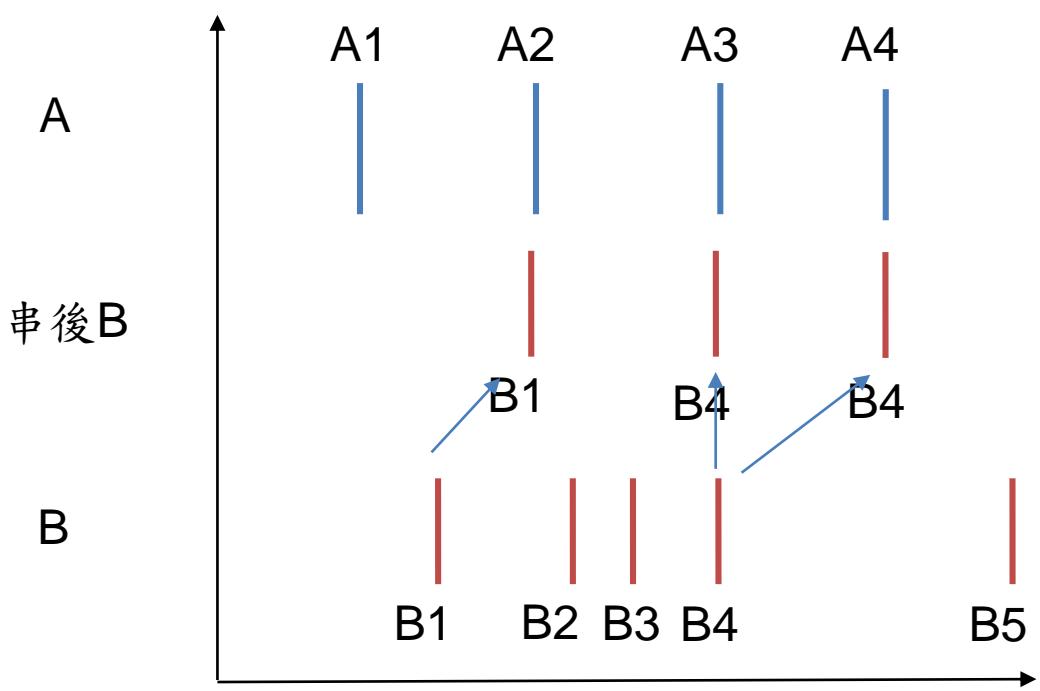
比較表	以 Event 為基準做串聯	以 Periodic 為基準做串聯
記錄方式	有"事件"才記錄。 例如機台換模、停機、人為調機等	固定"週期"記錄。 例如1小時一次
串接前 表單特性	資料筆數通常較少且稀疏	資料筆數通常較完整
串接後優點	資料較完整 (串接後可能遺漏值較少)	可觀察週期性變化
串接後缺點	可能有某"長"時間區段無資料	資料可能有部分缺失 (串Event會造成大量遺漏值)
建議串接 方法	Rolling Forward Rolling Backward Nearest time	Rolling Forward Rolling Backward Nearest time
目的或 使用時機	Troubleshooting	Monitoring

Lee and Dong (2019)

資料合併- Rolling forward (過去歷史資料當中離現在最近的填進來)

	Date	A
1	2016-01-01	A1
2	2016-04-01	A2
3	2016-07-01	A3
4	2016-10-01	A4

	Date	B
1	2016-02-20	B1
2	2016-05-01	B2
3	2016-06-15	B3
4	2016-07-01	B4
5	2016-12-31	B5

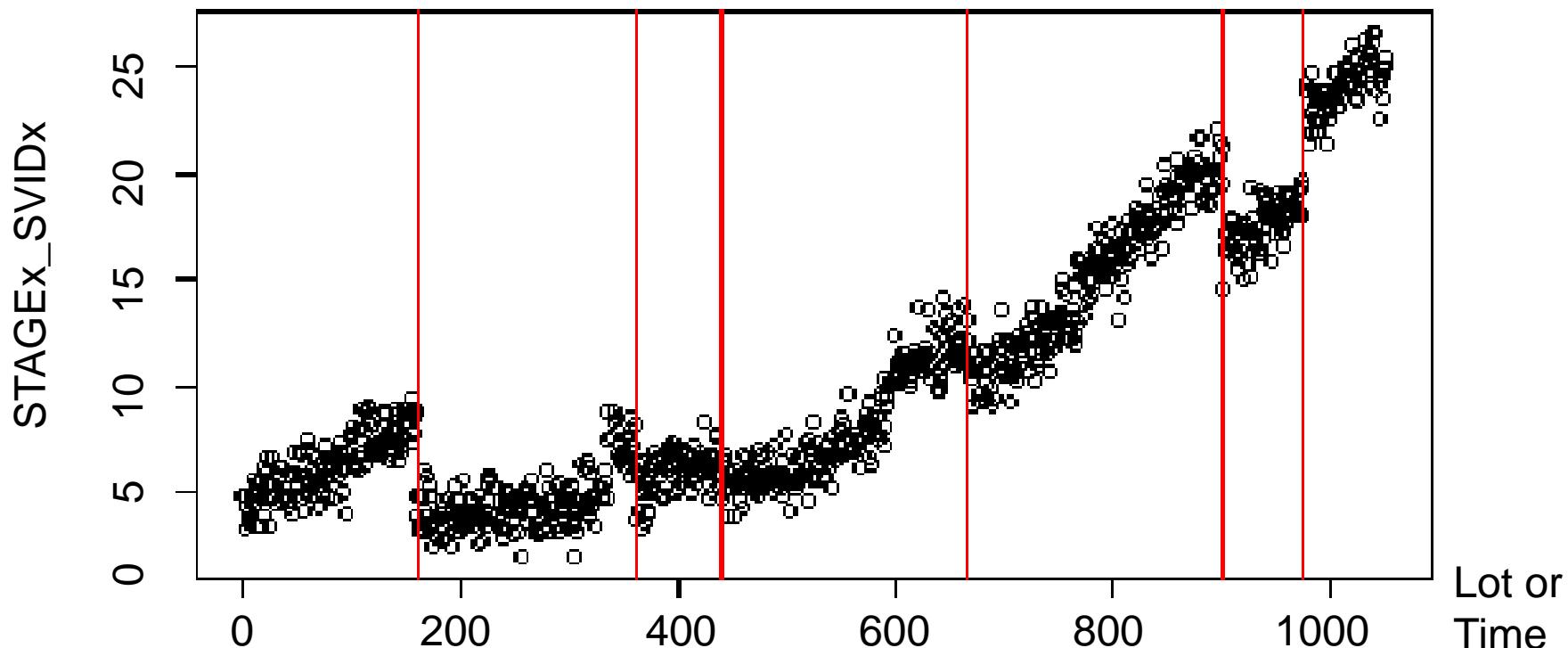


	Date	A	B
1	2016-01-01	A1	
2	2016-04-01	A2	B1
3	2016-07-01	A3	B4
4	2016-10-01	A4	B4

□ 維修保養

- 定期保養(年保、季保、月保、機台生產10,000產品...)
- 若有收集保養資料，可直接找出**保養時間點**。若無，可藉由推估
 - 機台up與down的時間 (Overall Equipment Effectiveness, OEE)
 - 產品的queue time
 - 機台參數(e.g. status variable identification, SVID)的監控

再與相關部門(例如設備)進行確認。



□ 維修保養troubleshooting

- 機台故障表單 + MES (含Recipe跟使用的零組件材料) + 良率
- 針對某一機台ID，用”時間”進行資料的合併串接

機台故障表單			MES			良率		
Time	Down code	Repair		Recipe	Part	Material		Yield
2017-05-07 14:05:28	Run	No或NA		Recip18	Part01	Mater05		94.3%
2017-05-07 16:12:14	Run	No		Recip18	Part01	Mater05		93.1%
2017-05-07 17:41:30	Down04	Part19 (換零件)	...	Recip18	Part19	Mater05	...	82.5% 或內插
2017-05-07 19:22:43	Run	No		Recip18	Part19	Mater05		82.5%
2017-05-07 20:18:17	Run	No		Recip02	Part19	Mater10		76.7%

$93.1\% \times [(17:41:30) - (16:12:14)] + 82.5\% \times [(19:22:43) - (17:41:30)]$

$(19:22:43) - (16:12:14)$

□ Data/Class Imbalance原因

- # of qualified product extremely dominates the # of defective product
- 資料不平衡大多發生於類別型態的資料上(一般泛指兩類)，若以連續分佈的資料來說，資料不平衡代表資料可能集中在某些區段，而這些區段也可以稱作”群/類別”。
- 資料不平衡的情況可能出現在獨立變數或是相依變數。

□ 資料多不平衡才算不平衡？

- For the two classes (0 and 1), rule of thumb...
 - 10% vs. 90%? 5% vs 95%? or 1% vs. 99%?
 - It depends... on your industry applications.
- From a theoretical viewpoint, it occurs if it skews the model training for prediction...
- 也就是說，如果你訓練的模型準確率”異常地高”
 - Overfitting? Class Imbalance?

Lot ID	X1	...	X100	Inspection
Lot01				PASS
Lot02				PASS
Lot03				PASS
Lot04				PASS
Lot05				PASS
Lot06				PASS
Lot07				FAIL
Lot08				PASS
Lot09				PASS
Lot10				PASS
Lot11				PASS
Lot12				PASS

□ 預測Inspection的結果

- 由於只有1筆FAIL
- 預測模型全部都猜PASS
 - 不需要分析變數X1~X100
 - 準確度可達 $11/12 = 91.7\%$

攝氏溫度200度是100度的兩倍？

衡量的層次	內容說明	案例
名目尺度 (nominal scale)	衡量的數字僅是作為代碼，數字大小不具任何意義，也不能做數學運算	機台的編號、貨批的編號等。
類別尺度 (categorical scale)	衡量的數字僅是用來表示歸屬的類別，因此類別尺度的資料可以重複	先對縣市編碼，再進一步歸類成北、中、南、東地區。
順序尺度 (ordinal scale)	衡量的數字表示方案之間的大小順序關係	各種Bin值的分類
間距尺度 (interval scale)	衡量的數字可有意義地描述並比較數字之間的差距大小。無固定原點，也可以調整分隔的間距大小	機台的溫度、量測的參數、學業成績
比率尺度 (ratio scale)	衡量的數字可做比率倍數的比較。有固定原點	化學藥劑使用量、重量
絕對尺度 (absolute scale)	所衡量的數字具有絕對的意義，無法再做其他有意義的轉換	機率、自然數

(簡禎富、許嘉裕，2014)

資料整合與清理

問題	原因	步驟
不正確的資料	資料的值超出合理範圍	
不一致的資料	不同來源資料整合後所出現的分歧 數值不一致、資料內容不一致、欄位不一致	
重複的資料 (Duplication)	重複記錄的欄位或數值 (data type: single, double...) (同樣的資料卻不同的寫法， “做了36顆”，“打出36粒”，“生產36個”，“左上角區塊有產生 defects”，“defects發現於左上方區域”)	資料整合
冗餘的資料 (Redundant)	出現相同意義的資料或欄位 具有相同意義或彼此間存有已知數學關係的欄位，此變數 的屬性或意義可由另一變數推導而得 (有些冗餘資料可以經 由相關分析偵測到) eg. 地址vs.地區	
遺漏值	量測設備或人為因素所造成的資料遺漏	
雜訊	資料本身的誤差或資料輸入的偏差	資料清理
離群值	資料本身的特性、不當量測或資料輸入錯誤	
資料尺度不適	資料格式不符合挖礦工具的假設	資料轉換 (正規化)
資料太多	資料或維度過高	資料化約

(簡禎富、許嘉裕，2014)

Data Source	Scale	Issues
Production data (MES)	Categorical/continuous <i>/time</i>	High dimension, multicollinearity, class imbalance, missing value
Equipment data	Categorical/continuous	High dimension, too many categorical levels, time series, missing value
Parts/Supplier data	Categorical	Too many categorical levels
Transportation data	Categorical/continuous	too many categorical levels, time series, missing value
Maintenance/Repair Data	Binary/categorical/continuous	Typing error, text, missing value, Choosing “others” or “NA”
Testing/Inspection Metrology data	Binary/continuous/ <i>figure</i>	Sampling data, time series, multi-response, metrology delay

Revised from Chen (2015)

遺漏值填補是...補「資料」？

	English	Math
Student_A	80	76
Student_B	80	91
Student_C	80	83
Student_D	80	62
Student_E	80	?
Avg.	80	

Max: 91

Min: 62

Avg: 78

- 填補遺漏值一般會造成部分失真或偏差
- 使用者應根據製造”資料特性”以及”分析目的”，來決定填補遺漏值的方法，以避免忽略原本應有的資訊
- 方法(補值一定要找關係!!!!)
 - 忽略變數值 (“N/A” 與 “0” 是不一樣的!!)
 - 移除觀測值 (remove the tuple)：當依變數Y遺漏時
 - 人工填寫遺失值
 - 使用一個全域常數填充遺漏值 (eg. N/A)
 - 使用屬性平均值
 - 使用與給定變數值屬於同一類別的所有樣本之平均值
 - 模型: 簡單/多元線性迴歸、類神經網路、最鄰近估計法K-Nearest Neighbor (KNN)、Random Forest..."MICE" ..

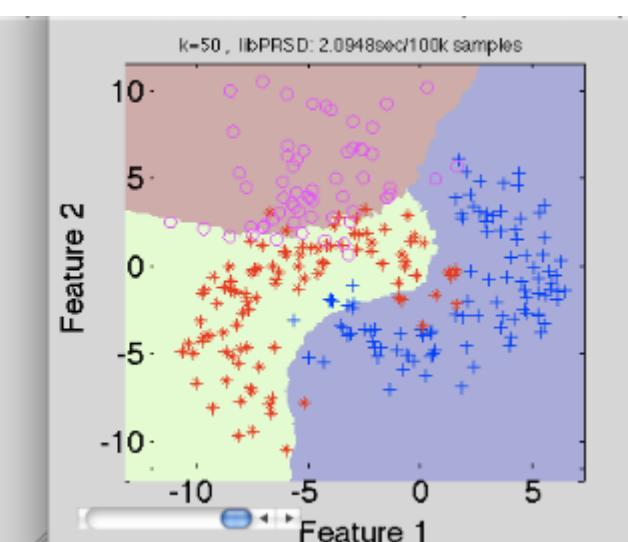
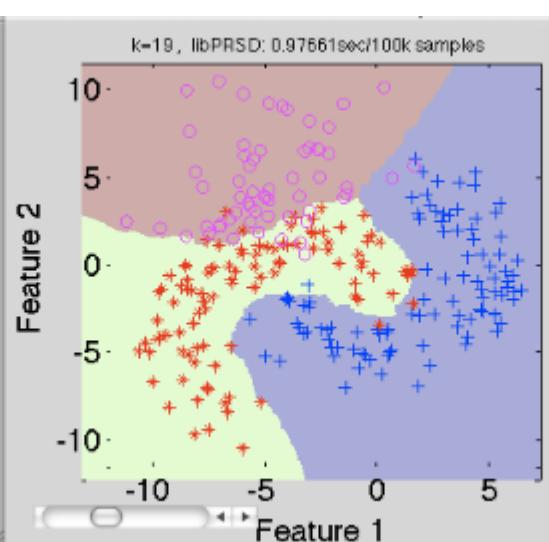
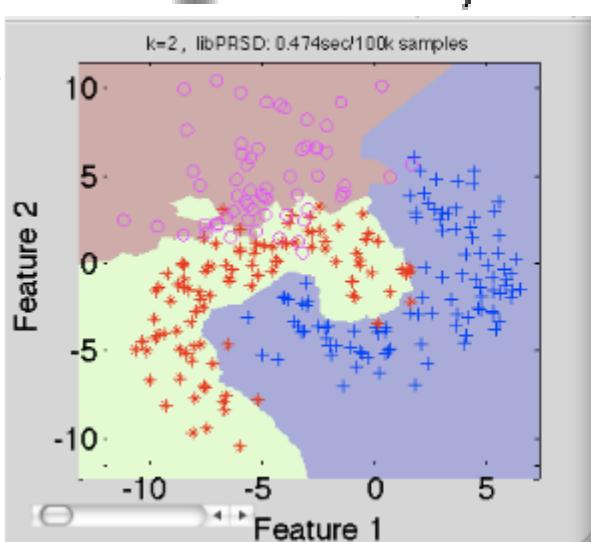
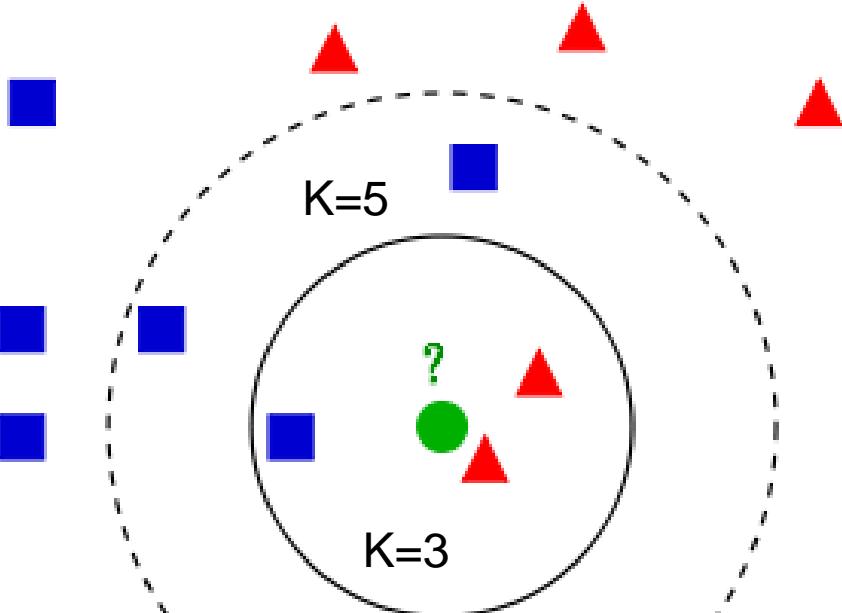
□ 補值一定要找關係!!!!

- 利用其他變數與遺漏值之間的關係來估計遺漏值
- 補值：可利用其他變數與遺漏值之間的關係來估計遺漏值
- 例如，若「收入水準」變數發生遺漏值，或許可能用「房子坪數」這變數來做預測

□ 假設在現有的資料庫中發現某一顧客其購買反應的態度為一遺漏值

顧客	性別	年齡	薪水	購買反應
A	女	27	\$19,000	No
B	男	51	\$64,000	Yes
C	男	52	\$105,000	Yes
D	女	33	\$55,000	Yes
E	男	45	\$45,000	No
F	女	45	\$100,000	?

最鄰近估計法K-Nearest Neighbor (KNN)



Wikipedia, <https://zh.wikipedia.org/wiki/%E6%9C%80%E8%BF%91%E9%84%B0%E5%B1%85%E6%B3%95>

perClass, 2017. kb16: Visualize the effect of a change of parameters in a trained classifier. <http://perclass.com/doc/kb/16.html>

□ 不偏估計量 VS. 變異程度

觀測值	原始資料值	第 11 筆遺漏	利用平均數估計	利用標準差估計
1	0.0886	0.0886	0.0886	0.0886
2	0.0684	0.0684	0.0684	0.0684
3	0.3515	0.3515	0.3515	0.3515
4	0.9874	0.9874	0.9874	0.9874
5	0.4713	0.4713	0.4713	0.4713
6	0.6115	0.6115	0.6115	0.6115
7	0.2573	0.2573	0.2573	0.2573
8	0.2914	0.2914	0.2914	0.2914
9	0.1662	0.1662	0.1662	0.1662
10	0.44	0.44	0.44	0.44
11	0.6939	?	0.3731	0.6622
平均值	0.4023	0.3731	0.3731	0.3994
標準差	0.2785	0.2753	0.2612	0.2753
誤差值			0.3208	0.0317

(簡禎富、許嘉裕，2014)

□ Difficulty in Data Collection

- Multiple sources

- Different data type/format
- Redundant dataset
- Difficulty in cross-functional/interdepartmental data collection (跨部門資料
非常難拿)

- 遺漏值

- 人為？系統？記錄格式不對？抽樣？

- 資料收集的時間與產品狀況

- 某系統可能某產品別有資料，其他產品沒有 (每個系統不一樣)
- 某系統可能兩小時收集一次，另一系統是每小時收集一次
- 某系統資料只有收集8am-6pm，另一系統是24hrs
- 同樣的資料卻不同的寫法，例如...
 - “做了36顆”、“打出36粒”、“生產36個”
 - “左上角區塊有產生defects”、“defects發現於左上方 區域”

□ 欄位變數遺漏值太多(e.g. 80%)該不該填？

- 通常可以刪除該欄位，但如果欄位個數p太少，則要留意！

X_{14}

High Correlation

X_{70}

20%
有值

80%
遺漏

□ 建議方法

- X_{70} 先刪除，
- 如果分析後發現 X_{14} 是重要變數(e.g. β 的p-value<0.01)
- 但 X_{14} 不具解釋Y的物理特性與因果關係
- 此時應留意是否考慮用 X_{70} 來做解釋

□ How about 觀測值遺漏值太多(e.g. 80%)該不該填？

- 如果觀測值個數n夠多的話，可考慮刪除
- 如果進行80%的填補，若兩個觀測值在80%填補一模一樣(e.g. 填平均)
- 則會造成兩個觀測值在向量空間幾乎重疊
- 例如迴歸中很多觀測值落在同一個點附近，造成建模出現偏差

Lee (2019)

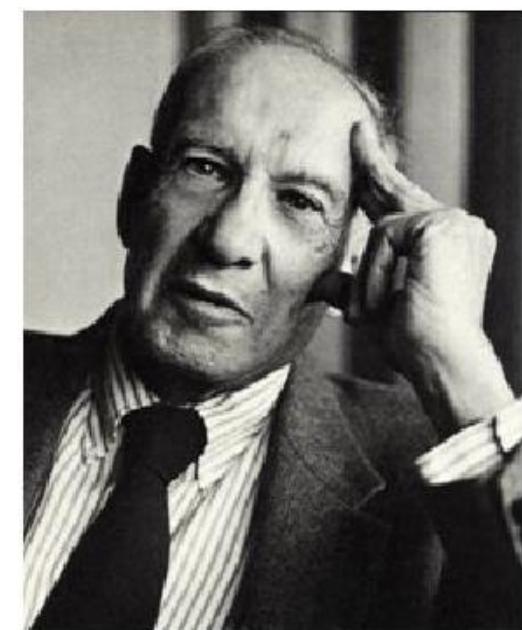


“In God we trust, all others must bring data”

- Edward Deming (1900-1993)

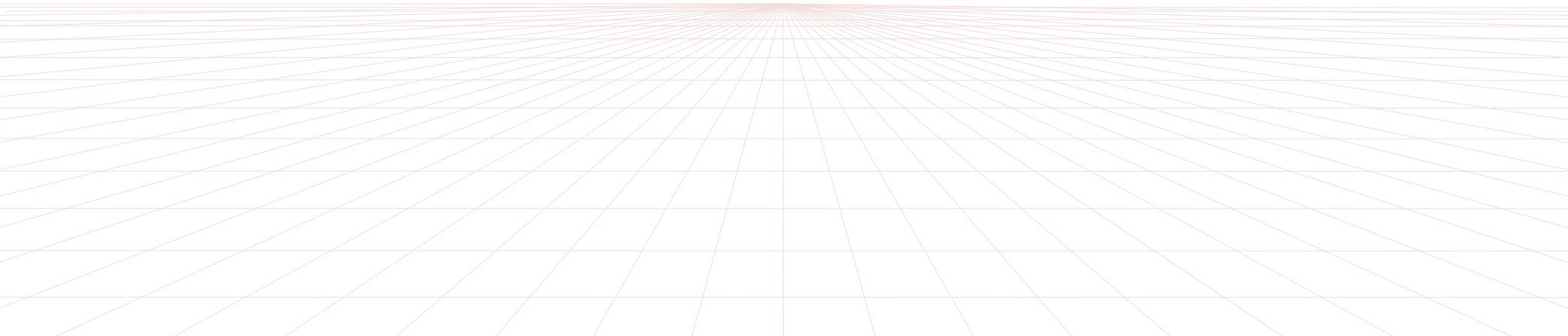
“What gets measured, gets managed”

- Peter Drucker (1909-2005)



重要參數篩選與精度預測

Feature Selection and Quality Prediction



- 老師...大家正在談工業4.0，我老闆想做，所以...
- 老師...如何觀察產線上不易察覺的浪費？
- 老師...庫存或在製品WIP太多該如何處理？
- 老師...機台稼動率不高怎麼辦？
- 老師...我們想做排程(或想替代現有的人工排程)...
- 老師...需求該如何預測？
- 老師...我有一堆資料，來Mining一下吧~
- 老師...製程影響因子太多，如何找出關鍵因子？
- 老師...資料量太大了，無法分析(一秒25,600筆資料...)
- 老師...我不敢刪資料因為怕失真..但是又太大了...
- 老師...模型挑選關鍵參數與工程經驗不一致如何處理？
- 老師...資料量少，因為現場都是有問題後才開始收集資料

案例一

Process Diagnosis / Troubleshooting (製程診斷)

案例二

Feature Selection (重要工程參數篩選)

案例三

Quality/Yield/Spec Prediction (良率/精度預測)

案例四

Predictive Maintenance (預測保養)

案例一：製程診斷

□ 不良品產生了，怎麼辦？

- 先別急著追究責任
- 是否需要“停線”以減少不良品的持續產生？

□ 屬正常損耗

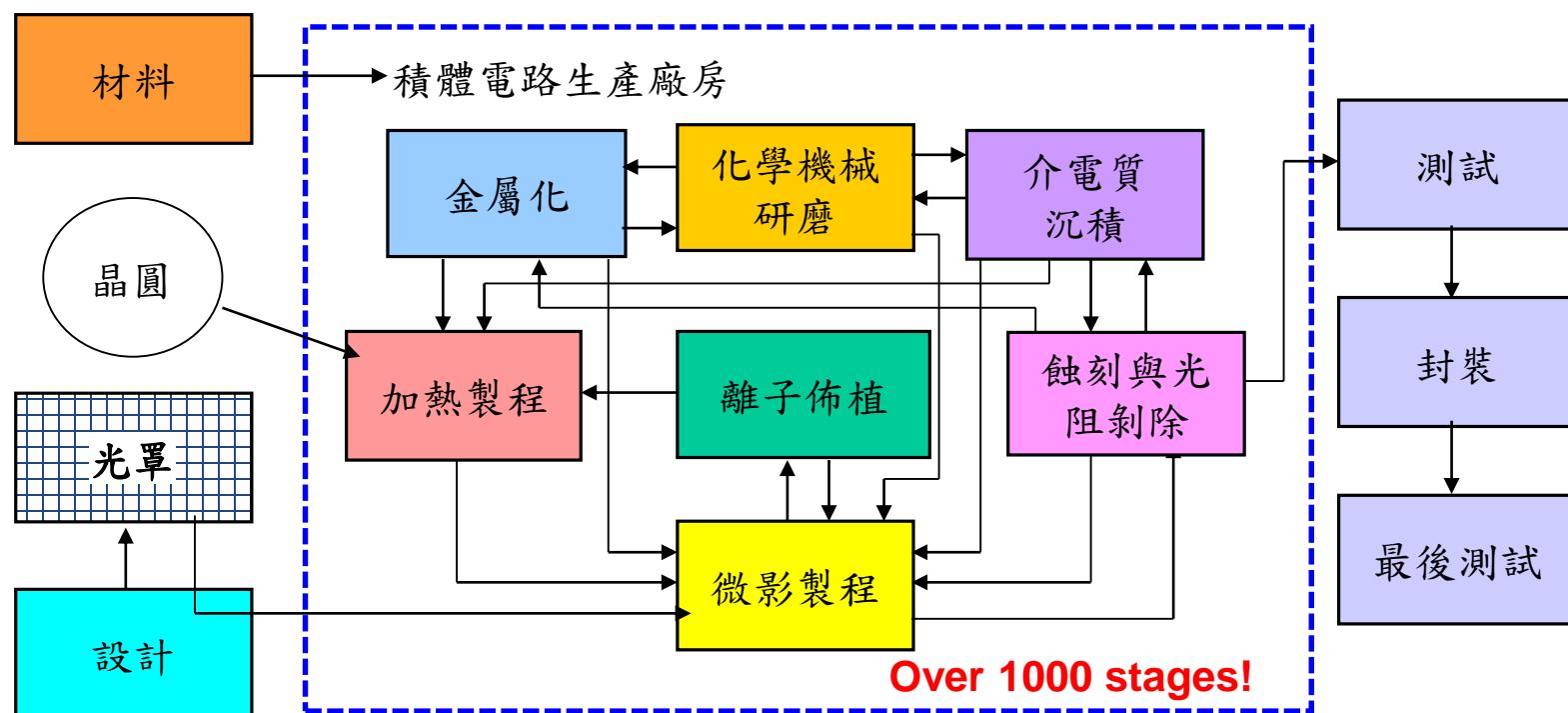
- 例如：調規格、換模換線初期 → 技術員的經驗與教育訓練
- 例如：不可控制的隨機因子、空氣中的particle → 良好習慣(軟實力)或無塵室規模更新(硬實力)
- 例如：新產品的開發與設計、實驗貨 → RD人員 實驗設計(DOE)與教育訓練
- 應作“長期規畫”改善，降低精實七大浪費，透過“制度設計”讓組織成長

□ 為不正常損耗

- 例如：Recipe設置不當、機台汙染、MO發生 → 教育訓練
- 例如：設備、模具、零組件的磨損 → 例行維修保養是否落實
- 例如：原材料品質變異(eg.長期存放) → 驗收入庫、先進先出
- 當下應透過“人機料法環”探究根本原因(root-cause)進行問題排除。如何透過“制度設計”以避免下次再發生…

□ 半導體製程良率的挑戰

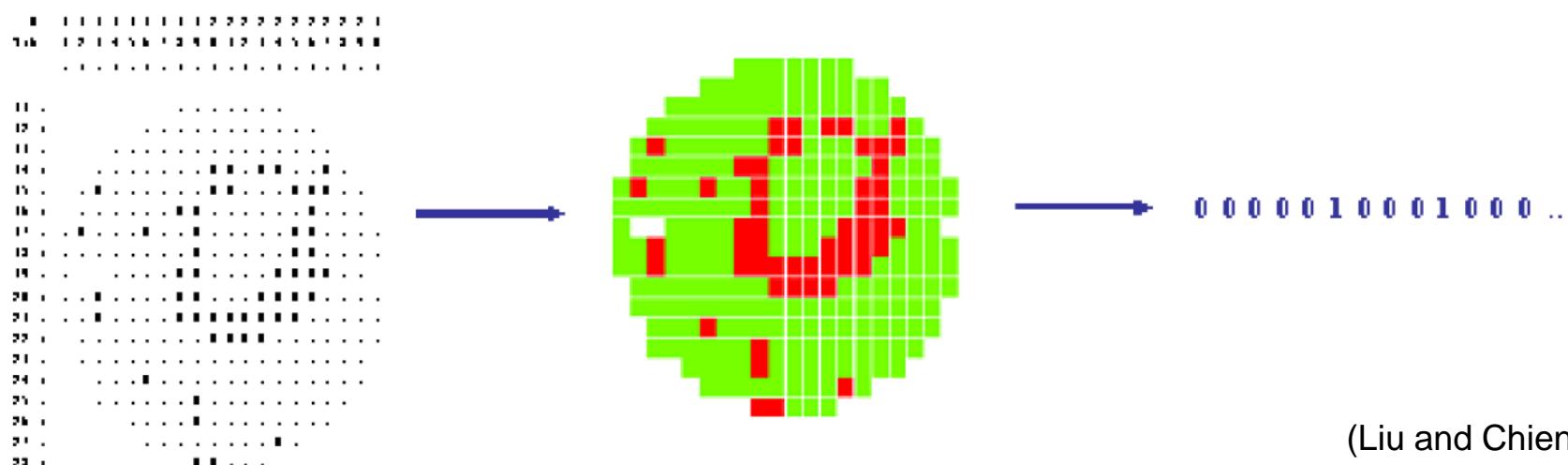
- 生產製程冗長且複雜，前後製程具**交互影響**
- 線寬的微縮製程容忍度下降
- 人、機、料、法、AMHS的整合困難
- 量測機台的有限產能及生產週期時間(cycle time)的要求，採用抽樣檢測(sampling)的方式不易即時偵錯



蕭宏(2004)，半導體製程技術導論，Prentice Hall.

□ Wafer Bin Map (晶圓圖分類)

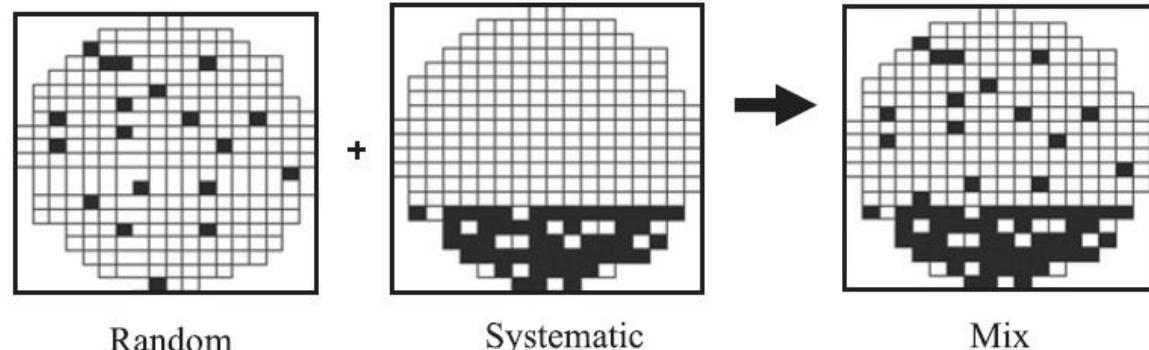
- Wafer bin maps that show specific spatial patterns can provide clue to identify process failures in the semiconductor manufacturing.
- In practice, most companies rely on experienced for troubleshooting. However, as IC feature size is continuously shrinking, WBM patterns become complicated and thus human judgments become inconsistent and unreliable.
- In the semiconductor fabrication process, the circuit probe (CP) test is used to detect specific failures of each die and thus indicate the test results with the corresponding bin values.



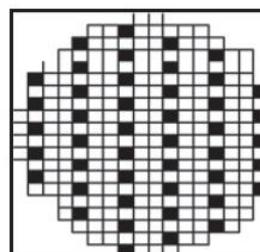
(Liu and Chien, 2013)

案例1: Process Diagnosis/ Troubleshooting (製程診斷) POOLab

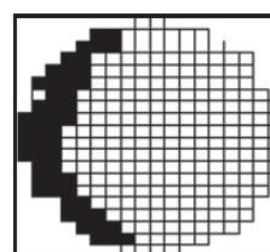
3 3 4
1 4 6 4 2 4 4 6 4 4
4 1 4 2 2 2 4 1 6 6 6 3 4 4
4 4 4 4 5 1 5 4 4 1 4 3 4 4 3 6
3 2 4 2 2 4 4 4 1 4 4 4 5 5 4 4
3 4 1 4 1 1 1 1 4 4 4 4 5 4 4 4 4
1 1 1 3 4 4 3 4 4 1 5 1 4 4 4 6 4 4
1 1 1 4 4 4 1 1 1 1 1 1 4 1 4 3 4 4
1 4 4 5 1 1 1 1 1 1 1 1 4 4 5 4 4
1 4 1 1 4 1 1 1 1 1 1 1 5 4 4 4 4 4
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 4 3 3
4 1 1 1 1 1 1 1 1 4 4 4 5 5 4 4 4
1 1 1 1 1 1 1 1 1 4 4 4 4 4 1 4 3
1 1 1 1 1 1 1 1 1 4 4 4 4 5 4 1 4
4 1 1 1 1 1 1 1 1 1 1 1 4 4 3 4 3
1 1 1 1 1 1 1 1 1 5 4 3 4 4
1 1 1 1 1 1 1 1 1 1 1 4



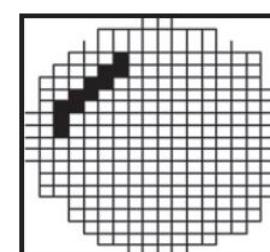
Root Cause



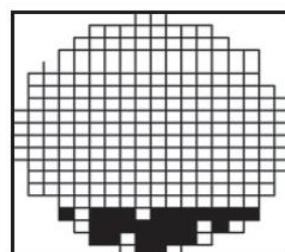
(Mask Error)



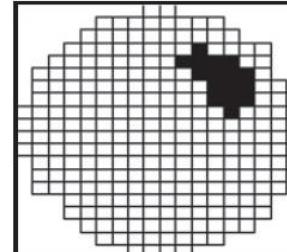
(Process Error)



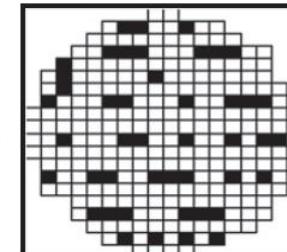
(Scrape Error)



(Probe-Card Error)



(Probe-Pin Error)



(Test-Spec. Error)

□ As-Is

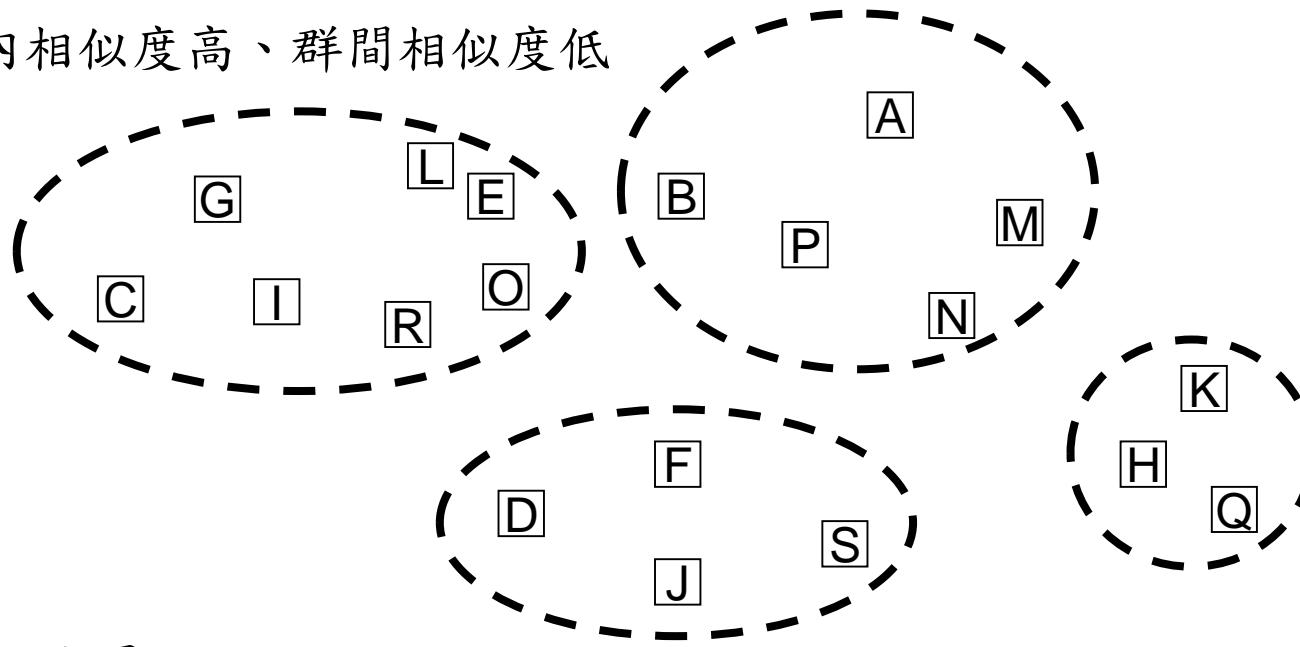
- Engineering Experience
- Troubleshooting Team

□ To-Be

- Data Science
- Automation

□ 分群 (Clustering) (eg. Ward's Method 華德法, 1963)

- 群內相似度高、群間相似度低



□ 空間統計量

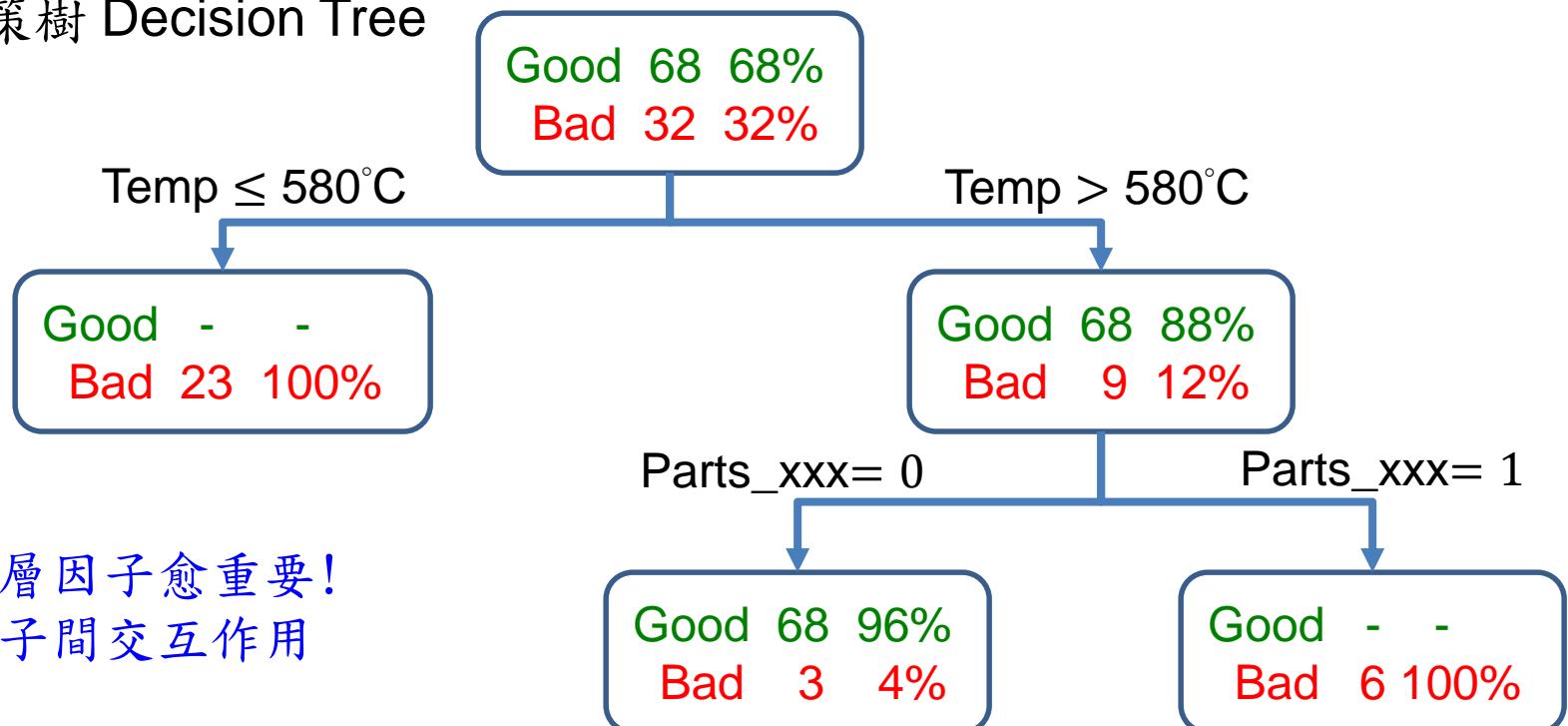
- The spatial correlation of two groups of data can be tested by the odd ratio hypothesis test
- H_0 : 晶圓圖上故障晶片或正常晶片呈現隨機
- H_1 : 呈現非隨機分佈 (即發現有特殊群聚或是離散現象出現)

J (0.5)	J(1)	J(0.5)
J(1)	I	J(1)
J(0.5)	J(1)	J(0.5)

(Hsu and Chien, 2007)

□ 分類 (Classification) (eg. 分類與迴歸樹, CART, Breiman et al., 1984)

- 決策樹 Decision Tree

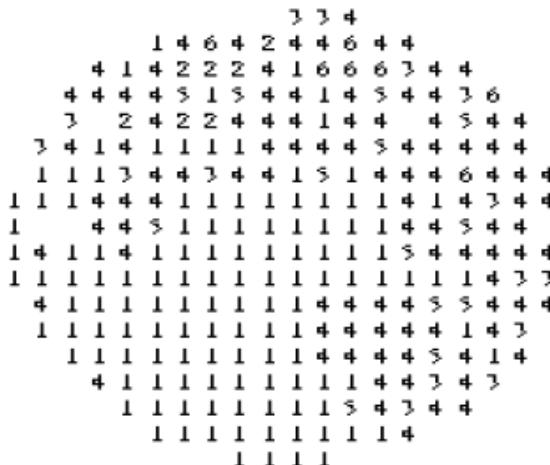


1. 愈往上層因子愈重要!
2. 考慮因子間交互作用

□ Rule Extraction

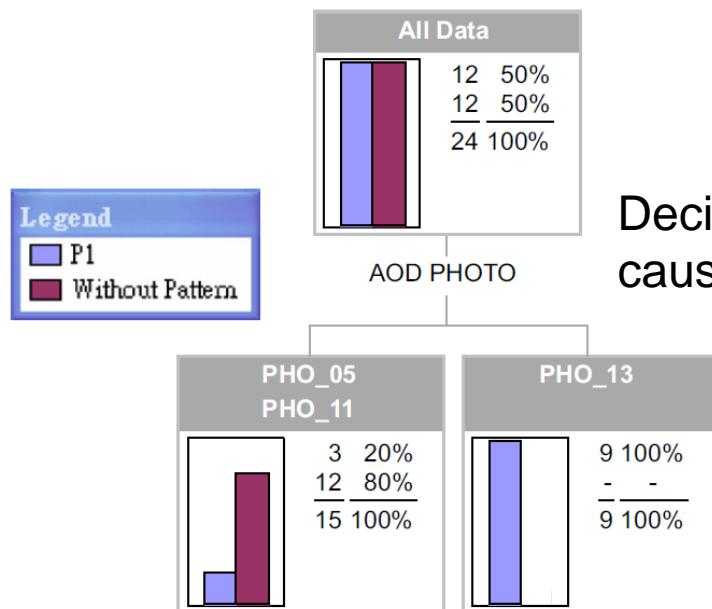
- If $\text{Temp} \leq 580^{\circ}\text{C}$, then “Bad”. (Accuracy: 100%)
- If $\text{Temp} > 580^{\circ}\text{C}$ and $\text{Parts_xxx} = 1$, then “Bad”. (Accuracy: 100%)
- If $\text{Temp} > 580^{\circ}\text{C}$ and $\text{Parts_xxx} = 0$, then “Good”. (Accuracy: 96%)

案例1: Process Diagnosis/ Troubleshooting (製程診斷) PQLab

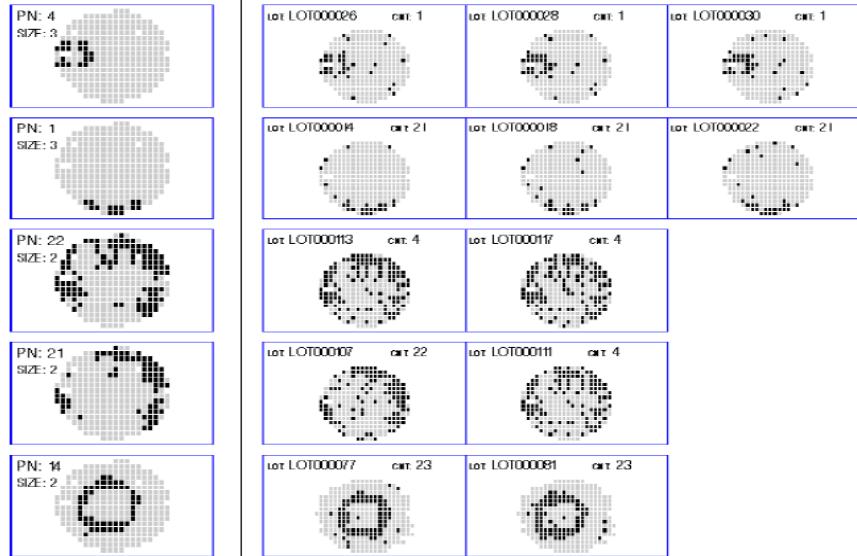


Map clustering for systematic defects

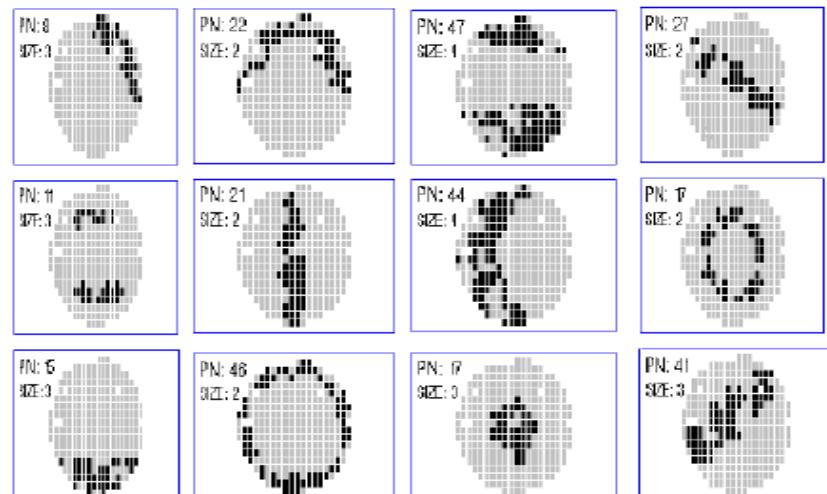
(Hsu and Chien, 2007)



Decision tree for root cause classification

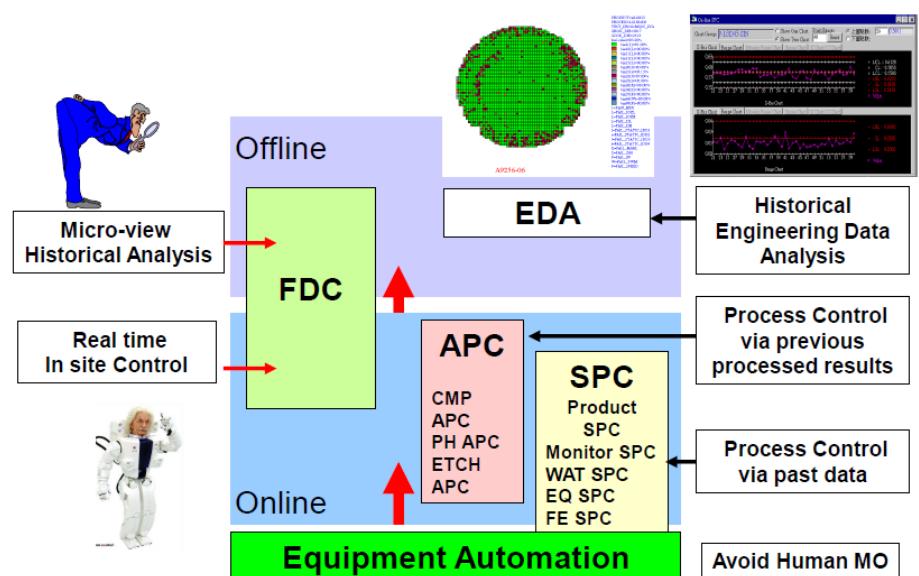
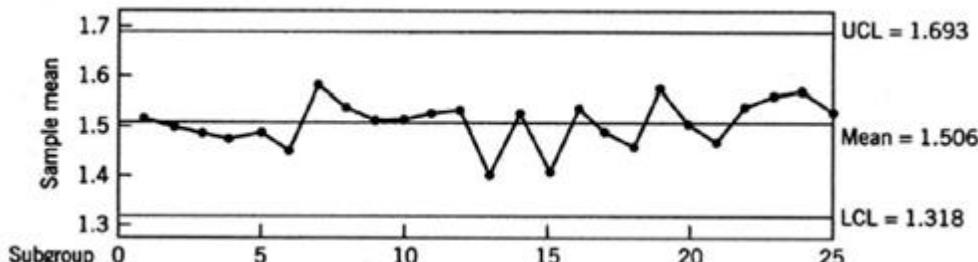


12 defect patterns and build WBM bank



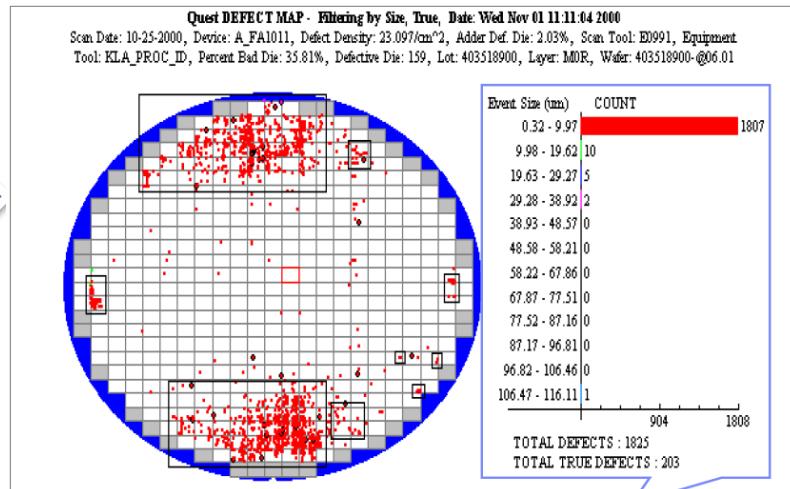
案例1: Process Diagnosis/ Troubleshooting (製程診斷) POLab

□ 良率提升方法論
SPC 製程監控

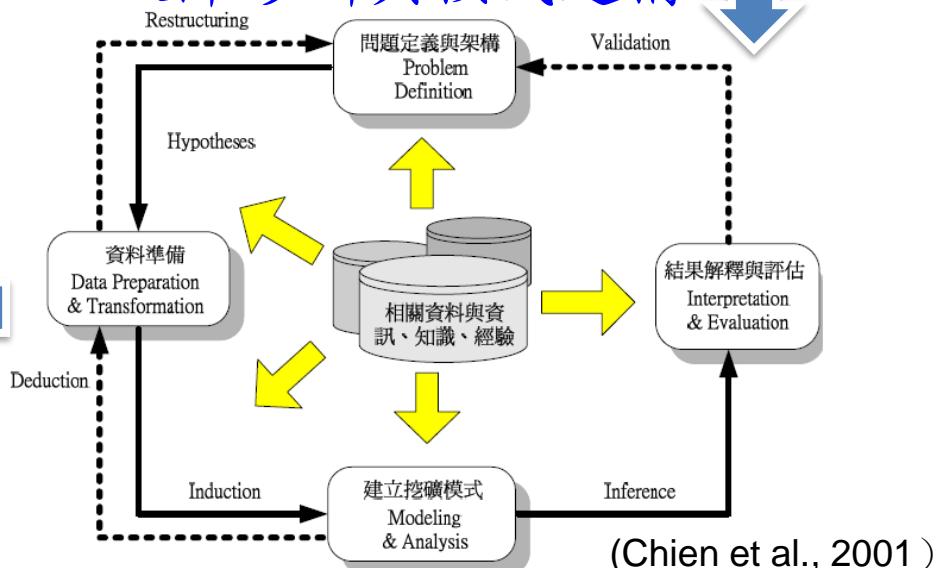


(Wang , 2012) 自動化工程系統

Defect Wafer

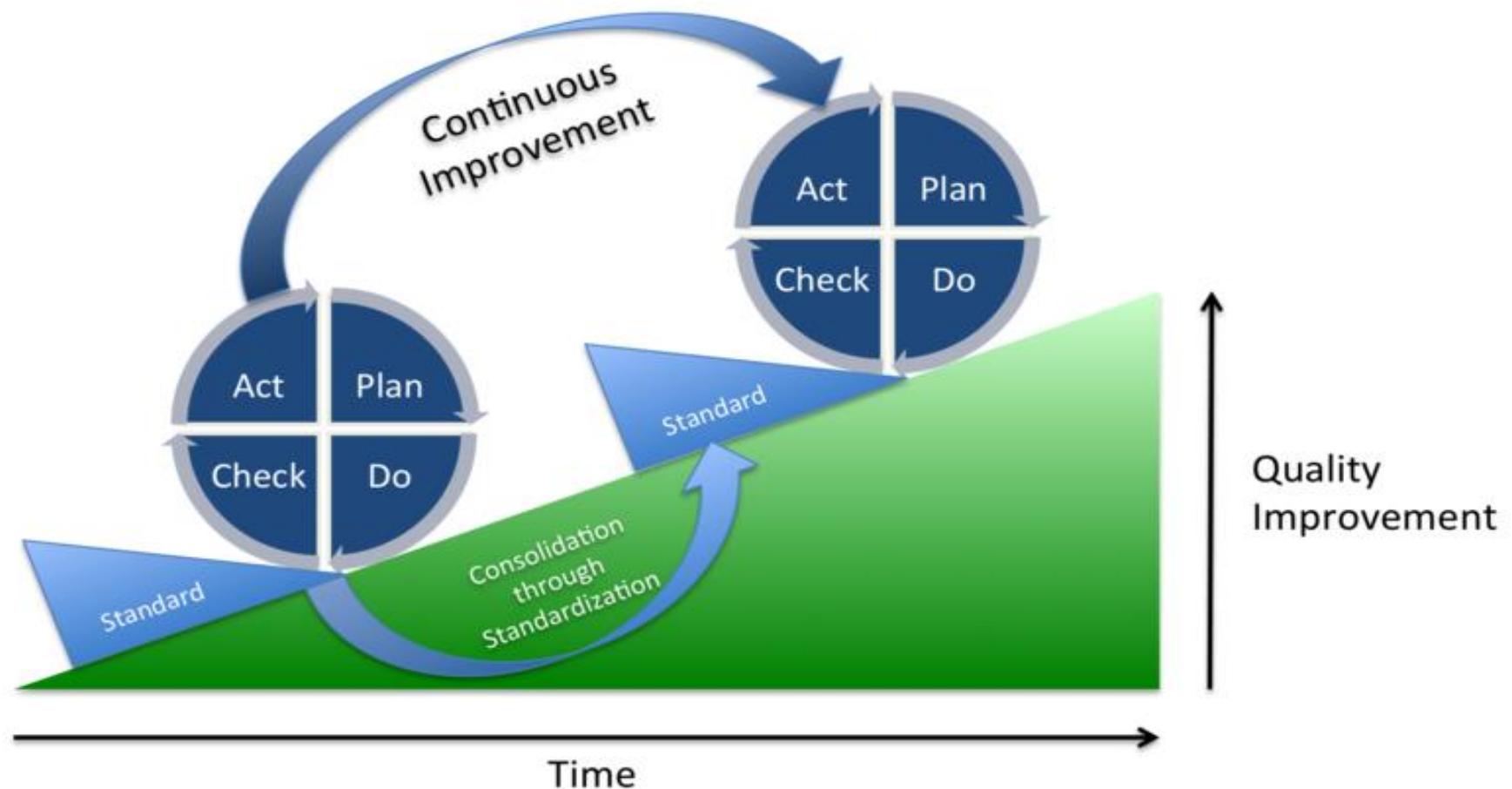


故障診斷與模式建構



(Chien et al., 2001)

Continuous Improvement by PDCA



標準化(standardization) → 制度化(institutionalization)

降低浪費、標準化、與制度化 → 成本降低cost down

Wikiwand (2016). <http://www.wikiwand.com/en/PDCA>

案例二：重要工程參數篩選

□ 為什麼要選重要參數/因子?

- 易於決策判斷

- 看“_____”就知道要不要決定出去旅遊
- 看“_____”就知道這衣服適不適合
- 量“_____”就知道身體健康狀況

□ 製造現場篩選重要因子的目的

- Troubleshooting

- 掌握影響機台品質變異的主要因子、**上下游因子的交互作用**
- Engineering Process Control (EPC)

- 建立管理機制、簡單法則

- 看Bottleneck就可推估現場WIP level

- 精度預測?

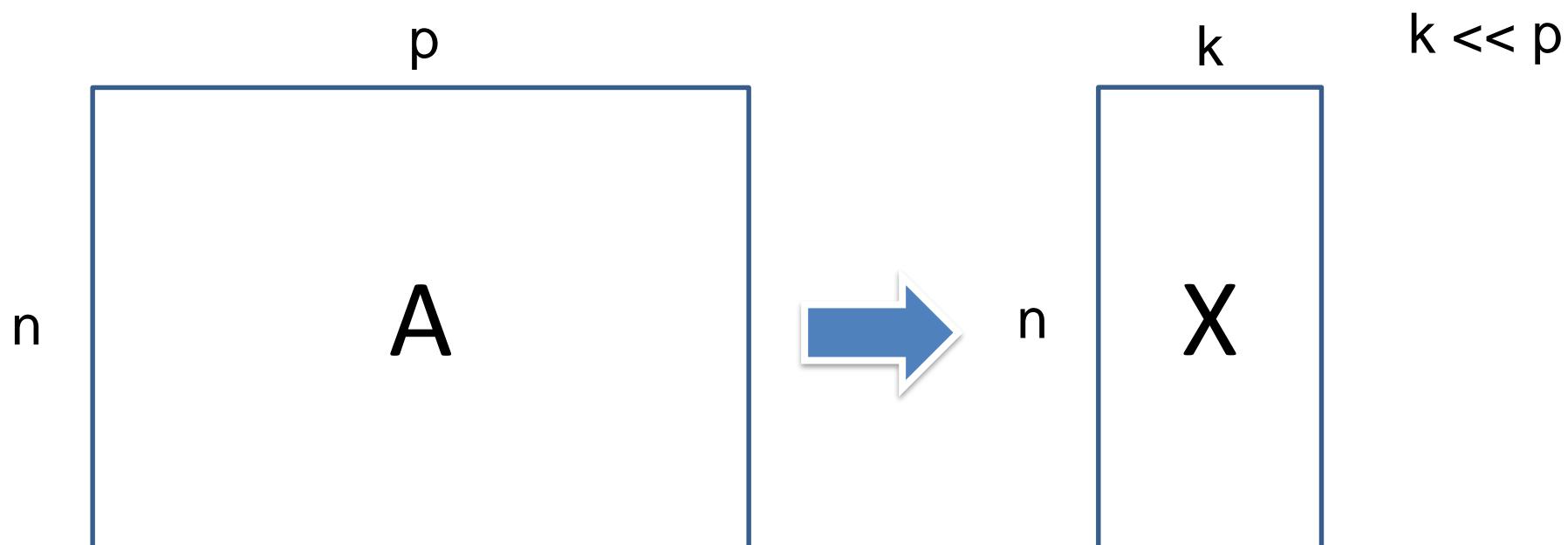
- 提升預測準確度
- On-line real-time prediction

- 監控Monitoring → 用較低成本/較少管制圖

- 環境因子監控、機台參數監控

□ Feature Selection

- In a large-scale dataset, select a few important variables (i.e. columns)
- In particular, # of variables (p) much larger than # of observations (n), i.e. $p \gg n$ issue
- Address “Curse of Dimensionality”
 - The number of observations required exponentially grows to estimate the function or model parameters.



□ Objective (Guyon and Elisseeff, 2003)

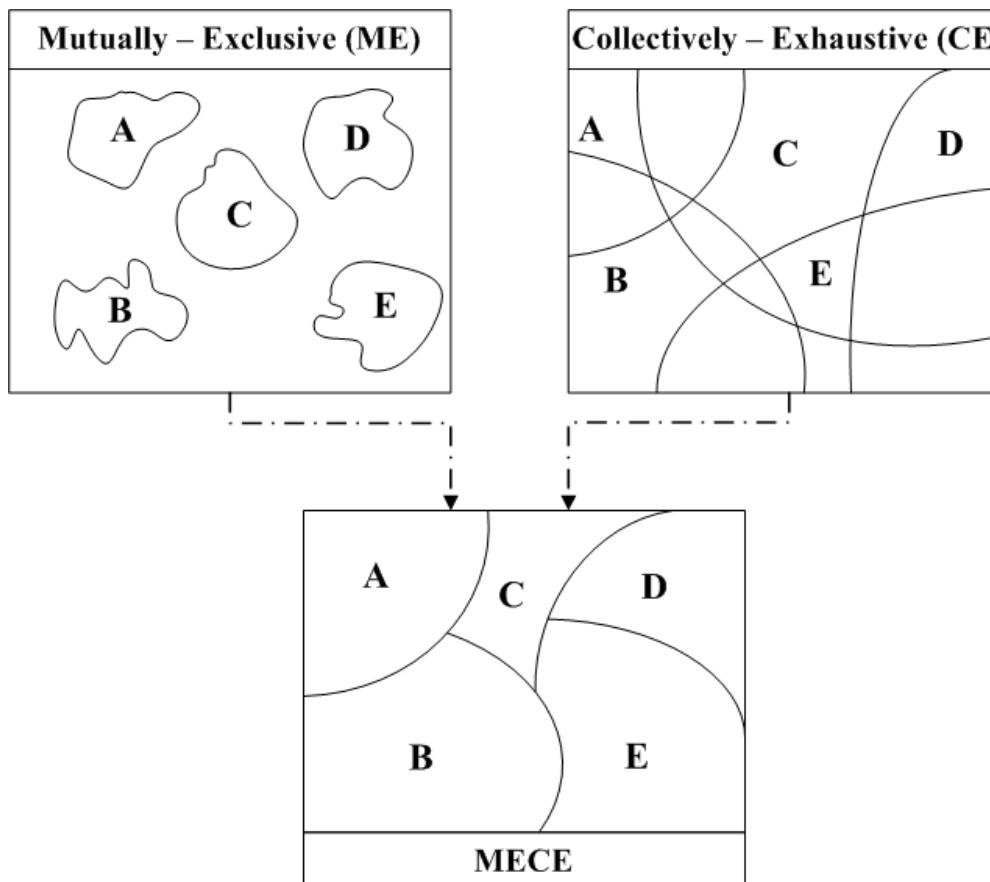
- improving the **prediction performance** of the predictors
- providing faster and more **cost-effective** predictors
- providing a better **understanding** of the underlying process that generated the data. (eg. for process monitoring)

□ Types

- **Variable selection:** select the “best” subset of the existing variables/features without a transformation.
 - Supervised Learning (監督式學習) with “Y” as label
 - Eg. stepwise regression, LASSO, random forest, etc.
- **Feature extraction (variable transformation):** transforming the existing variables into a lower dimensional space
 - Unsupervised Learning (非監督式學習) with only “X”
 - Eg. independent component analysis (ICA), principal component analysis (PCA), etc.

□ 篩選原則(Selection Principle)

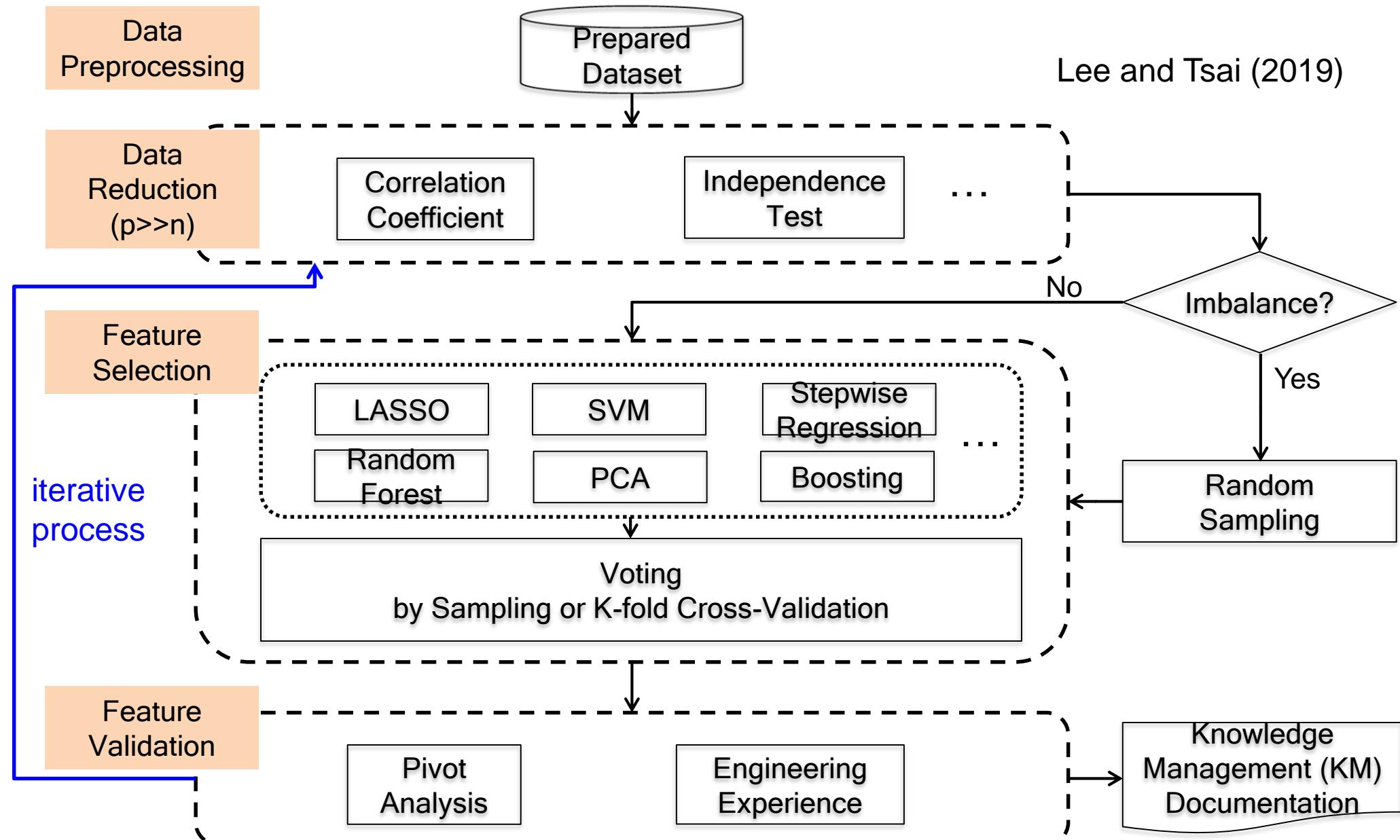
- Mutually-Exclusive Collectively-Exhaustive (MECE) Feature Selection
- claims that the **important** variables should not contain **overlapping** information and should provide **sufficient** information.



Lee and Chen (2018)

案例2: Feature Selection (重要工程參數篩選)

Lee and Tsai (2019)



□ Data Reduction

- 可先假設因子之間無交互作用 (強假設, 要注意!), 先專注於重要變數 (i.e., main effect)
- 根據實際應用個案, 可做以下調整
- 可刪除: 進行x與y的相關係數檢定, 低於某個門檻值
- 可刪除: 中介變數z ($x \rightarrow z \rightarrow y$), 不是z影響y, 實際是x
- 可刪除: 累加或計數變數 (counter), 計數但非製程參數, 不影響z
- 可刪除: 某欄位觀測值都一樣的
- 可刪除: x間高相關可能共線性的變數 (視所使用的模型而定)

□ 變數挑選的邏輯

- 天下雨 → 地濕 (若A則B : A稱為B的充分條件，B稱為A的必要條件)
- 地濕 → 不一定天下雨 (B不一定則A)
- 地沒濕 → 天一定沒下雨 (非B則非A)

□ 同理...

- 因果 → 相關
- 相關 → 不一定因果
- 沒相關 → 絕對沒因果

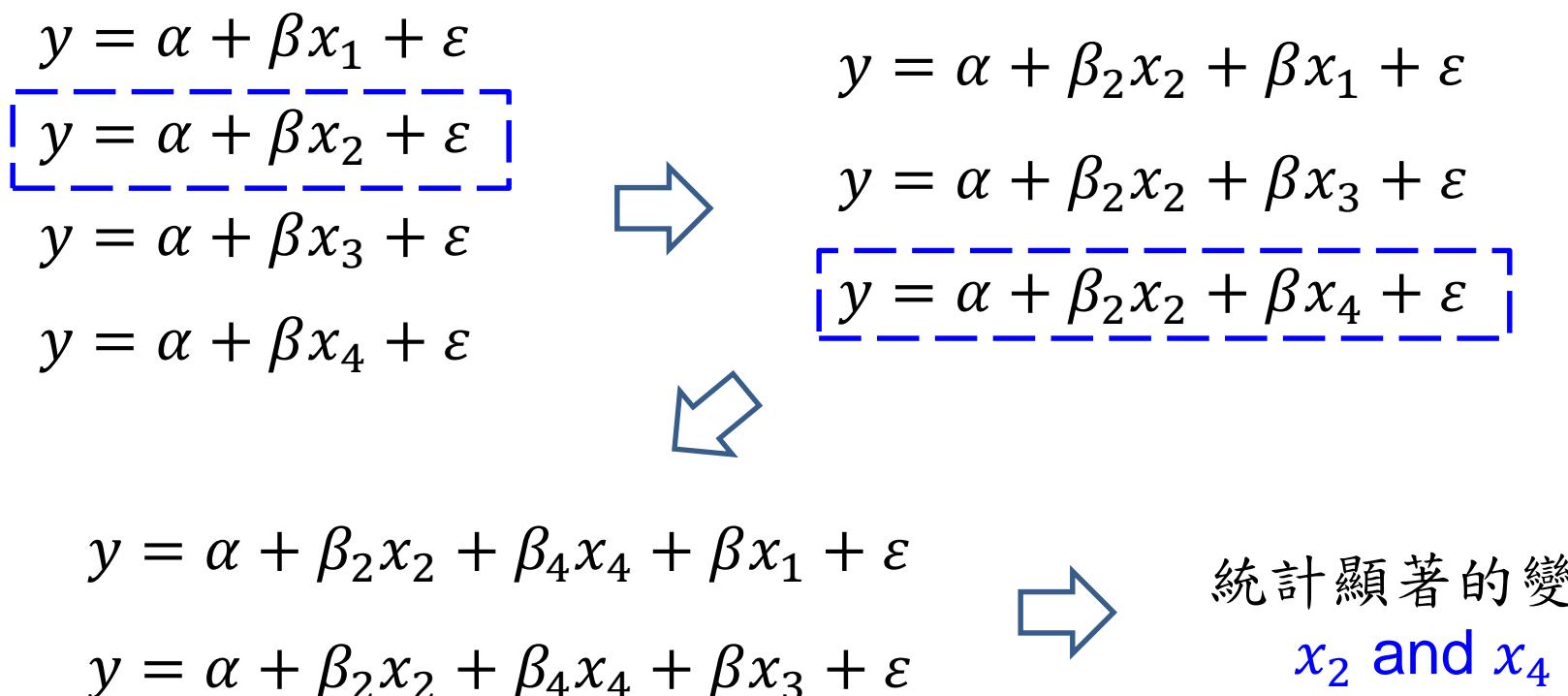
□ 結論

- 當變數欄位太多時，可試著透過X跟Y的相關係數，來刪除不必要的變數欄位 (強假設：因子之間無交互作用)
- 接著應該做 實驗設計(DOE)、Fused Lasso、Tree-based Method, etc. 確認交互作用影響

(Lee, 2019)

□ Stepwise Regression (i.e. 逐步迴歸) (Hocking, 1976)

- Starting with no variables in the model, then adding the variable (if any) that improves the model the most (with smallest p-value).
- Forward selection (y, x_1, x_2, x_3, x_4)

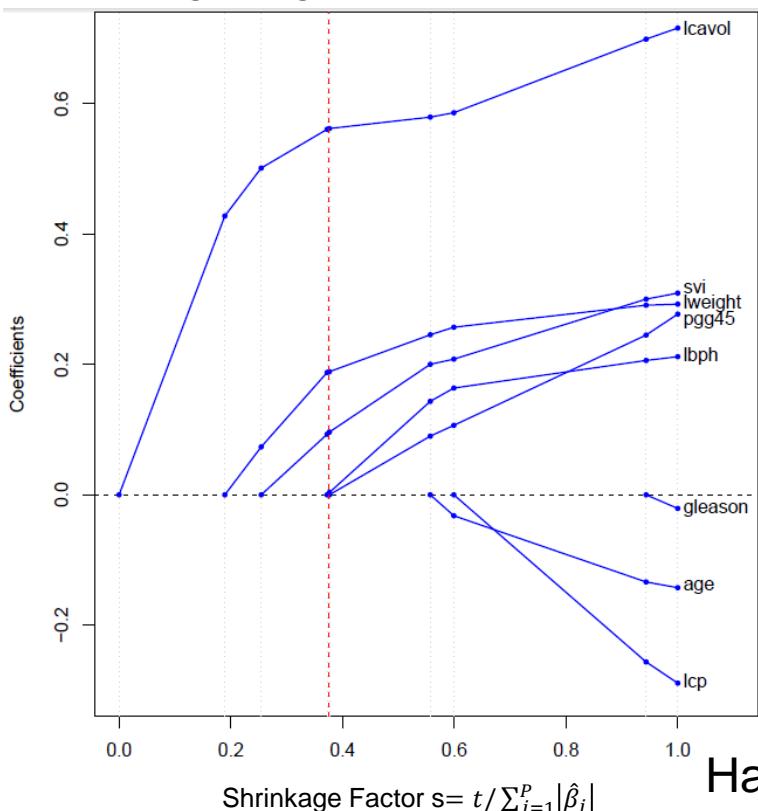


□ LASSO Regression (Tibshirani, 1996)

- $\hat{\beta}^{lasso} = \operatorname{argmin}_{\beta} \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^P x_{ij}\beta_j)^2$

Subject to $\sum_{j=1}^P |\beta_j| \leq t$

- Lagrangian form $\hat{\beta}^{lasso} = \operatorname{argmin}_{\beta} \left\{ \frac{1}{2} \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^P x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^P |\beta_j| \right\}$



Penalty(λ)	x1	x2	x3	x4	x5	x6	x7
$\lambda = 0$	*	*	*	*	*	*	*
$\lambda = 40$	*	*	*	*	*	*	*
$\lambda = 41.8$		*	*	*	*	*	*
$\lambda = 41.9$		*	*	*	*		*
$\lambda = 42$			*	*	*		*
$\lambda = 50$			*	*			*

Hastie et al. (2008)

Lee and Cai (2019)

□ Voting (投票法)

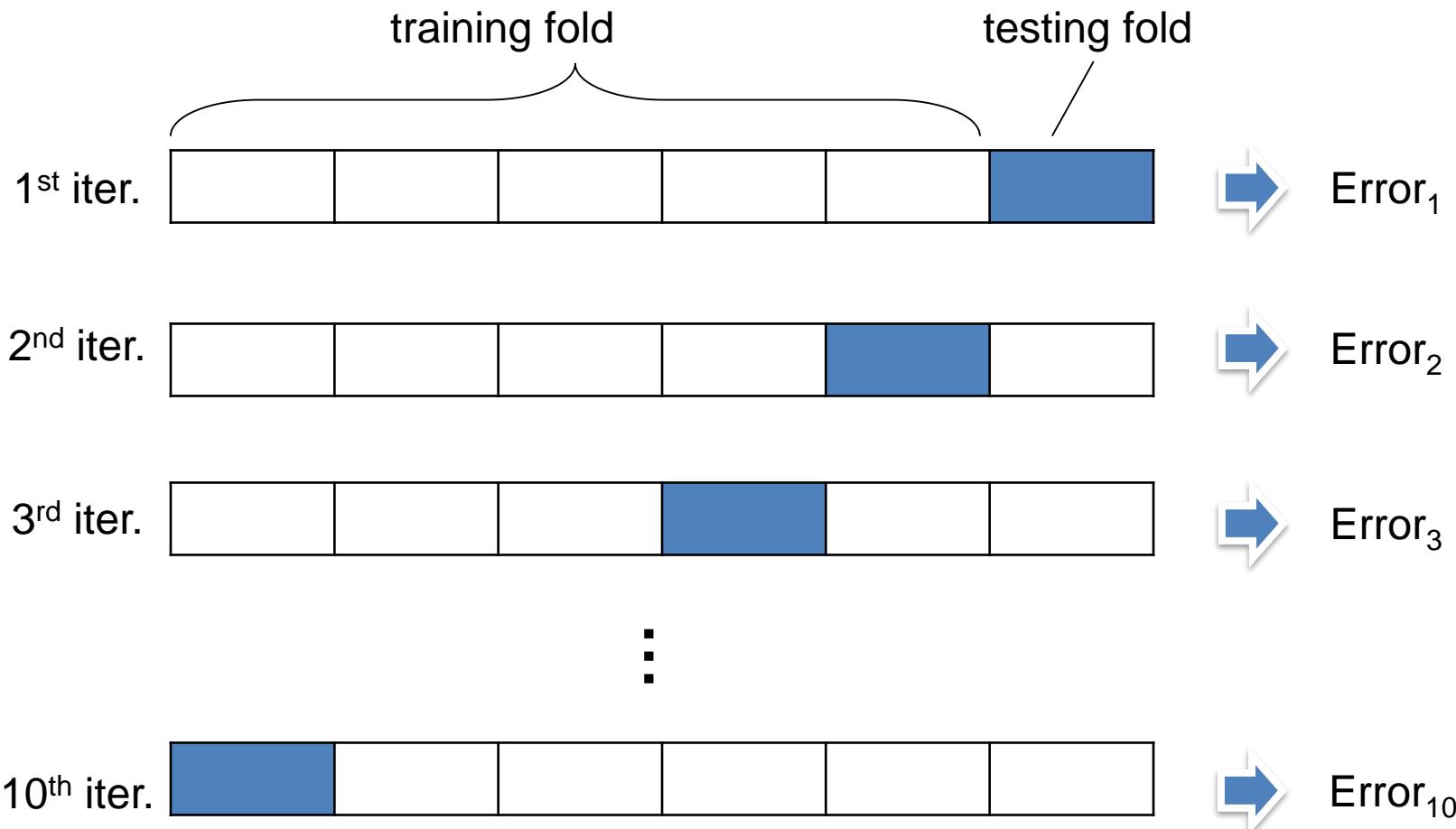
- When using several selection methods, calculate the selecting frequency for “**robust**” variable selection.

SVID	Stepwise	LASSO	Random Forest	Boosting	Voting
SVID_003	○	○	○	○	4
SVID_101		○	○	○	3
SVID_021		○	○	○	3
SVID_040	○		○	○	3
SVID_002	○	○		○	3
SVID_128	○		○		2
SVID_062	○	○			2
SVID_077		○		○	2
:	:				:

□ K-fold Cross Validation

- eg. 10-fold cross validation

$$\text{Minimize } \text{Error} = \frac{1}{10} \sum_{i=1}^{10} \text{Error}_i$$



□ Class Imbalance Solutions

- Random sampling deals with the issue. We focus on undersampling which samples a subset of the majority class.
- The main deficiency is that many majority class examples are ignored.
- Thus, we sample several subsets from the majority class (resampling).
- Others: oversampling, cost-sensitive, SMOTE, ensemble-based...

□ Example

- For Y label, 良品 vs. 不良品 = 1000 : 50
- Samples 50 良品 at a time for model training
- # of replications: 20 times
- Rank the variables by the “voting”
- Hint: 1:1 can be properly extended to 5:1

□ Pros and Cons

- Improve running time and storage problem
- Neglect potential useful information

SVID	Voting by Undersampling
SVID_003	19
SVID_101	18
SVID_021	18
SVID_040	18
SVID_002	17
SVID_128	17
SVID_062	17
SVID_077	17
:	:

□ Feature Selection by Pivot Analysis (樞紐分析)

SVID	良品_Avg	不良品_avg	Range/Avg.
SVID_003	0.9487	0.2583	1.1439
SVID_101	0.2078	0.5434	0.8934
SVID_021	0.4105	0.8377	0.6845
SVID_040	52.08	71.69	0.3170
SVID_002	2.7256	2.1025	0.2581
SVID_128	8.0523	8.5336	0.0580
SVID_062	0.9430	0.9747	0.0330
SVID_077	1569	1603	0.0216
:	:		:

$$\frac{|\text{良品}avg - \text{不良品}avg|}{Avg.}$$

Mach_ID	FAIL	TOTAL	RATIO
Photo_06	12	384	0.0313
Etch_10	3	295	0.0102
Photo_32	10	1011	0.0099
Photo_17	4	410	0.0098
PVD_02	1	123	0.0081
CVD_14	9	3456	0.0026
Diff_09	1	495	0.0020
Etch_07	5	2769	0.0018
:	:		:

$$RATIO = FAIL / TOTAL$$

只有1筆，沒有”再現性”!

□ Engineering Experience Validation (工程驗證)

● 重要工程參數的說明

- 變數名稱
- 物理意義
- 幾何圖形
- 調升與調降如何影響製程結果Y



BuzzHand (2014), 戚風蛋糕失敗的六大真相, http://www.buzzhand.com/post_118913.html

□ Engineering Experience Validation (工程驗證)

- 檢討篩選排序後的參數

- 每次可檢討30個參數為基礎，刪除物理上不具意義的參數
 - 可執行3~4回合

- 例如：

- 第一回合刪除不具意義的參數: SVID_051及SVID_064等共計 38個參數
 - 第二回合刪除不具意義的參數: SVID_034及SVID_082等共計 11個參數
 - 第三回合刪除不具意義的參數: SVID_115共計 1個參數

工程驗證刪除參數
邊際效應遞減
3~4回合

Lee and Tsai (2018)

□ 專案效益

- 找出過往經驗沒有發掘的未知重要因子
- 發展透過少數的因子建立線上量測機制
- 建立起製程因子的分析流程，以及工程參數驗證流程
- 建構知識庫(KM)後有助於工程人員快速學習與瞭解

□ 小議題

- How about "主成分分析PCA"進行維度縮減?
- It depends, 應根據個案討論。基本上因參數的物理特性，通常用轉換後的主成分，需要"命名"且物理意義"不易解釋"。然而，主成分用於預測模型時有時效果不錯。

愈重要的工程參數(變數) 統計方法愈容易找出來?

How about 某一個欄位所有的觀測值都一樣...

資料收集要多少量？

時間區間？

某個小時段 → troubleshooting based on the same product-mix
模型可能可以tune的不錯、找出關鍵參數

廣大時間區間 → robust forecasting based on the different product-mix
模型MSE不會太好看、換 x 或recipe來預測 y 時，主要不是預測”值”(!?)，
而是關注於”趨勢”(上升 or 下降)為主
也就是獨立變數的係數 (coefficient of independent variable)

案例三：品質/良率/精度預測

□ 製程預測

- 透過製程的處方設定 (recipe)來進行反應變數(Y)的推論，eg. 長寬預測、厚度預測、精度預測、虛擬量測...

□ 預測？

- 良品/不良品
- 膜厚(thickness)、線寬(Critical Dimension, CD)、覆蓋(Overlay)、缺陷數(defects)、缺陷面積大小(area)、缺陷位置層(layer), 良率(yield)等...

□ 方法

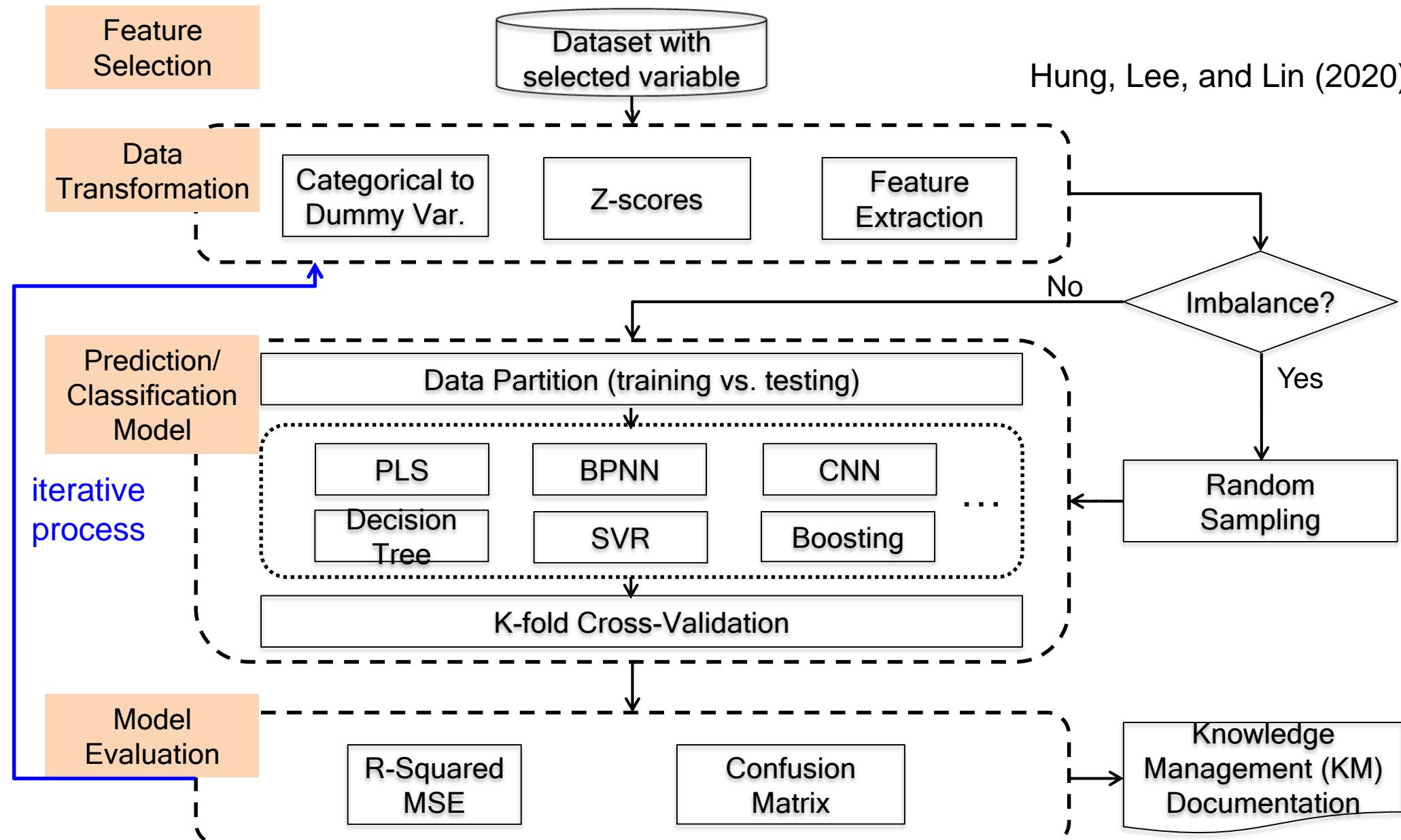
- 推論 → 就是進行「因果關係」的釐清

□ Benefits

- 協助機台校正或產品開發
- 降低抽樣頻率
- 發展預測保養 (predictive maintenance)

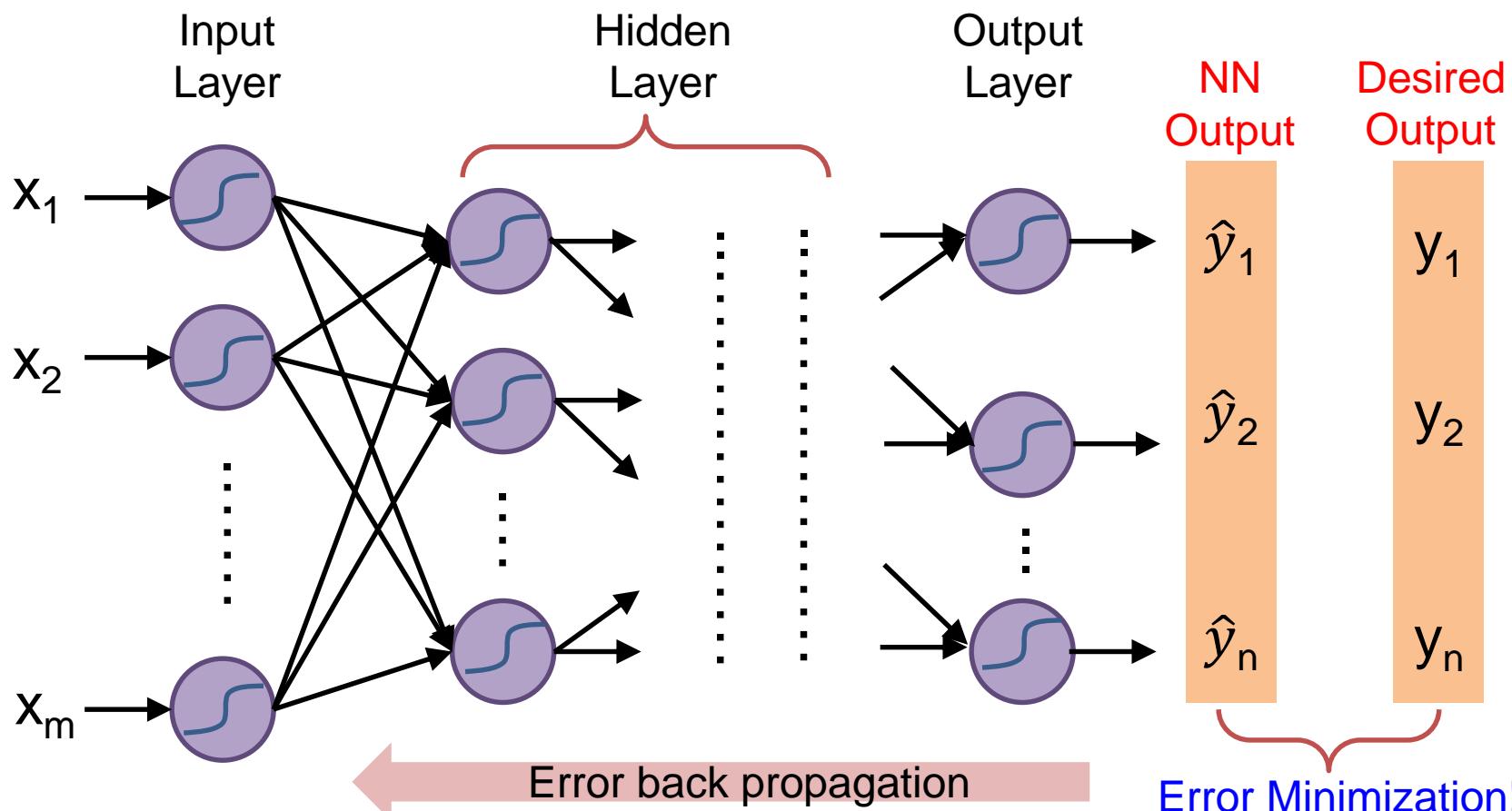
案例3: Yield Prediction (良率/精度預測)

Hung, Lee, and Lin (2020)



□ Backpropagation Network (BPNN) (Werbos, 1974; Rumelhart & McClelland, 1985)

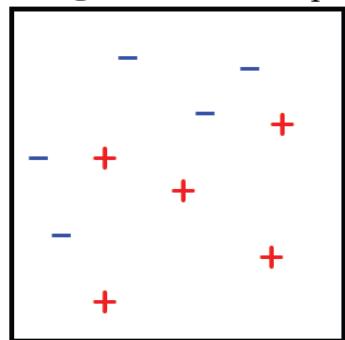
- Multi-Layer Feedforward Network + Error Back Propagation (EBP)
- Minimize the error between NN output and desired output by “(stochastic) gradient descent”. (a large number of hidden layers → Deep Learning)



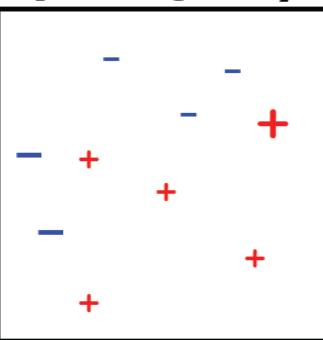
案例3: Yield Prediction (良率/精度預測)

□ Boosting (eg. Gradient Boosting Machine, GBM)

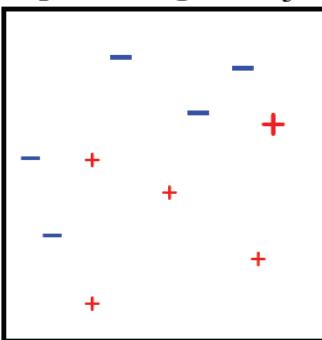
Original data set, D_1



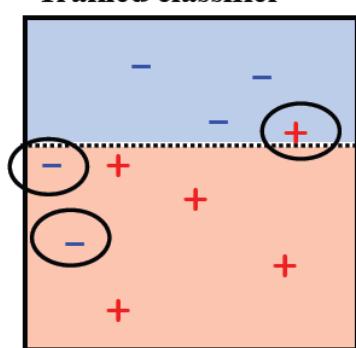
Update weights, D_2



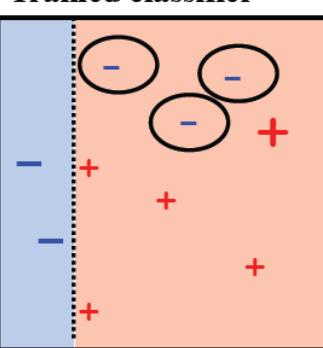
Update weights, D_3



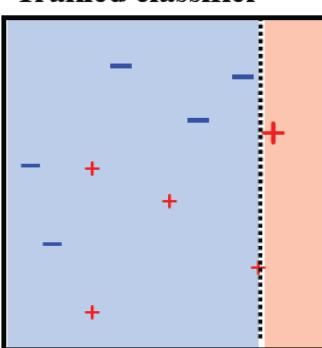
Trained classifier



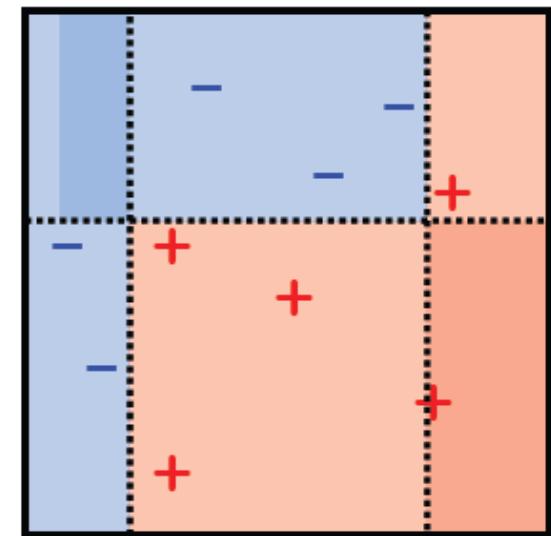
Trained classifier



Trained classifier



Combined classifier



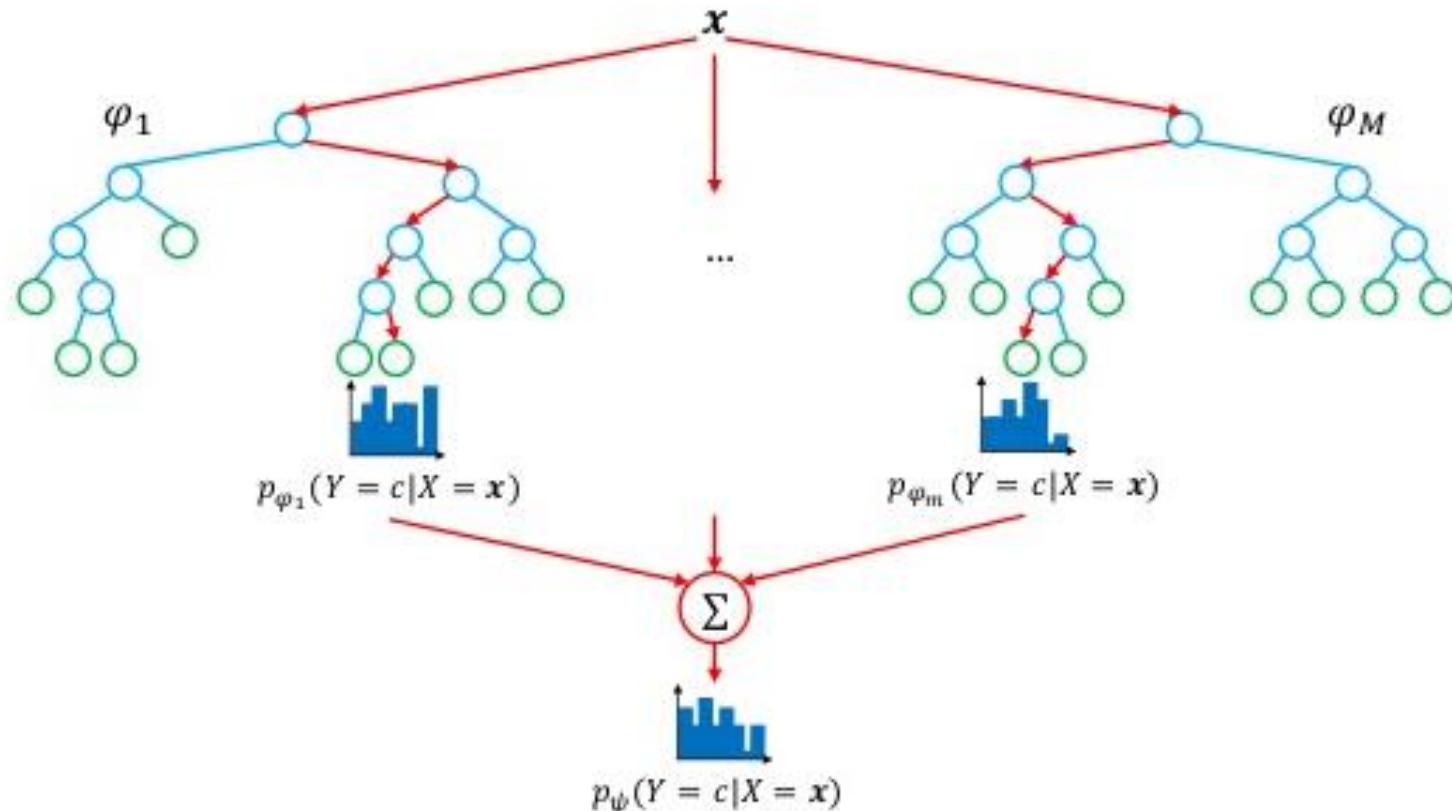
Weight each classifier and combine them:

$$.33 * \begin{array}{|c|} \hline \text{blue} \\ \hline \text{orange} \\ \hline \end{array} + .57 * \begin{array}{|c|} \hline \text{blue} \\ \hline \text{orange} \\ \hline \end{array} + .42 * \begin{array}{|c|} \hline \text{blue} \\ \hline \text{orange} \\ \hline \end{array} \geq 0$$

1-node decision trees
“decision stumps”
very simple classifiers

Alexander Ihler (2012). <http://sli.ics.uci.edu/Classes/2012F-273a?action=download&upname=10-ensembles.pdf>

Random forests



Randomization

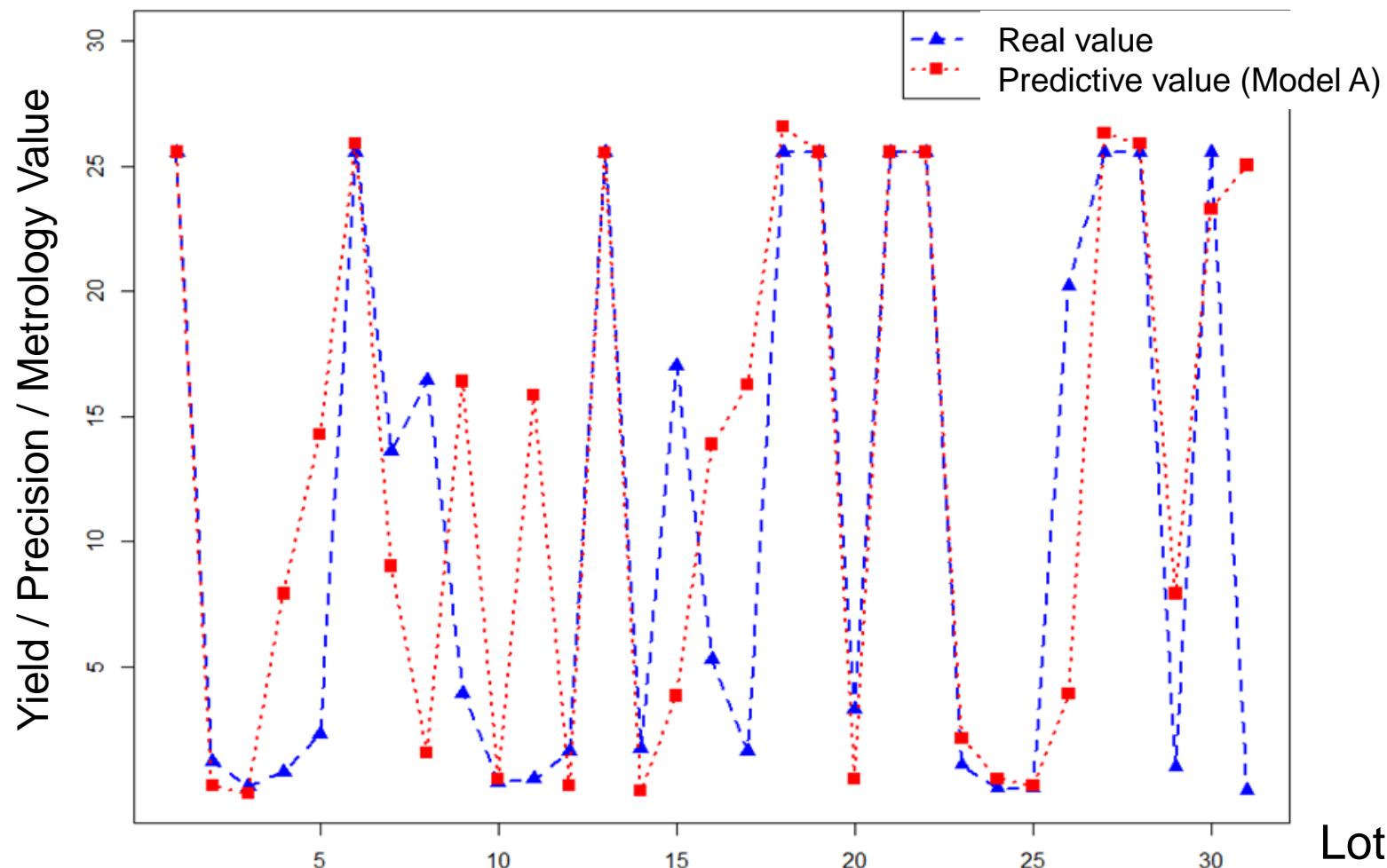
- Bootstrap samples
 - Random selection of $K \leq p$ split variables
 - Random selection of the threshold
- $\left. \begin{array}{c} \text{Bootstrap samples} \\ \text{Random selection of } K \leq p \text{ split variables} \\ \text{Random selection of the threshold} \end{array} \right\}$ Random Forests

$\left. \begin{array}{c} \text{Bootstrap samples} \\ \text{Random selection of } K \leq p \text{ split variables} \\ \text{Random selection of the threshold} \end{array} \right\}$ Extra-Trees

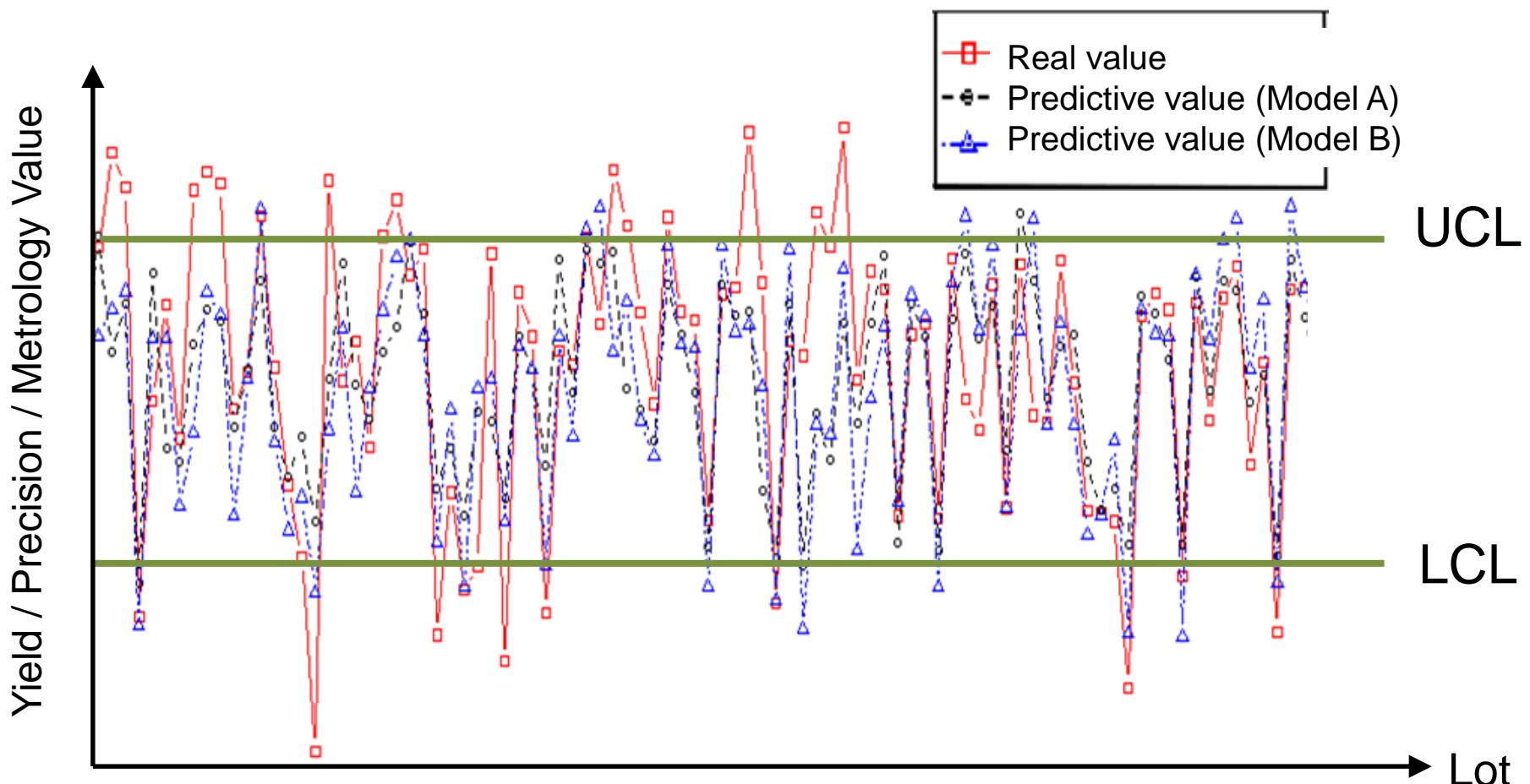
<https://www.kdnuggets.com/2017/10/random-forests-explained.html>

□ 預測的結果

- Testing Dataset with R^2 , Adjusted R^2 , Mean Squared Error(MSE), etc.



□ Statistical Control Chart for Classification (分類)



- 預測一個”值”，本身沒有對錯的問題。
- 預測一個”類別”，就有對錯的問題。

□ 分類結果- Confusion Matrix

- 以測試組資料的結果來客觀評估較佳的分類模型
- 假設是二元分類模型，也就是只有兩種類別的模型，例如：
— (良品／不良品) (陽性／陰性) (有病／沒病) (發生／不發生) 等。

		預測類別	
		類別1 (不良品)	類別2 (良品)
實際 類別	類別1 (不良品)	TP (true positive)	FN (false negative) (Type II error) (miss)
	類別2 (良品)	FP (false positive) (Type I error) (false alarm)	TN (true negative)

□ 根據分類結果，可計算出正確率 $Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$

□ 敏感度 (Sensitivity, Recall)

- 為類別1(不良品)當中，被正確預測的比率

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

- Miss (漏報) Rate = 1 – Sensitivity

		預測	
		類別1	類別2
實際	類別1	TP	FN
	類別2	FP	TN

□ 明確度 (Specificity)

- 為類別2(良品)當中，被正確預測的比率

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

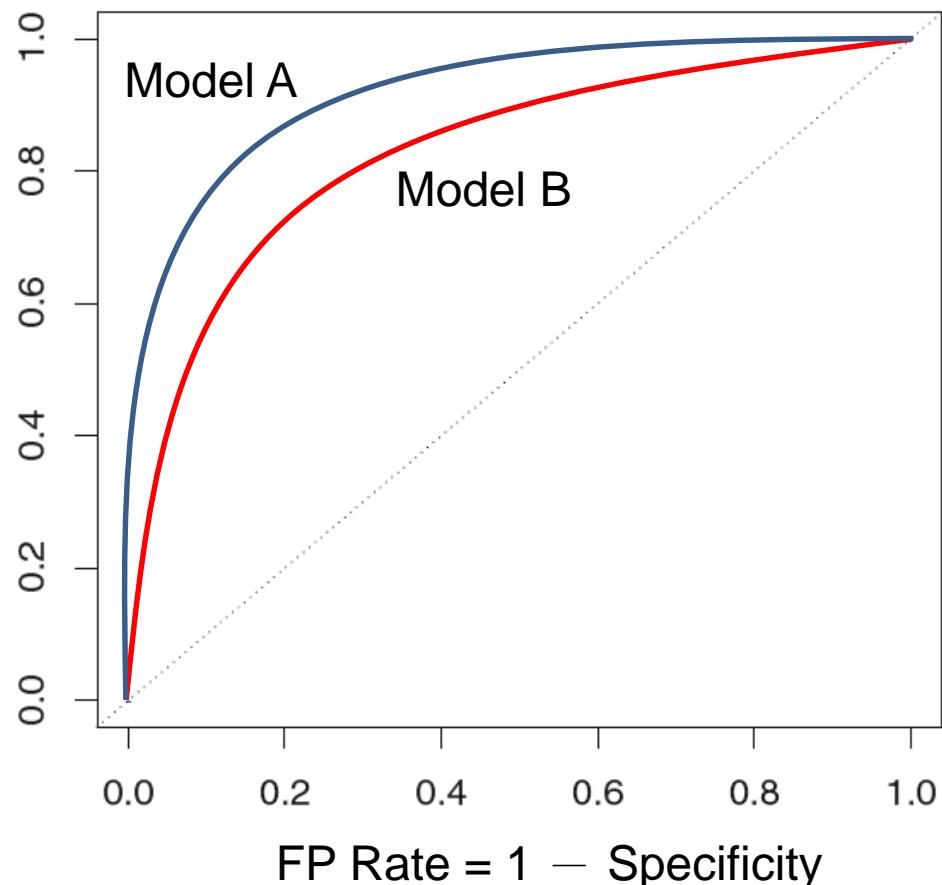
- False Alarm (假警報) Rate = 1 – Specificity

□ ROC曲線

- Receiver Operating Characteristic curve
- 可作為衡量不同 FP Rate 下 TP Rate 的變化
- 好的模型 TP rate 越大越好且 FP rate 越小越好
- TP Rate 與 FP Rate 呈現一種 Trade-off (權衡取捨)

描述當資料屬於類別1(不良品)時，
被正確判斷的機率

TP Rate
= Sensitivity



□ 一般來說，右下方的面積愈大，該模型的分類效果愈佳！

當資料不屬於類別1(不良品)時，
被誤判的機率 (Type I error)

□ 精度/良率預測

- Model performance (128個觀測值)

Model A		預測	
		Bad	Good
實際	Bad	61	7
	Good	29	31

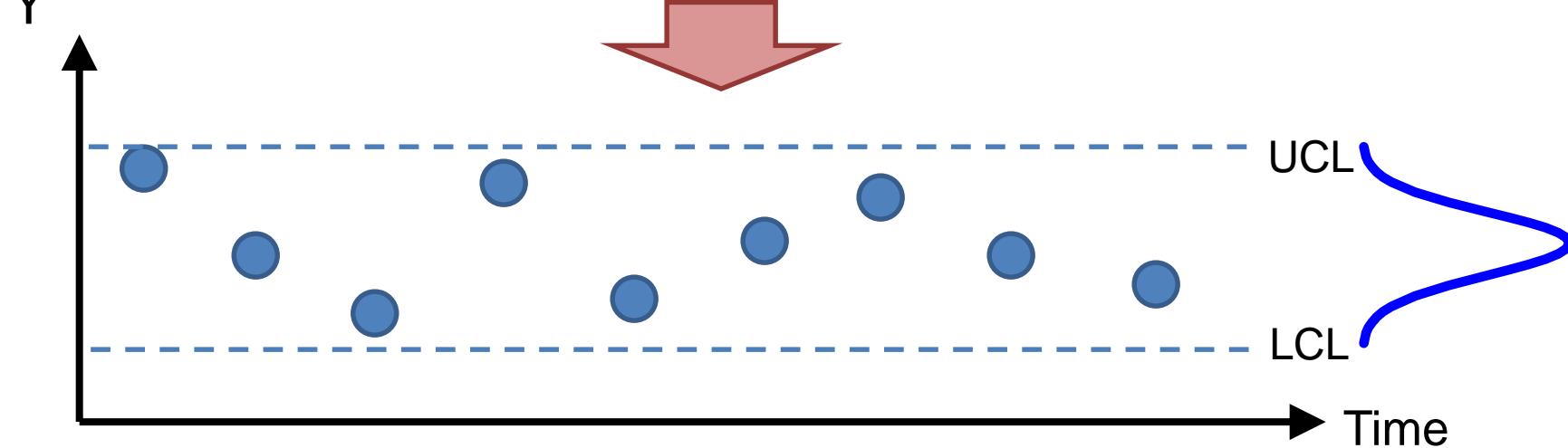
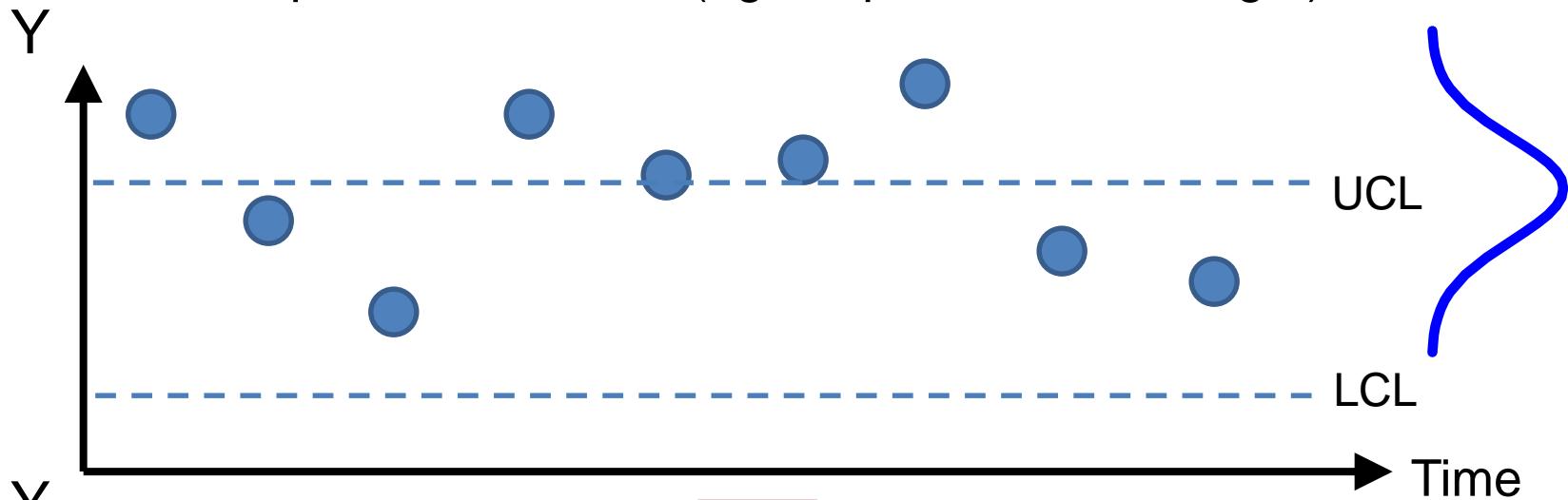
Model B		預測	
		Bad	Good
實際	Bad	47	21
	Good	7	53

	Testing			
	Accuracy	TP Rate	TN Rate	AUC
Model A	71.9%	89.7%	51.7%	70.2%
Model B	78.1%	69.1%	88.3%	78.9%

AUC: Area under the Curve of ROC

□ Process Adjustment (製程控管與調整)

- Manipulate the compensating variables of a process to achieve the **desired** process behavior (eg. output close to a target)



案例四：預測保養

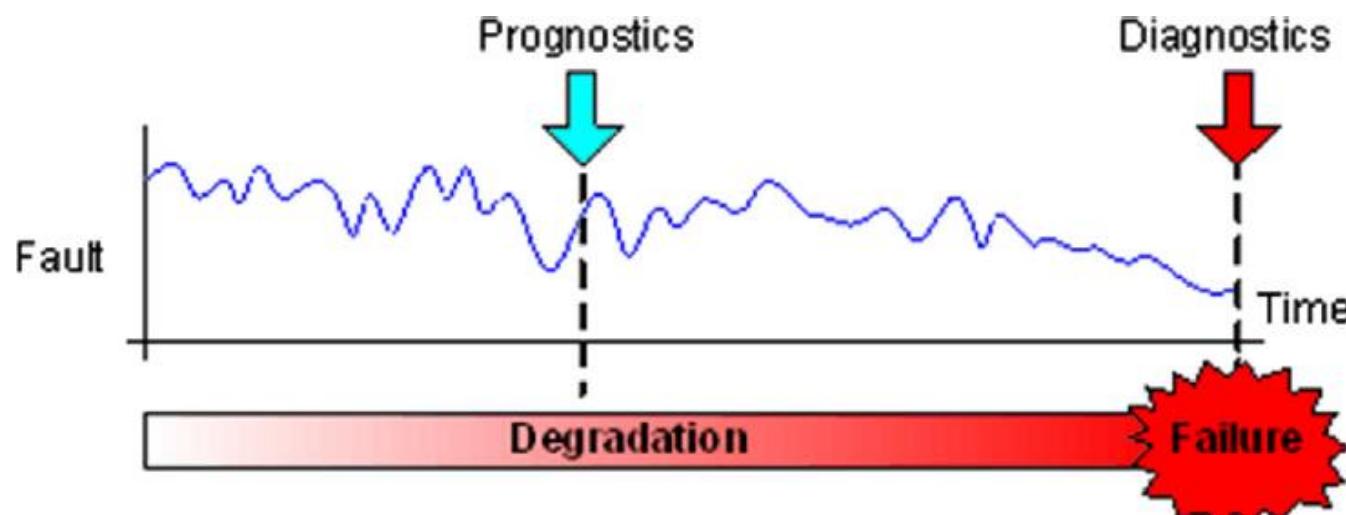
案例四：預測保養(PdM)

□ 因為工業4.0...

- 機台預測保養(Predictive Maintenance, PdM)

□ Prognostics

- (1) How is the machine operating now? (Performance assessment)
- (2) When will the machine break down? (Remaining useful life)
- (3) What will be the primary faults that cause downtime?
- (4) Why does the fault occur?



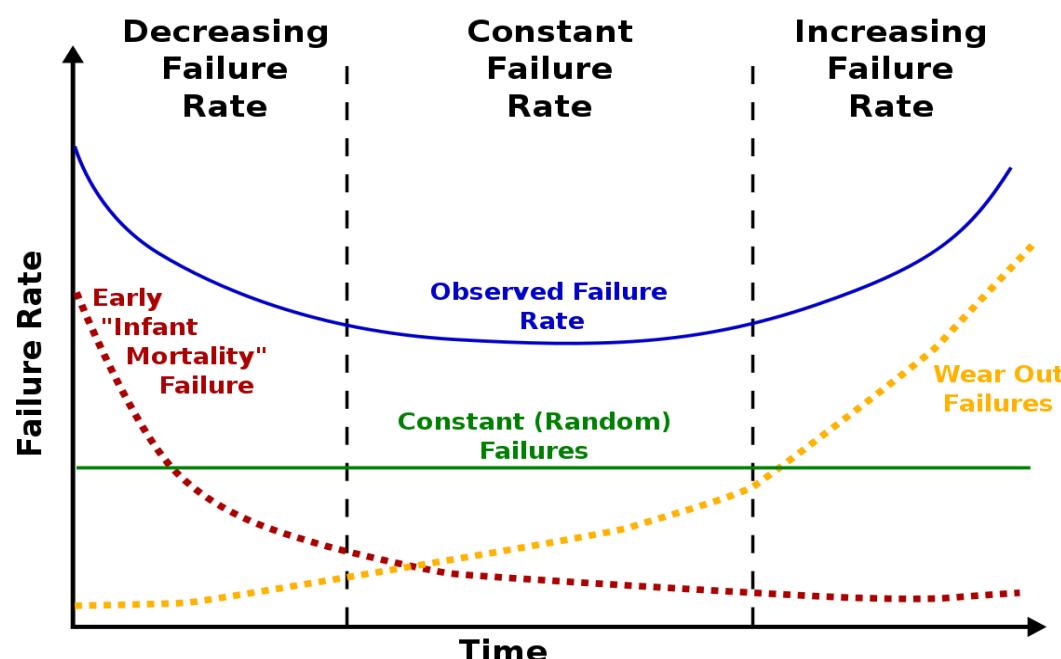
(Lee et al., 2014)

Lee, et al. (2014). Prognostics and health management design for rotary machinery systems- Reviews, methodology and applications. Mechanical Systems and Signal Processing, 42, 314-334.

案例四：預測保養(PdM)

□ 維修保養

- **Reactive:** Service or replace equipment after it fails
- **Preventive:** Service or replace equipment according to the manufacturer's suggested schedule, or the amount of time it has been in service, or based on operational observations
- **Condition-based:** Service or replace equipment based on monitoring performed to assess its current condition
- **Predictive maintenance (PdM):** Maintaining components or assets according to fact-based expectations for when they will fail or require service, i.e., **remaining useful life estimation!**



□ 機台預測保養(Predictive Maintenance, PdM)

- 雙週、月保、季保、年保check list..
 - 都是手寫表單，要key-in到系統裡...
 - 字跡潦草 → 「其他」
- 年保資料最完整，但...
 - 資料量少(因為一年一次)
 - 一年一次，解析度差，看不到一年內機台狀態的改變
 - 只有保養後的資料，沒有保養前的狀態...
- 雙週、月保資料相對多，但零散難以拼湊...
 - 這台機台拆一個零件下來裝到另一台上
 - 憑經驗調參數，不一定要恢復到”最佳狀態”，先”可用”救火再說...
 - 有些保養後70分...有些保養後50分...
 - 到底換了哪個零組件？不清楚 → 帳料不符
 - 就像修電腦一樣，拆了一堆power、主機板、RAM
 - 反正RAM一堆，隨便插，插上去能用再說，管他從哪裡來...

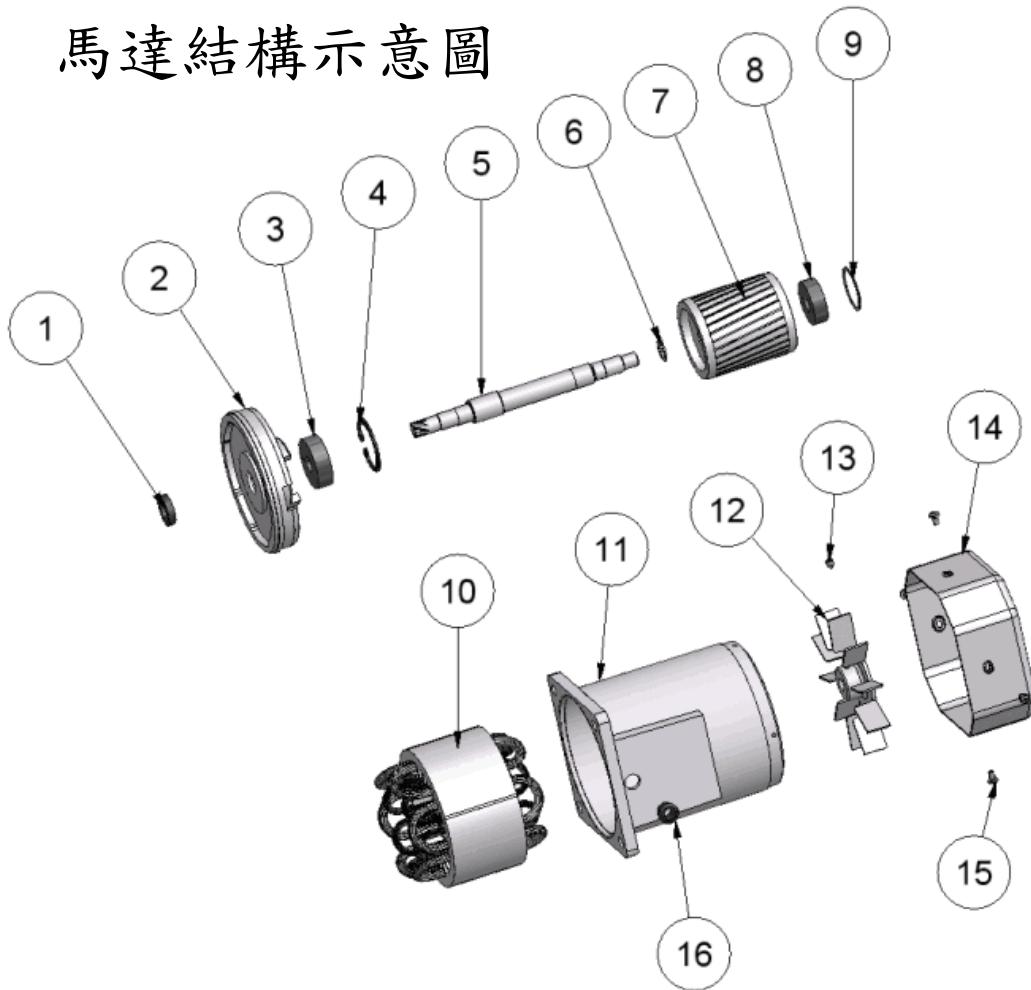
當你瞭解”現場”後，
你覺得這案子還要不要GO?

MephaGO (沒法GO)!?

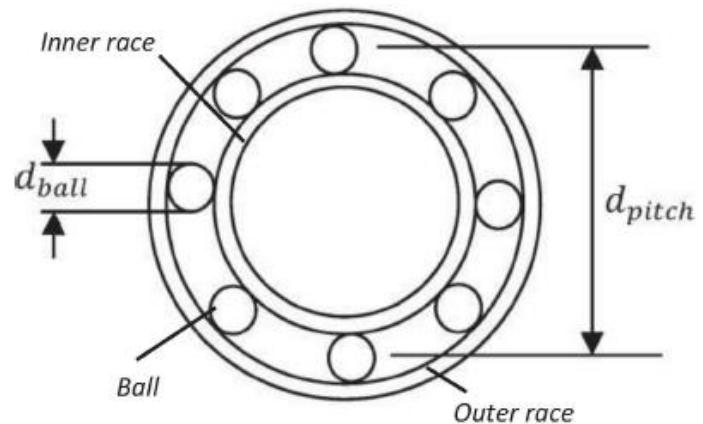
- 如果還是想GO，想試試看的話...
- 可能建議有二
 - 從解析PdM資料面下手(通常牽涉到流程與制度)...
 - PdM需要哪些資料？為什麼資料不完整呢？
 - 目前哪些人為寫？哪些系統有紀錄？
 - 表單要從手寫變成資訊系統(掃條碼)？是否要裝新的sensor？
 - 未來還是會用到這些資料...
 - 缺點：問題很簡單、績效端不上檯面
 - 從分析PdM相關的題目...(且資料要"相對"完整...)
 - 有沒有什麼"替代資料"也"間接"地量測機台狀態？
 - 預測保養 → 關鍵零組件 → 關鍵零組件的可靠度分析？庫存管理？
 - 現有"預防保養"制度的檢討 eg. 年保改季保？動態保養？
 - 除了時間，還可考慮機台使用耗材(eg. 刀具、線材)的狀況來推估零件磨損的速度

□ 馬達軸承壽命預測 Remaining Useful Life (RUL) Prediction

馬達結構示意圖



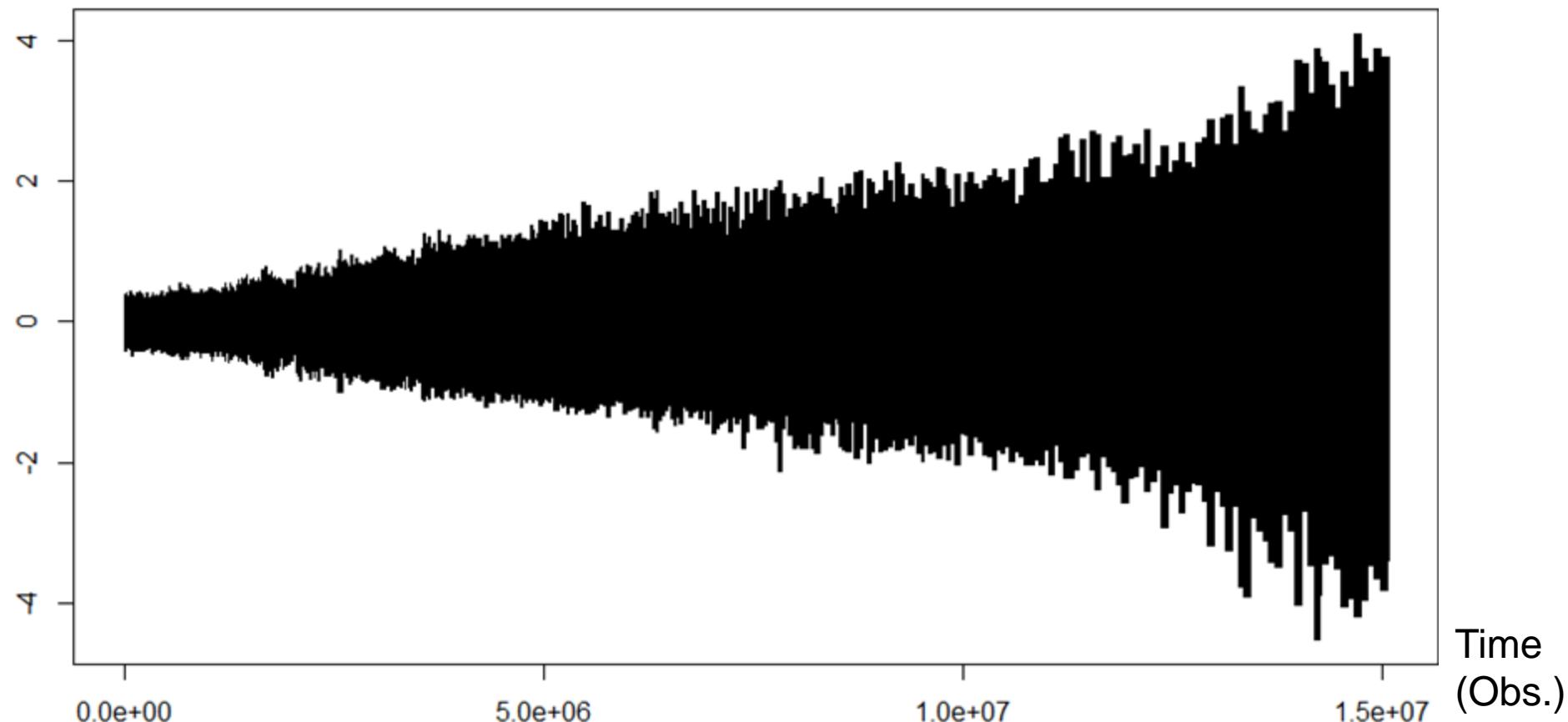
No	名稱	No	名稱
1	油封	9	波型華司
2	馬達前蓋	10	定子線圈
3	滾珠軸承	11	馬達機殼
4	R扣環	12	風扇
5	馬達齒軸	13	固定螺絲
6	S扣環	14	風罩
7	轉子	15	十字頭螺絲
8	滾珠軸承	16	護線扣環



http://www.li-ming.com.tw/index.php?option=com_content&view=article&id=225&Itemid=351&lang=zh-TW

□ Vibration Dataset

Amplitude

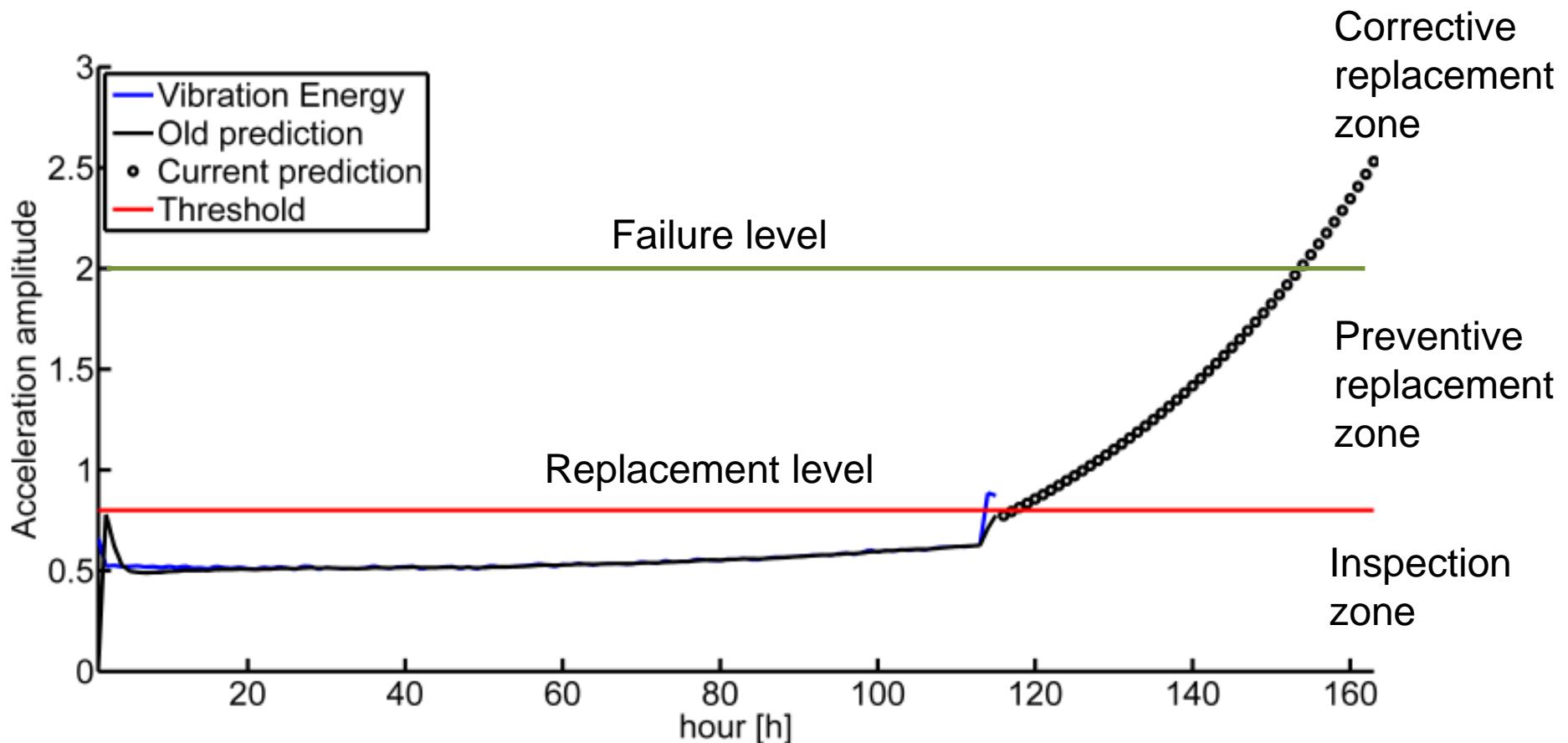


Lee et al. (2019)

Lee, C.-Y., T.-S. Huang, M.-K. Liu, and C.-Y. Lan. 2019. Data Science for Vibration Heteroscedasticity and Predictive Maintenance of Rotary Bearings. *Energies.*, 12 (5), 801.

□ Prediction of the Deterioration

- Vibration acceleration signal



revised from Grall et al. (2002) and Rocchi et al. (2014)

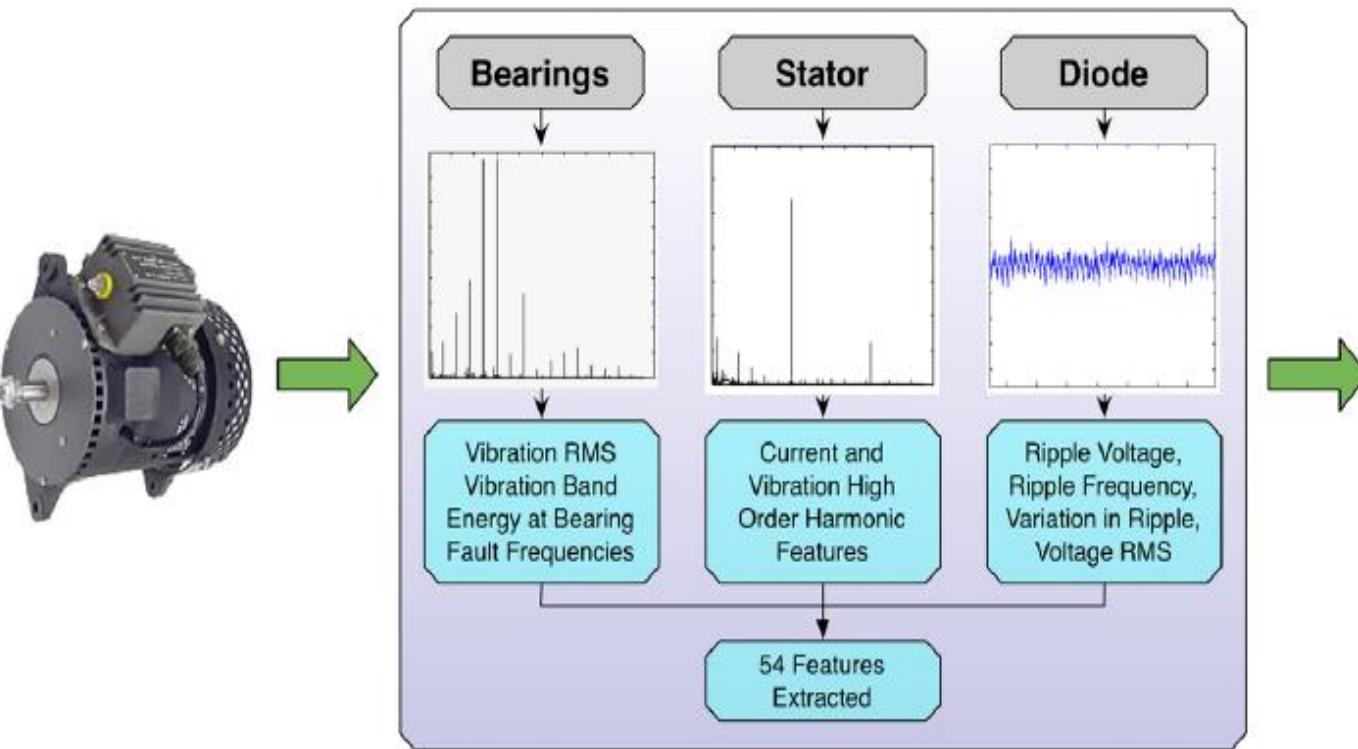
案例四：預測保養(PdM)

Equipment

Sensor Data Collection

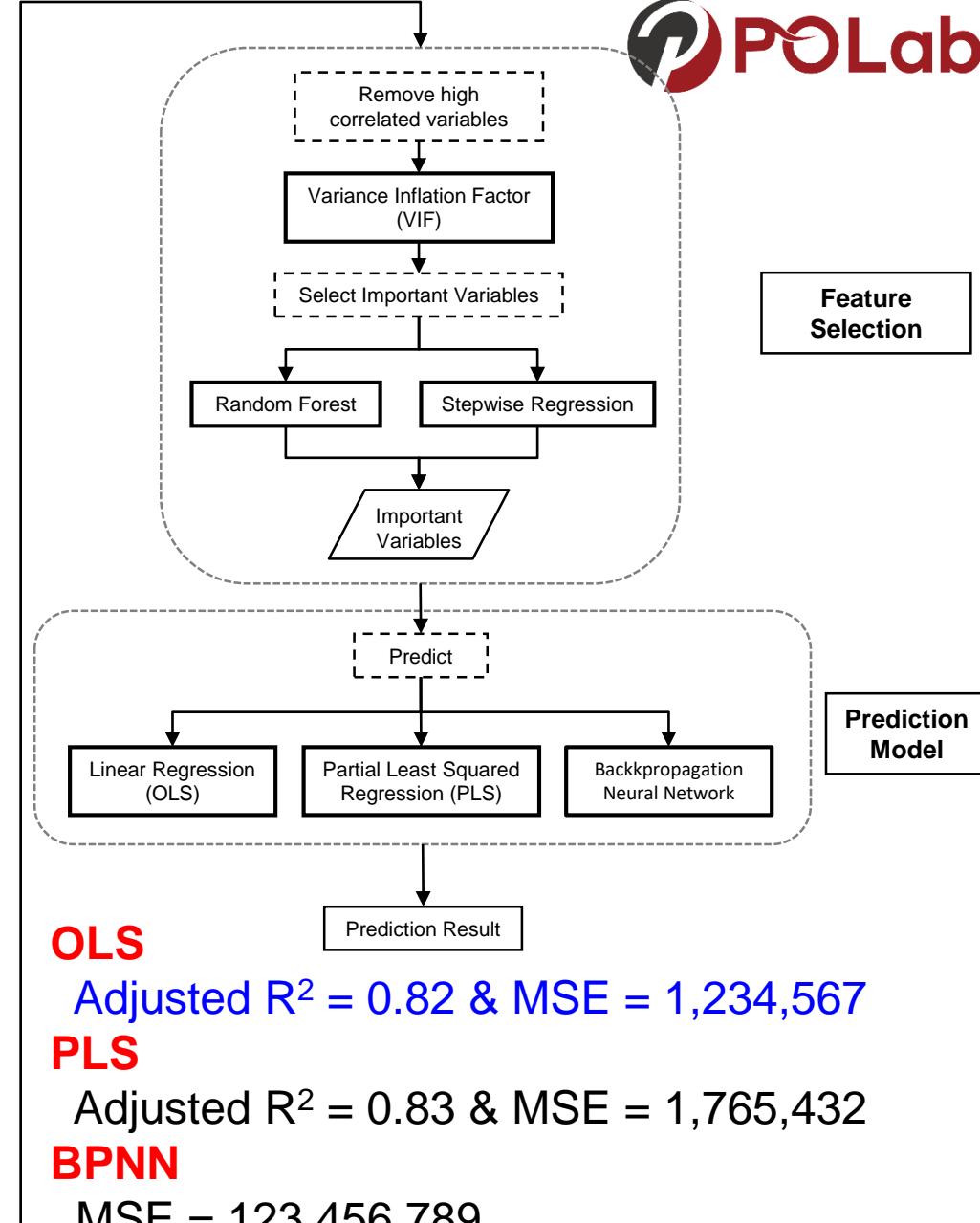
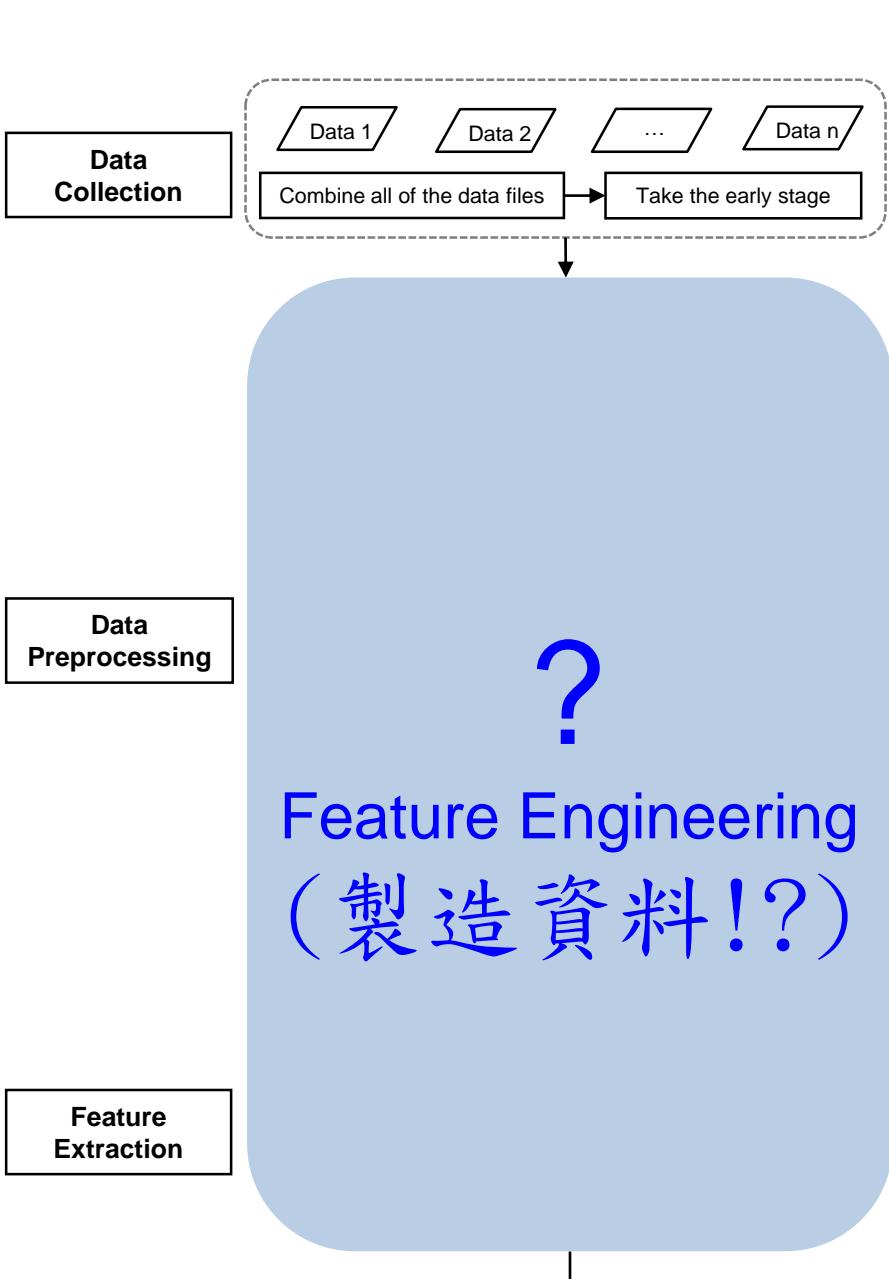
Modelling

Visualization



(Lee et al., 2014)

Lee, et al. (2014). Prognostics and health management design for rotary machinery systems- Reviews, methodology and applications. Mechanical Systems and Signal Processing, 42, 314-334.



Lee, C.-Y., T.-S. Huang, M.-K. Liu, and C.-Y. Lan. 2019. Data Science for Vibration Heteroscedasticity and Predictive Maintenance of Rotary Bearings. *Energies.*, 12 (5), 801.

案例四：預測保養(PdM)

□ Feature Engineering (Time Domain, Frequency Domain, Time-frequency Domain)

- Root Mean Square (RMS), Mean, Variance, Skewness, Kurtosis, and Crest Factor/Max/Min, ...
- Time synchronous averaged (TSA) signal, **sliding window**, residual signal (RES), difference signal (DIF), ...
- Time series decomposition, **Hilbert vibration decomposition**, empirical mode decomposition, intrinsic mode functions (IMF), ...
- Regression (intercept, slope), **exponential degradation $\theta \exp(\beta t)$** (amplitude θ , slope β), autoregression, ...

Component	Issue & failure	Characteristic	Common measures	Common features	Common algorithms used
Bearing	Outer-race, inner-race, roller, and cage failures	Raw data does not contain insightful information; low amplitude; high noise	Vibration, oil debris, acoustic emission	Vibration characteristic frequency, time domain statistical characteristics, metallic debris shape, size, quantity, sharp pulses and rate of development of stress-waves propagation	Fourier Transform (FT) [18,19], Short Time Frequency Transform (STFT) [20], Wavelet Transform (WT) [21], Empirical Mode Decomposition (EMD) [22], Bispectrum [23], Autoregression (AR) Frequency Spectra [24], Hilbert Spectrum [25], Instantaneous Power Spectrum [26], Hilbert–Huang Transform (HHT) [27], Neural

Lee, et al. (2014). Prognostics and health management design for rotary machinery systems- Reviews, methodology and applications. Mechanical Systems and Signal Processing, 42, 314-334.

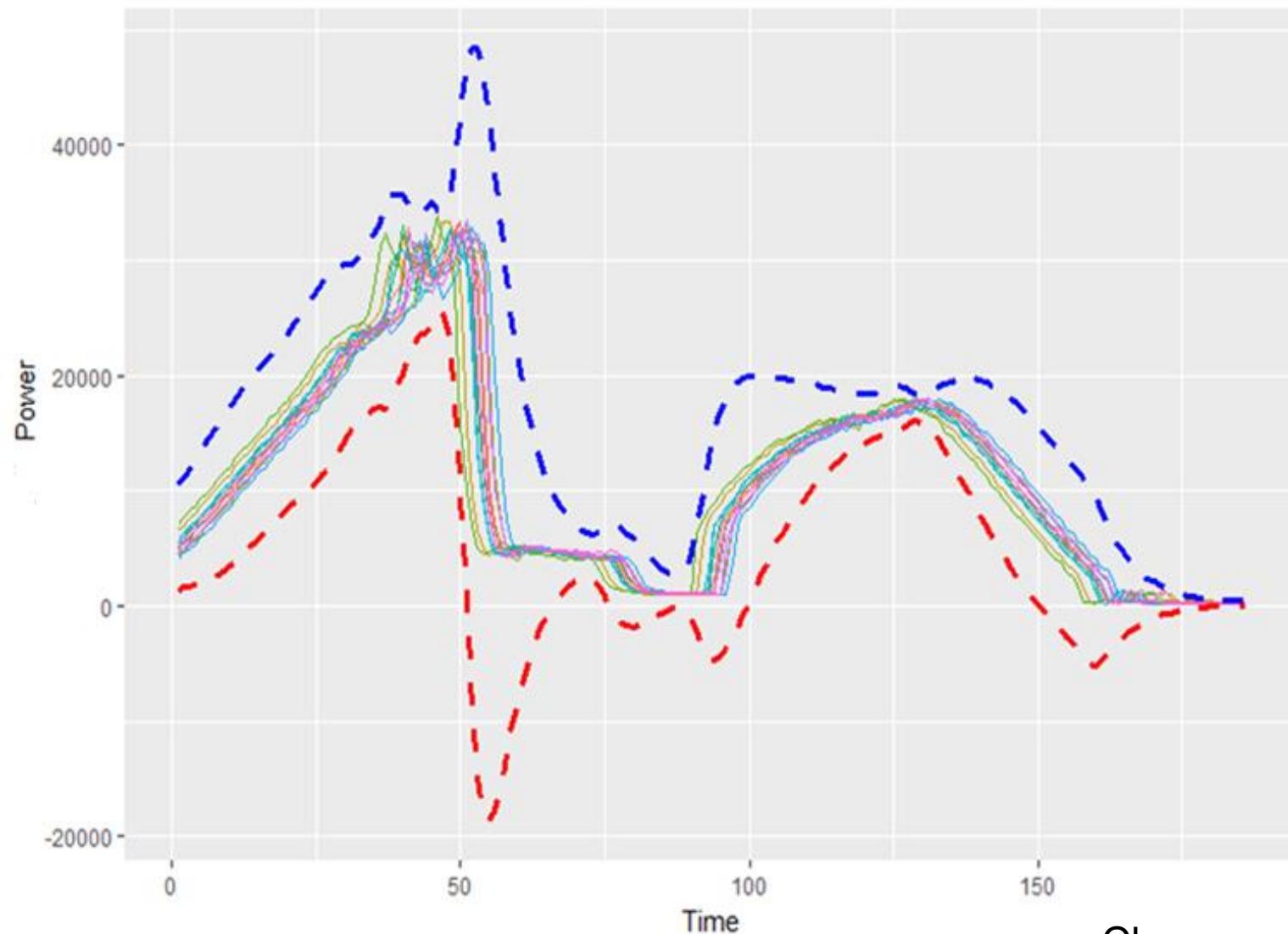
給我們的啟發是...

如果能夠製造出來好的Feature...

預測可簡單易解釋且效果好

事實上，只有少數幾個sensors或features才真正提供有用的信號...

□ MCEWMA



Chang and Lee (2018)

□ 資料科學模型建構的議題還有...

- 預測要預測什麼？
- 如果樣本數不夠怎麼辦？
- 如何決定資料收集的時間區間與數量？
- 如果資料有些偏差，該如何發展穩健(robust)的模型？
- Training dataset與testing dataset如何切割與決定？
- 抽樣所造成的資料不完整與不平衡如何處理？
- 產品組合持續改變該如何分析？
- 前後工作站間的交互作用如何衡量？
- 預測模型何時需要重新修正或重建(re-train)？
- 預測模型可以實際上線的標準如何訂定？
- 模型挑選關鍵參數與工程經驗不一致如何處理？
- 如何確認該關鍵參數是潛在因子？
-

□ 然而，製造現場的議題還有...

- 在製品WIP太多如何處理？
- 批量大小該如何決定？
- 機台轉速或throughput的調整？
- 生產排程與派工的法則？
- 操作員或員工排班的最佳化？
- 產能彈性如何建構以應變需求波動？
- 每年度的生產力target該如何訂定與提升？
- 產品組合product-mix如何決定？
- 如何減少浪費？看板式管理如何進行？
- 專案管理如何Continuous Delivery？敏捷開發？
- 抽樣檢驗計畫該如何動態調整？
- 多種零組件與原物料，庫存管理如何訂定？
-

資料科學可否協助
解決這些議題？

創業者的世界裡只有兩樣東西

有待解決的問題、正在嘗試的方法

這兩個東西合起來就是一個詞

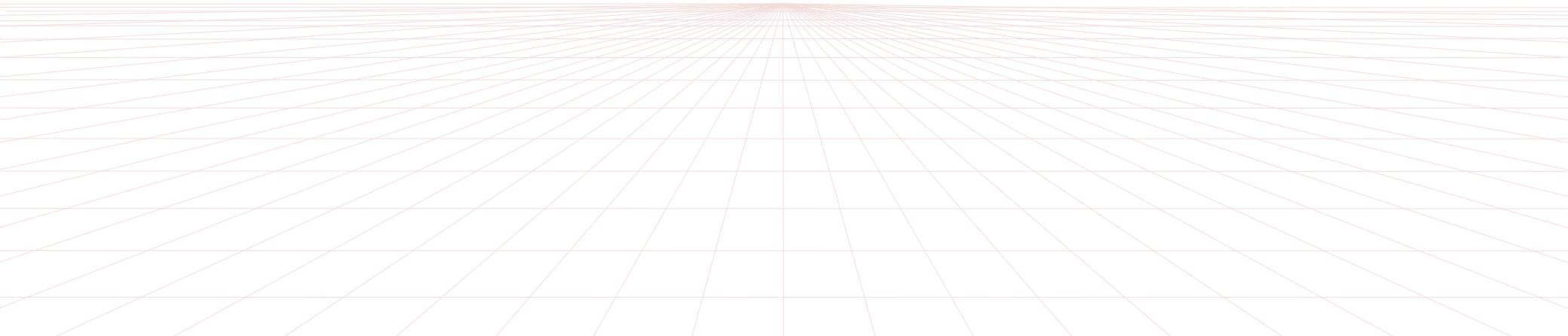
“機會”!!



罗振宇 (Dec. 31, 2016)，罗辑思维 "时间的朋友2016" 跨年演讲

自動化生產排程與演算法

Automated Production Scheduling



事實上...

很多時候，我們..
也不知道自己要做什麼題目？

現場改善

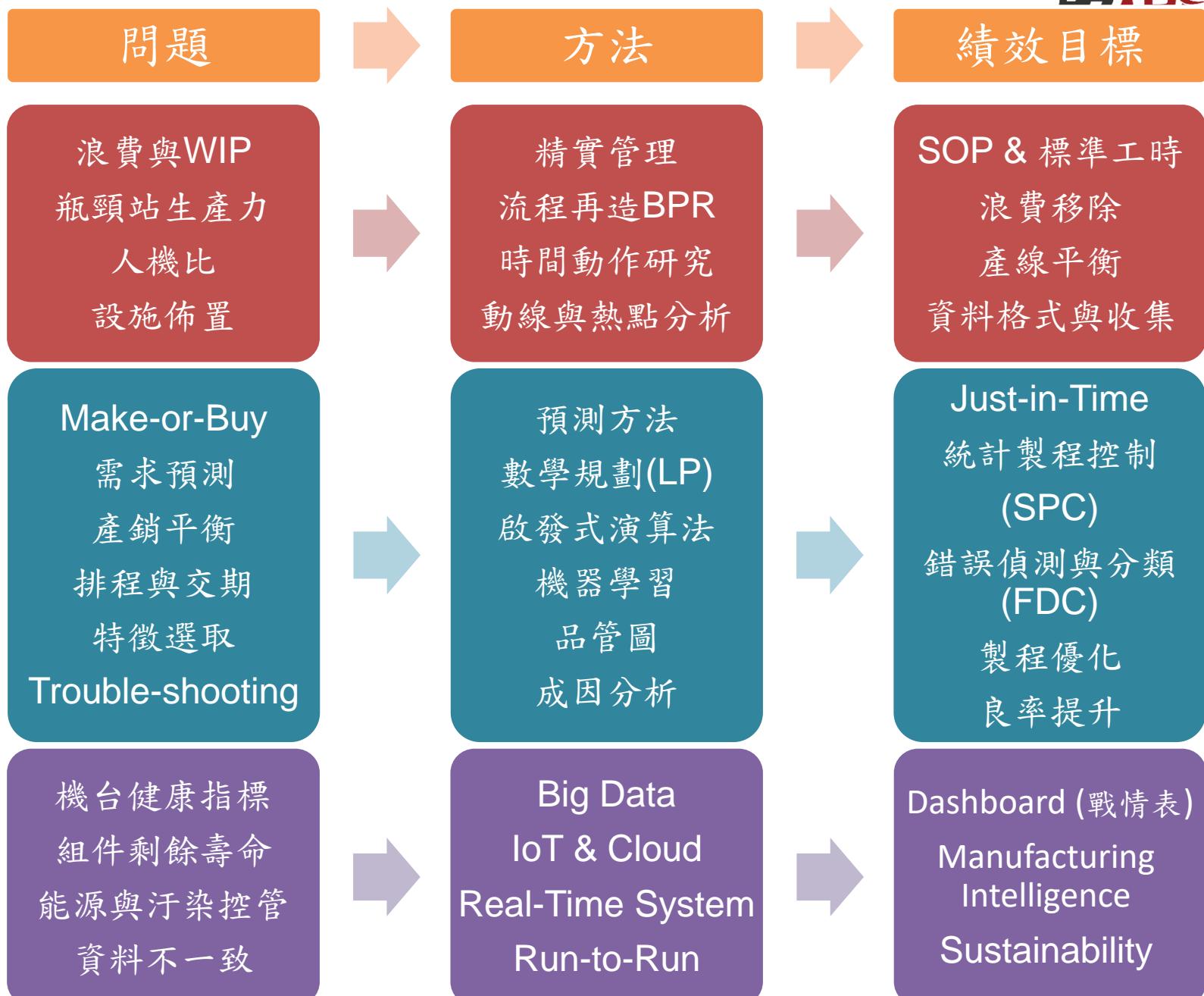


資料科學



系統整合

Lee (2019)



那...那如果我們什麼資料都沒有，
或資料還沒有標準化地著手收集，
那我們要怎麼開始呢？

- 資料科學並非要有一堆數學模型或一大堆資料...
- 一開始可以先嘗試，找到問題
- 為了改善，針對這問題無論如何都要收集到資料!
- 方法
 - 資料庫 (但資料可能時間過久、現況已與當時不同、久未校正或有誤差)
 - 詢問專家 (但可能已經離開專案小組、或轉調不同部門組織)
 - 現場觀察
- 現場觀察是一非常強且客觀的方法 → 眼見為憑

□ 於201X年，企業參訪，觀察工廠製造現場與流程

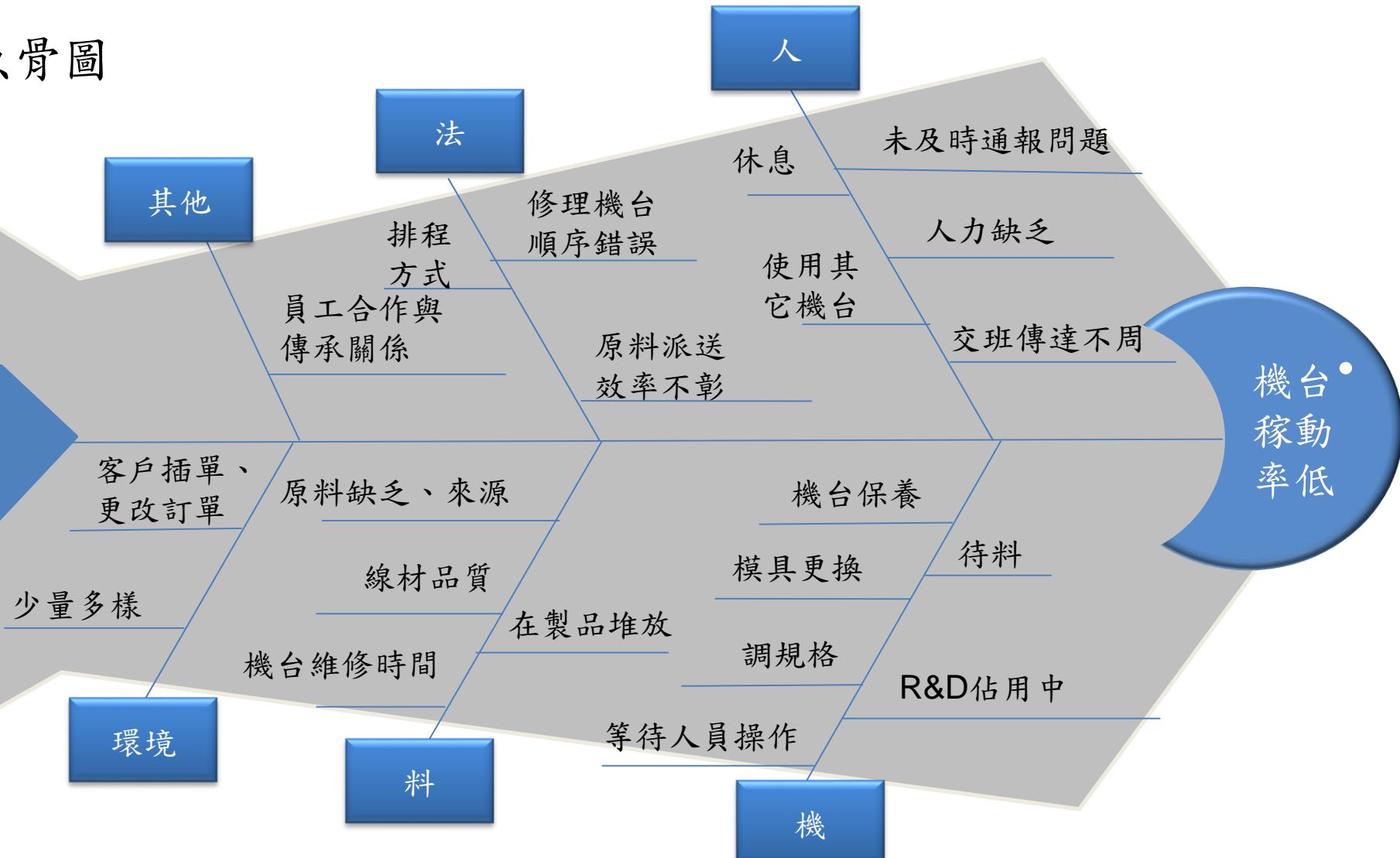
□ 某廠區稼動率 $116/200 = 58\%$

□ 初步參訪的疑惑

- 為何機台(資金)閒置？
- 是否是景氣不佳、訂單需求不足？
- 現場人員數目不足操作機台？
- 過多機台換模具、調整、維修保養中？
- 老舊機台(閒置)設備過多？

□ 工廠稼動率的合理性驗證

□ 魚骨圖



陳建旭、鄭佩怡、李家岩 (2015)

□ 人機程序圖

- 紀錄人和機台之間的動作
- 案例：
 - 時間：2017/02/11, 2pm 開始觀察3小時 (此處資料為模擬產生，非實際數據)
 - 對象：製造部A員操作3台機台
 - 記錄人的操作行為與機台做動的時間
 - 計算機台稼動率 = (機台作業生產時間) / (總觀察時間)
$$\text{eg. } (56+160+138 \text{ mins})/(180+180+180 \text{ mins}) \\ = 65.6\%$$

□ 人材



數據)

□ 工廠稼動率條件

- 強假設：人員動 → 車台動 → 物料動
- 車台的稼動率基本上會低於人員的工作效率

□ 觀察操作員的時間分解

- 人員工作效率：80% ~ 85%
- 因此，機台稼動率之上限即為人員工作效率(80% ~ 85%)
- $200 * 80\% = 160$ (台) 到 $200 * 85\% = 170$ (台) 之間為理想機台運作數

□ 小結

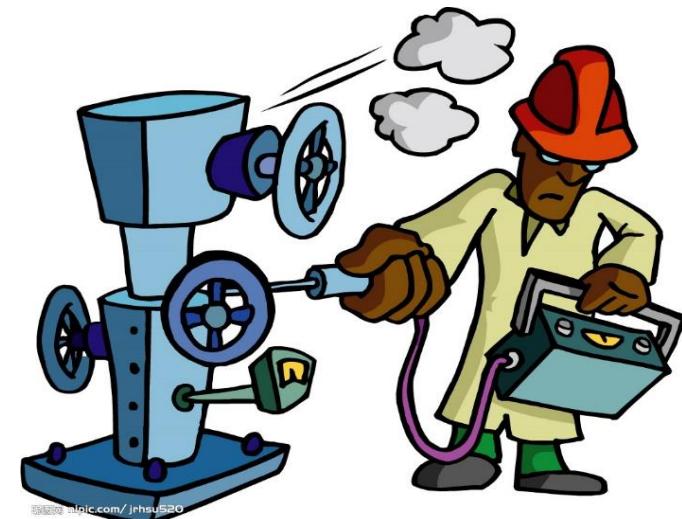
- 每天若在戰情表(Dashboard)上看到機台只有160台在Run (並不糟喔)

□ 改善

- 改善人員操作的作業順序 (eg. 有機台停機時要先處理，而非填表單)

□ 人機比計算方法

- $= \frac{\text{供給} \rightarrow \text{人}}{\text{需求} \rightarrow \text{機}}$
- $= \frac{\text{人力實際供給時數/天}}{\text{單位機台人力需求時數/天}}$
- $= \frac{\text{單位人力工作時長} \times \text{工作效率}}{\text{此時長} \times \text{機台於此時長內需要人力之比率}}$
- $= \frac{\text{工作效率}}{\text{機台需要人力之比率}}$
- $\cong \frac{\text{工作效率(Labor Eff)}}{1 - \text{稼動率(Machine Util)}}$ $= \frac{0.8}{1 - 0.656} = 2.33 \text{ (台)}$



- 因此長官可以知道買多少機台，需要搭配多少人力。

□ 收集資料 → 現場觀察 (工作研究 / time-motion study)

□ 優點

- 資料準確客觀
- 為當下現況資料 (較即時)

□ 缺點

- 很花時間
- 觀察的時間會很分散片段 (某製程/工序這次沒觀察到，要等下次)
- 缺少現場實務經驗，資料解釋不易 (需現場人員協助)

□ 建議可與學校合作~

- 學校不只是只有理論與方法的地方
- 更是培育人才的地方 (其實對現場來說，就是因為學生很多不懂 → 才能把觀察到的事實講出來 → 藉機學習、也與業界腦力激盪)

生產排程 Scheduling

□ 老師...我要做排程..

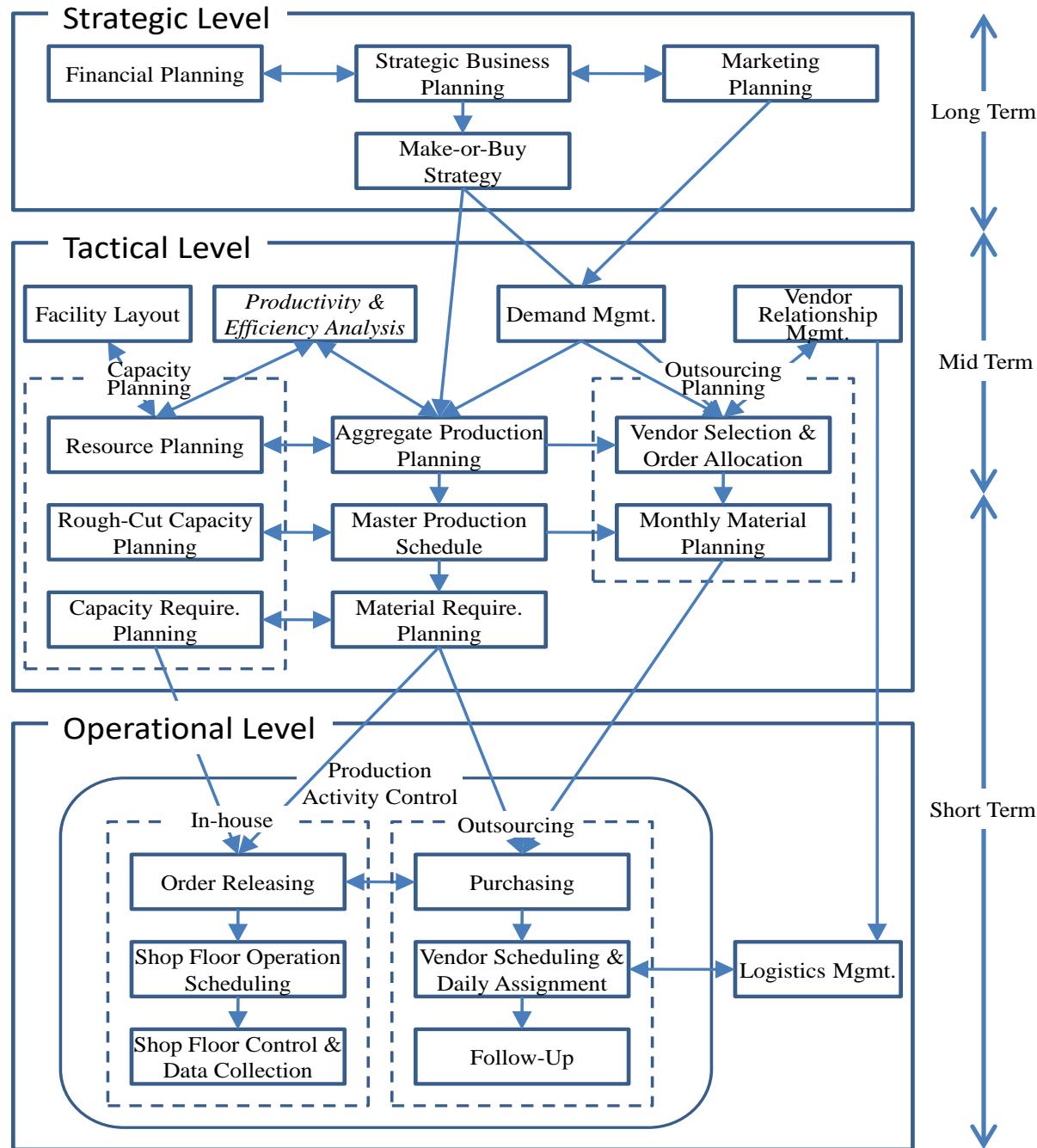
- Q: 那..那一個層級排程呢?
- A: 全部....
- 現場 → 瓶頸站
- 排程也不單只有現場
- 現場應該先導Lean...

□ Sales vs. 排程人員

- Sales → 只管接單
- 排程 → 只管達交
- 目標的衝突

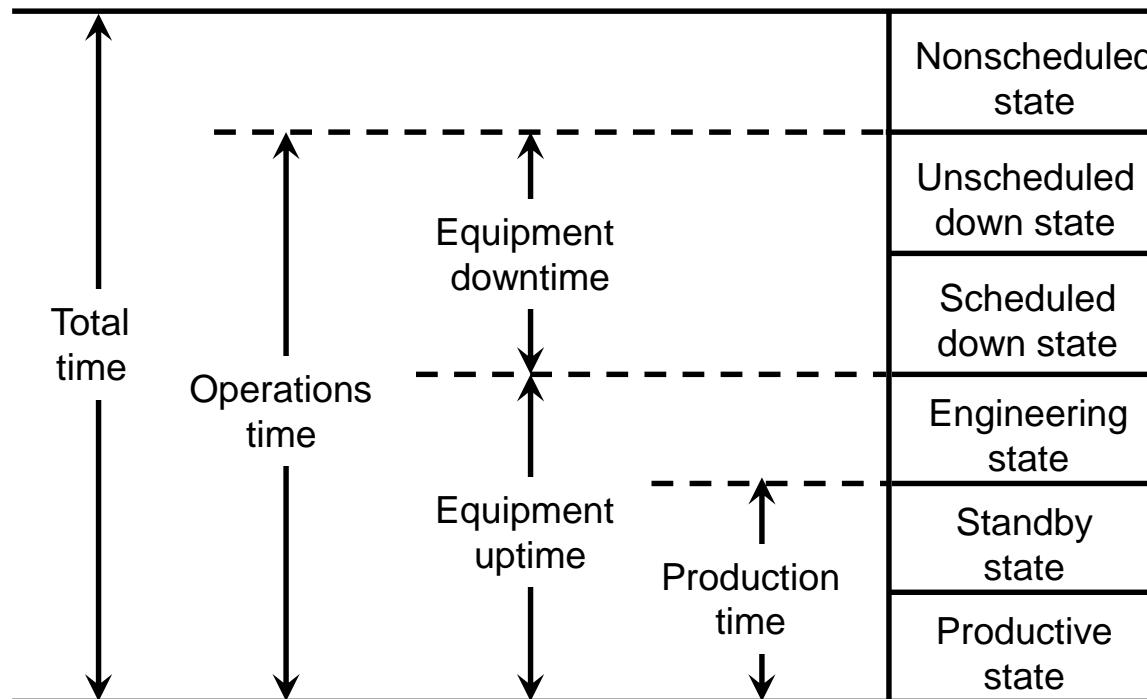
□ 排程 vs. 稼動率(utilization)

Lee and Johnson (2013)



□ Overall Equipment Effectiveness (OEE) revisited

— Nakajima (1988); de Ron and Rooda (2005):



$$OEE = \frac{\text{Theoretical production time for effective units}}{\text{Total time}}$$

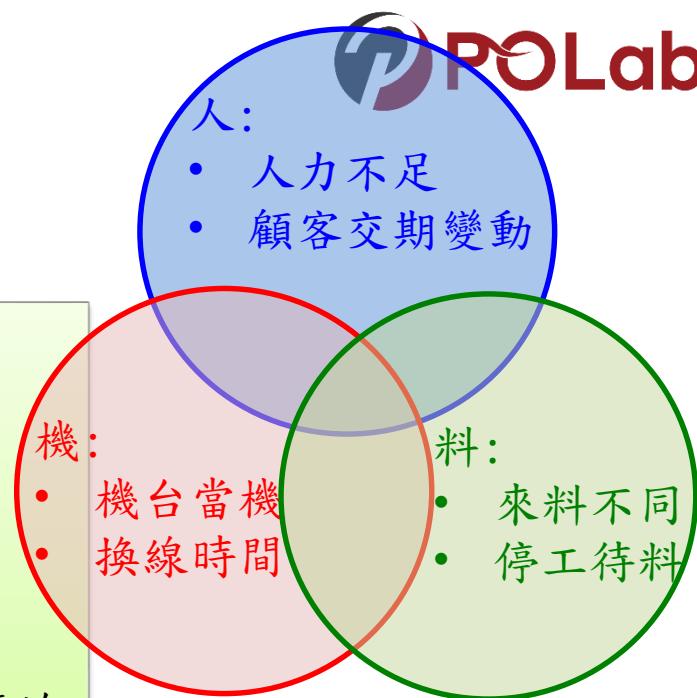
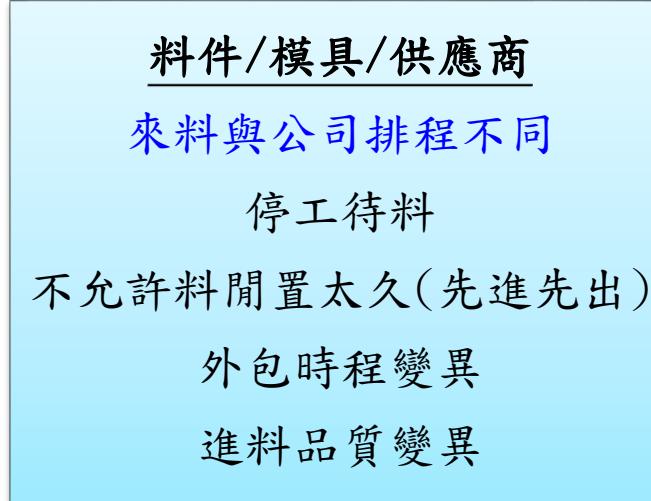
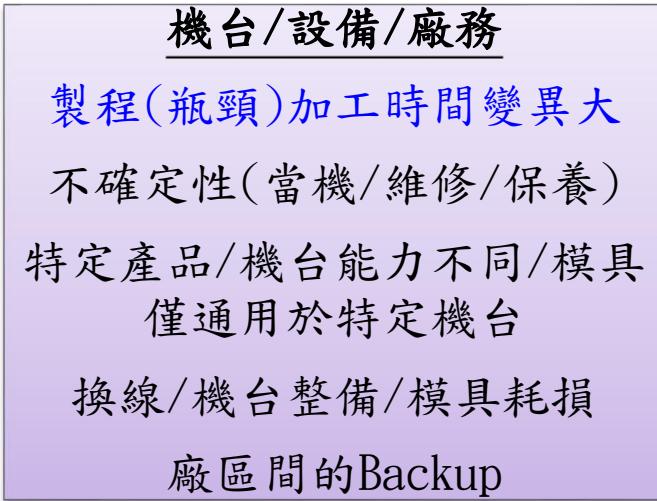
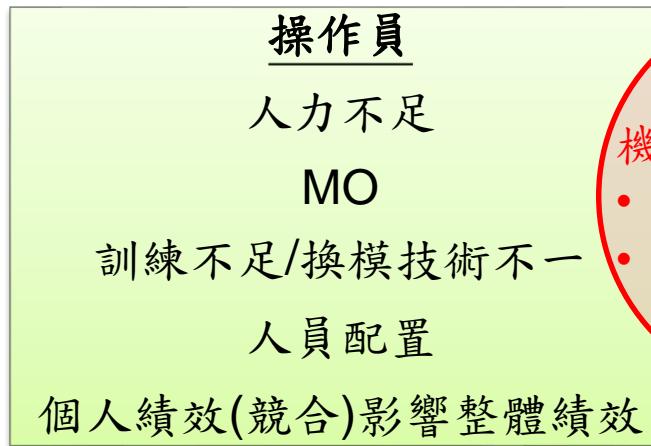
$$= \frac{\text{Equipment uptime}}{\text{Total time}} \times \frac{\text{Production time}}{\text{Equipment uptime}} \times \frac{\text{Theoretical production time for actual units}}{\text{Production time}} \times \frac{\text{Theoretical production time for effective units}}{\text{Theoretical production time for actual units}}$$

$$= \text{Available Eff} \times \text{Operational Eff} \times \text{Rate Eff} \times \text{Quality Eff}$$

— Machine **idleness** is identified explicitly through this definition

半自動化多目標生產排程

□ 現場生產排程系統困難與挑戰



□ 好的排程

- 有效提升機台稼動率/利用率 (utilization)
- 有效縮短生產週期/前置時間 (cycle time)
- Time-to-Market/ 產品達交

□ 排程 (Scheduling) (Graves, 1981)

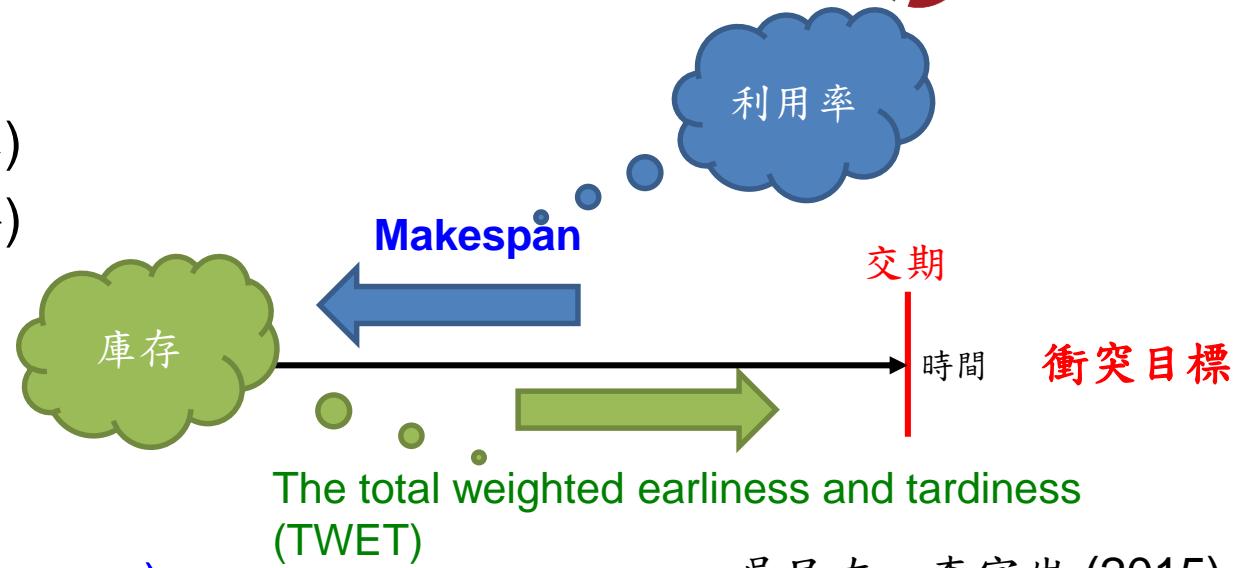
- 『在一段時間內，配置可用的資源來處理一群工作，以達成所設定的目標』
- Scheduling is the allocation of limited resources over time to perform a given set of jobs or activities. 換言之，是決定所有工作在機器上加工的起始時間以及加工順序，所以乃是同時考慮資源限制和執行限制之最佳化問題。
- 排程重點：針對對現場特定績效指標對不同工作做最佳的資源配置

Chien and Chen (2007)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
M1		T ₁₁					T ₂₄								
M2	T ₂₁				T ₁₂										
M3					T ₂₃		T ₁₃								
M4			T ₂₂							T ₁₄					

□ 目標概念

- 加工時間最短(每個訂單)
- 先進先出(未照排程到料)
- 優先順序
- 日期達交



□ 可量化(最小化)

- 最小最大完成時間(Makespan)
- 總加權完成時間(The total weighted completion time)
 - Cycle time/ machine idleness
- 生產線平衡(Line Balancing)
- 最少換模換線次數/時間
- 延遲訂單數量(The number of tardy jobs)
- 總加權延遲時間(The total weighted tardiness)
 - Meet due date
- 總加權提早及延遲時間(The total weighted earliness and tardiness)
 - On-time delivery/ WIP reduction

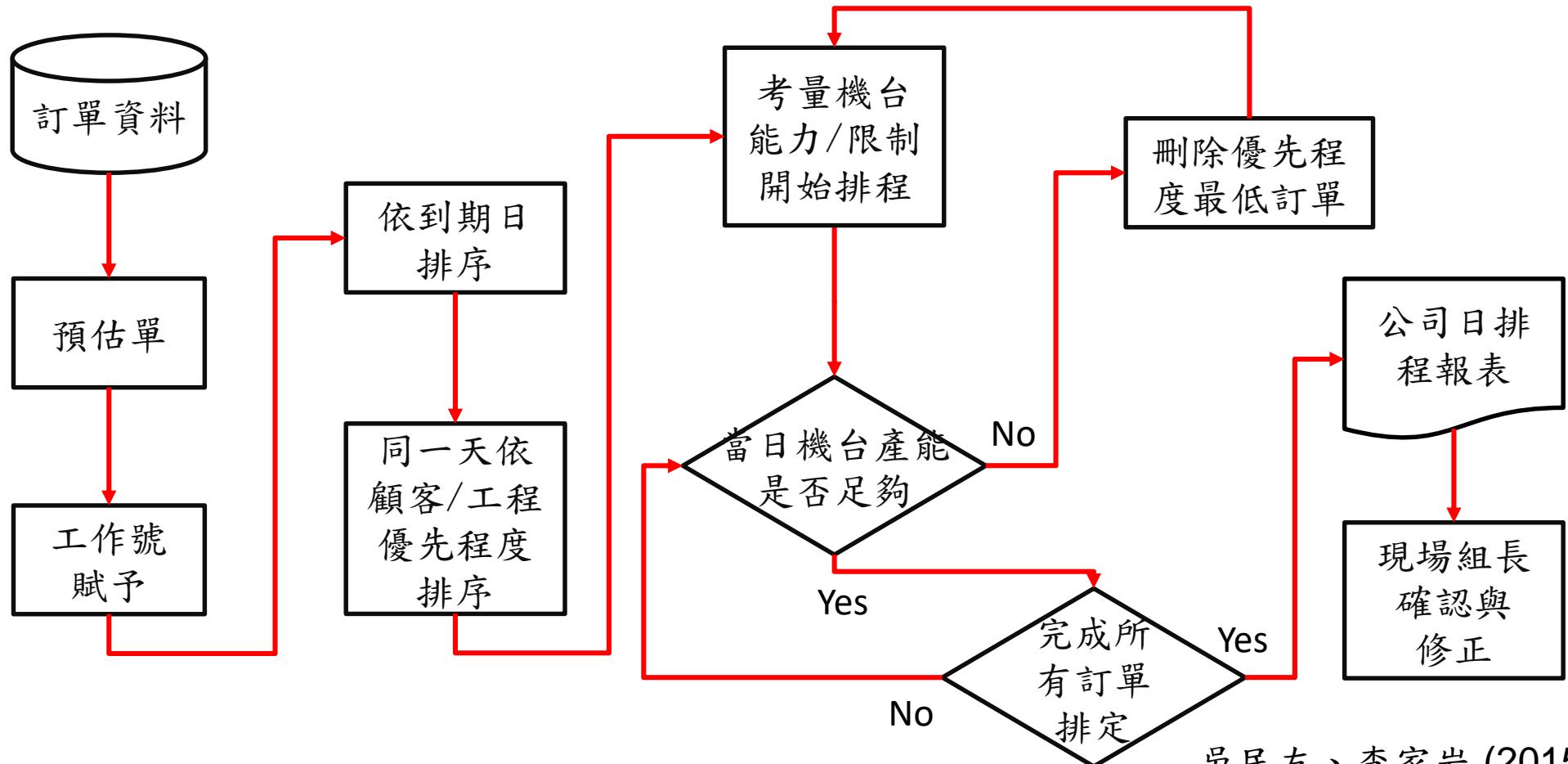
吳民友、李家岩 (2015)

□ 排程應用範疇

- 航空公司機師/空服員排班
- 機場飛機起降
- 車輛路徑派遣 (Vehicle Routing)
- 高鐵/火車/公車排班
- 電梯設定
- 微處理器系統
- 醫院排班
- 醫院復健排程
- 醫師手術排班
- 預約系統 (eg. 看病/餐廳/航班等)
- 對於服務顧客的管制
- 人機料的生產排程(協調多重資源的排程)

□ 現場人工排程

- MRP後，按照設定加工時間(processing time)，推算預計投料時間點
- 排程目標：滿足達交率 → 依到期日(Due day)排序工單



吳民友、李家岩 (2015)

□ Dispatching

- is to release of orders and instructions for starting of the production for an item in accordance with the 'route sheet' and schedule charts.

□ Dispatching Rule (優先順序法則): 用來選擇欲處理工作次序的一種簡單啟發式方法

- FCFS (First Come, First Served) 先到先服務
- SPT (Shortest Processing Time) 最短處理時間
- EDD (Earliest Due Date) 到期日
- Moore (Moore's Algorithm) 到期日⁺

- CR (Critical Ratio) 關鍵性比率

- S/O (Slack per Operation) 每個作業的寬裕時間

- Rush 緊急事件

$$CR = \frac{(到期時間) - (現在時間)}{\text{現場需要加工時間}}$$

$$S/O = \frac{(到期時間) - (現在時間) - \text{現場需要加工時間}}{\text{剩餘作業數}}$$

- Examining the scheduling of n jobs on 1 machine (or the handling of the **bottleneck machine**)

- Performance Metrics

- Mean Flow Time
- Average Job tardiness
 - Tardiness is nonnegative; however, the lateness can be negative.
- Number of Tardy Jobs

- Lateness vs. Tardiness

- $L_j = c_j - d_j$, where c_j is completion time and d_j is due date
- $T_j = \max\{0, L_j\}$

- Let's try FCFS, SPT, EDD and CR

□ Example (Nahmais, 2005)

- Machine shop has 5 unprocessed jobs (J1, J2, J3, J4, J5) numbers by order they entered Bottleneck machines queue:

Job #	Process Time (p_i)	Due Date (d_i)
1	11	61
2	29	45
3	31	31
4	1	33
5	2	32

- Notations

- p_j : process time
- c_j : completion time
- d_j : due date

□ FCFS

Sequence	P. Time	Comp. Time	D. Date	Tardiness
J1	11	11	61	0
J2	29	40	45	0
J3	31	71	31	40
J4	1	72	33	39
J5	2	74	32	42
Totals	(74)	268		121

- Mean Flow Time: $(268)/5 = 53.4$
- Avg Tardiness: $(121)/5 = 24.2$
- # Tardy Jobs: 3

The “makespan” is fixed in this case (N jobs on 1 machine). Only consider the “job sequence”.

□ SPT

Sequence	P. Time	Comp. Time	D. Date	Tardiness
J4	1	1	33	0
J5	2	3	32	0
J1	11	14	61	0
J2	29	43	45	0
J3	31	74	31	43
Totals	74	135		45

- Mean Flow Time: $(135)/5 = 27$
- Avg Tardiness: $(43)/5 = 8.6$
- # Tardy: 1

□ EDD

Sequence	P. Time	Comp. Time	D. Date	Tardiness
J3	31	31	31	0
J5	2	33	32	1
J4	1	34	33	1
J2	29	63	45	18
J1	11	74	61	13
Totals	74	235		33

- Mean Flow Time: $(235)/5 = 47$
- Avg Tardiness: $(33)/5 = 6.6$
- # Tardy: 4

□ CR: Iterative Process

- looks at time remaining between current time and due date
- considers processing time as a percentage of remaining time
 - CR = 1.0 means just enough time
 - CR > 1 .0 more than enough time
 - CR < 1.0 not enough time

- Compute:

$$CR = \frac{(Due_Date) - (Cur_Time)}{Processing_Time}$$

- Current time updates after each selection by adding scheduled Process Time to current time

□ CR

JOB	Pr. Time	D. Date	CR
Current Time = 0			
1	11	61	5.546
2	29	45	1.552
3	31	31	1.00
4	1	33	33
5	2	32	16

JOB	Pr. Time	D. Date	CR
Current Time = 31			
1	11	61	2.727
2	29	45	.483
4	1	33	2
5	2	32	0.5

□ CR

JOB	Pr. Time	D. Date	CR
Current Time = 60			
1	11	61	0.091 do last
4	1	33	-27*
5	2	32	-14**

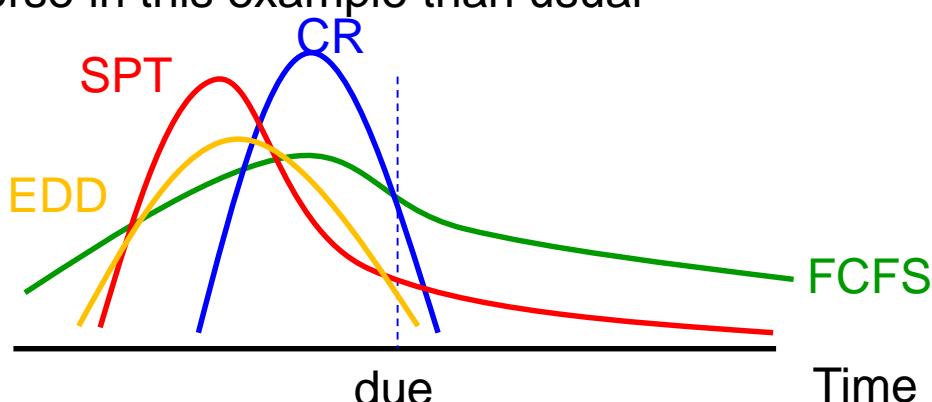
Tie: use SPT

- Mean F. Time: $(289)/5 = 57.8$
- Mean Tardiness: $(87)/5 = 17.4$
- # Tardy: 4

JOB	C. Time	D. Date	Tardy
Summary			
3	31	31	0
2	60	45	15
4	61	33	28
5	63	32	31
1	74	61	13
Total:	289		87

□ Performance Comparison

- FCFS:
 - Does not consider job attributes (due dates, processing times) and not do well
- SPT
 - Minimizes total time jobs spend in the system (mean flow time)
 - Given great mean flow time and does well relative to tardiness but does not consider due dates at all (Note that some job may queue forever!)
- EDD
 - Minimizes the Maximal Tardiness (T_{max}), $T_{max} = \max\{0, L_j\}$
 - Does well relative to due dates (not optimal in any case but intuitively attractive)
- CR
 - Does worse in this example than usual



然而...

- EDD or SPT是現場常用的rules
- 然而，各有缺點
 - EDD缺點：現場一直救火，急件插單，滿足大客戶，放棄小客戶
 - SPT缺點：有些job可能永遠都沒機會加工
- 此外，簡單的Dispatching rule未考慮到
 - "換線換模"的時間與成本
 - 生產線平衡(Line Balance)
- 延伸方法
 - 快速換模技術(Single-Minute Exchange of Die, SMED)
 - 生產線平衡技術 (Line Balancing)

Lee and Johnson (2013)

□ Single-Minute Exchange of Die, SMED (新鄉重夫 Shingo, 1985)

- Rapid changeover and setup time reduction in converting current manufacturing process to manufacture the next product; improves production flow and reduces lot sizes.
- Long setup time leads to a small number of setups, larger batch sizes, larger WIP inventories and poor process flow.

□ 思維的改變

- 新鄉重夫：「快速換模絕對不是 "技巧的問題"，而是 "思維的問題" 」
- 豐田汽車副社長大野耐一：「生產現場中，"人的智慧"是成敗關鍵」

Shingo, S. (1985). A Revolution in Manufacturing: The Smed System, Productivity Press.

□ SMED流程

- 1. 記錄整個換模(線)現況過程和時間
 - 工作研究：人機程序圖、時間與動作研究(time-motion study)
- 2. 區分內部換模與外部換模
 - 內部換模：指必須在機器停止生產狀態下，才能進行的換模動作。
 - 外部換模：指機器在生產運轉中，仍然可以進行的換模動作。
- 3. 將內部換模作業儘可能轉到外部換模作業
 - 工具、模具、零組件標準化，不尋找、不移動、不多用
 - 工具預熱、預裝配、預設定、預清潔、預熱等等
- 4. 改善內部換模
 - 平行作業、
 - 工具放置的方法/位置統一/尺寸統一
- 5. 改善外部作業
 - 把工具/規尺存放在機器旁邊
 - 備好工具更換準備的核查清單、提供詳細的工具更換流程記錄單
- 6. 標準化新的換模準備作業

科建顧問(2009)，"少量多樣生產的關鍵-快速換模(線)技術"，http://www.kind.com.tw/main.php?action=article_intro&id=52。

- JSP is among the hardest **combinatorial optimization** problems.

- NP-hard problem

- Heuristic Method (Priority Rule)

- Shortest processing time (SPT), earliest due date (EDD), etc.
- Pros: easy to understand
- Cons: poor performance for complicated production line

Local optimum
Short running time



- Meta-Heuristic Algorithm (**Tabu**, Simulated Annealing, **Genetic Algorithm**)

- Approximated-optimization approach
- Pros: provide a good solution efficiently
- Cons: cannot guarantee the global optimum

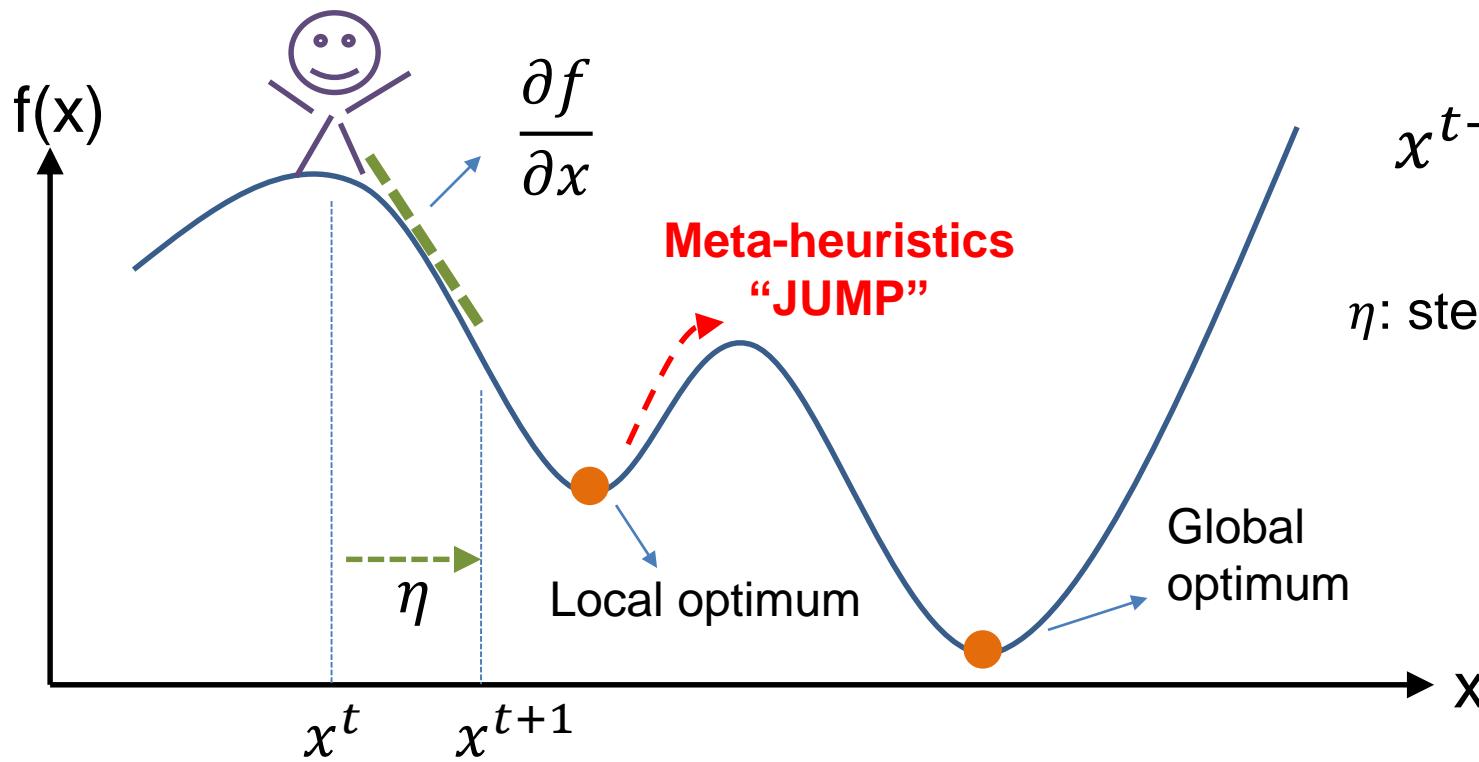
- Mathematical Programming

- Optimization-based approach formulated by mixed integer programming
- Pros: Guarantee global optimum
- Cons: computational burden for large-scale problem

Global optimum
Long running time

□ Definition of Meta-heuristics (Glover and Laguna 1997)

- A meta-heuristic refers to a **master strategy** that guides and modifies other heuristics to produce solutions beyond those that are normally generated in a quest for **local optimality**.
- For the “Minimization Problem”...



Gradient Descent (梯度下降法)

$$x^{t+1} \leftarrow x^t + \eta \frac{\partial f}{\partial x}$$

η : step size (learning rate)

□ Basis

- Tabu Search (TS) has its roots in methods that cross boundaries of feasibility and local optimality
- TS was first proposed by Glover (1977)
- The word tabu (or taboo) comes from Tongan, a language of Polynesia, where it indicates things that cannot be touched because they are sacred. Now it means “a **prohibition imposed by social custom**”
- In TS, tabu status of forbidden elements shift according to time and circumstance, based on an evolving **memory**.
- Tabu status can be overruled for a preferable alternative.
- Hence **TS uses adaptive (flexible) memory (long-term and short-term)**
- Neighborhood Search
- **exploitation (開採)** of good solutions and **exploration (探索)** of new promising regions
- Stopping criteria (to **terminate** the searching procedure)
 - The maximal number of iteration/ The objective function is unchanged given a number of iteration/ Running time

□ Tabu Search Algorithm (Glover and Laguna, 1997)

- 1. Find an initial solution $x_0 \in X$, set $x^{now} = x^{best} = x_0$, initialize memory
- 2. Intensification phase:
 - 2.1 If termination condition (e.g. simple iteration count, no admissible improving move, no change in x^{best} in so many iterations) is satisfied, then go to step 3
 - 2.2 Choose the best $x^{next} \in N(x^{now})$ such that x^{next} is not tabu or satisfies aspiration criterion
 - 2.3 Move from x^{now} to x^{next} , i.e. set $x^{now} = x^{next}$
 - 2.4 If x^{now} is better than x^{best} , then set $x^{best} = x^{now}$
 - 2.5 Update recency based memory (tabu classifications), frequency based memory and/or critical event memory (elite solutions), return to step 2.1
- 3. Diversification phase:
 - 3.1 If termination condition is satisfied, then stop
 - 3.2 Using frequency based memory and/or critical event memory, find a new starting point x^{now} , return to step 2

□ Single-machine total weighted tardiness problem (Chern, 2004)

Jobs	Process Time (p_j)	Due Date (d_j)	Weights (w_j)
1	10	4	14
2	10	2	12
3	13	1	1
4	4	12	12

- Exhaustive enumeration method (窮舉法): $4!$ Cases
- Iteration 1: Initial solution (2,1,4,3) with total weighted tardiness
 - $$T(2,1,4,3) = w_2 \max\{p_2 - d_2, 0\} + w_1 \max\{(p_2 + p_1) - d_1, 0\} + w_4 \max\{(p_2 + p_1 + p_4) - d_4, 0\} + w_3 \max\{(p_2 + p_1 + p_4 + p_3) - d_3, 0\}$$

$$= 12 \times (10 - 2) + 14 \times (10 + 10 - 4) + 12 \times (10 + 10 + 4 - 12) + 1 \times (10 + 10 + 4 + 13 - 1)$$

$$= 500$$
 - Set initial Tabu list, TABU = < >, and size of tabu list = 2

□ Iteration 2: neighborhood search at (2,1,4,3).

- Schedules obtained by pairwise interchanges are:
(1,2,4,3), (2,4,1,3), and (2,1,3,4)
- $T(1,2,4,3) = 480$, $T(2,4,1,3) = 436$, and $T(2,1,3,4) = 652$
- (2,4,1,3) is the best from the non-tabu results
- Update Tabu list, TABU = < (1,4) >

□ Iteration 3: neighborhood search at (2,4,1,3).

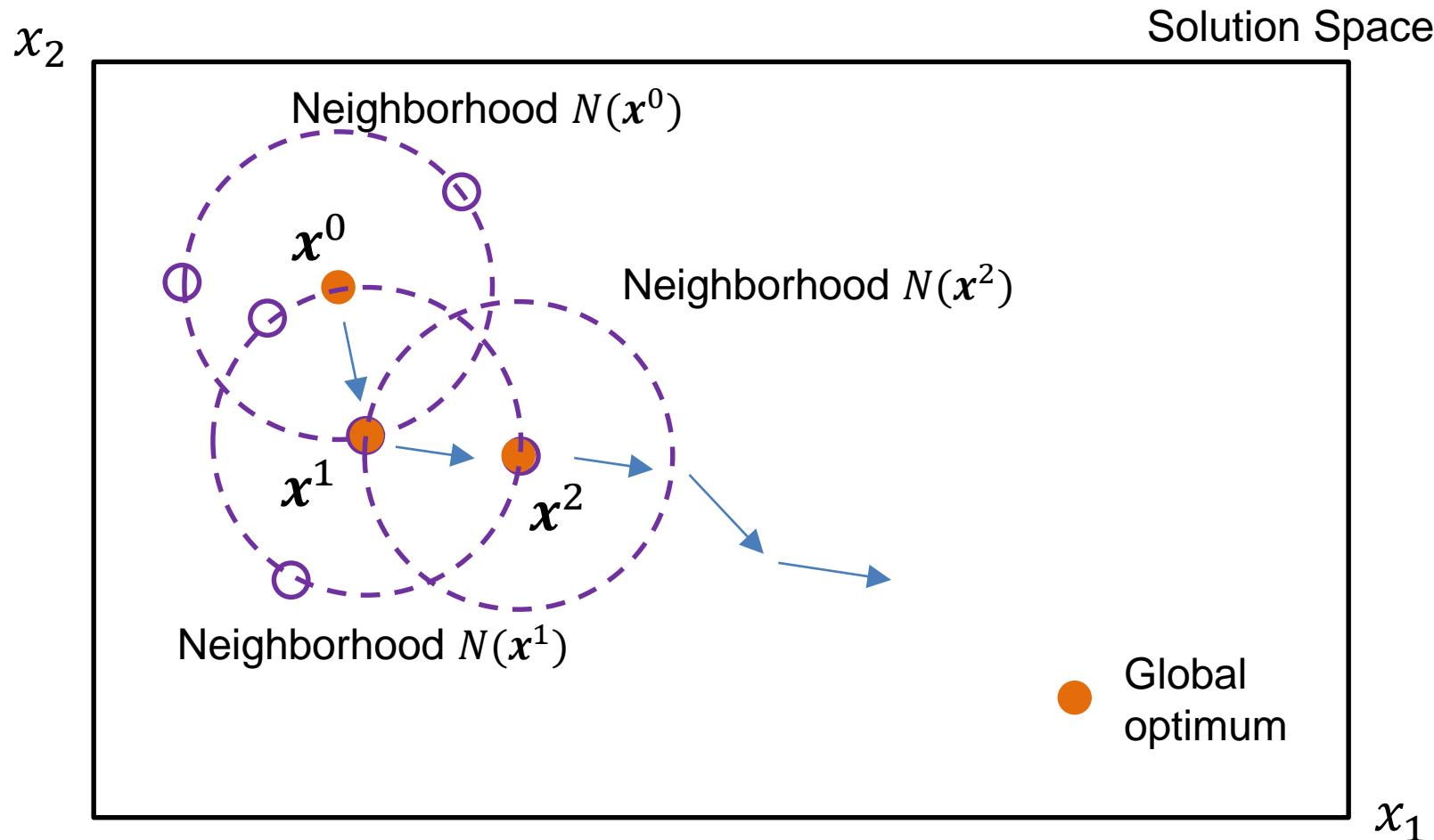
- Schedules obtained by pairwise interchanges are:
(4,2,1,3), (2,1,4,3) TABU!, and (2,4,3,1)
- $T(4,2,1,3) = 460$, $T(2,1,4,3) = 500$ TABU!, and $T(2,4,3,1) = 608$
- (4,2,1,3) is the best from the non-tabu results
- Update Tabu list, TABU = < (2,4), (1,4) > (size of tabu list = 2)

- Iteration 4: neighborhood search at (4,2,1,3).
 - Schedules obtained by pairwise interchanges are:
(2,4,1,3) **TABU!**, (4,1,2,3), and (4,2,3,1)

- $T(2,4,1,3) = 436$ **TABU!**, $T(4,1,2,3) = 440$, and $T(4,2,3,1) = 632$
- (4,1,2,3) is the best from the non-tabu results
- Update Tabu list, TABU = < (2,1), (2,4) >

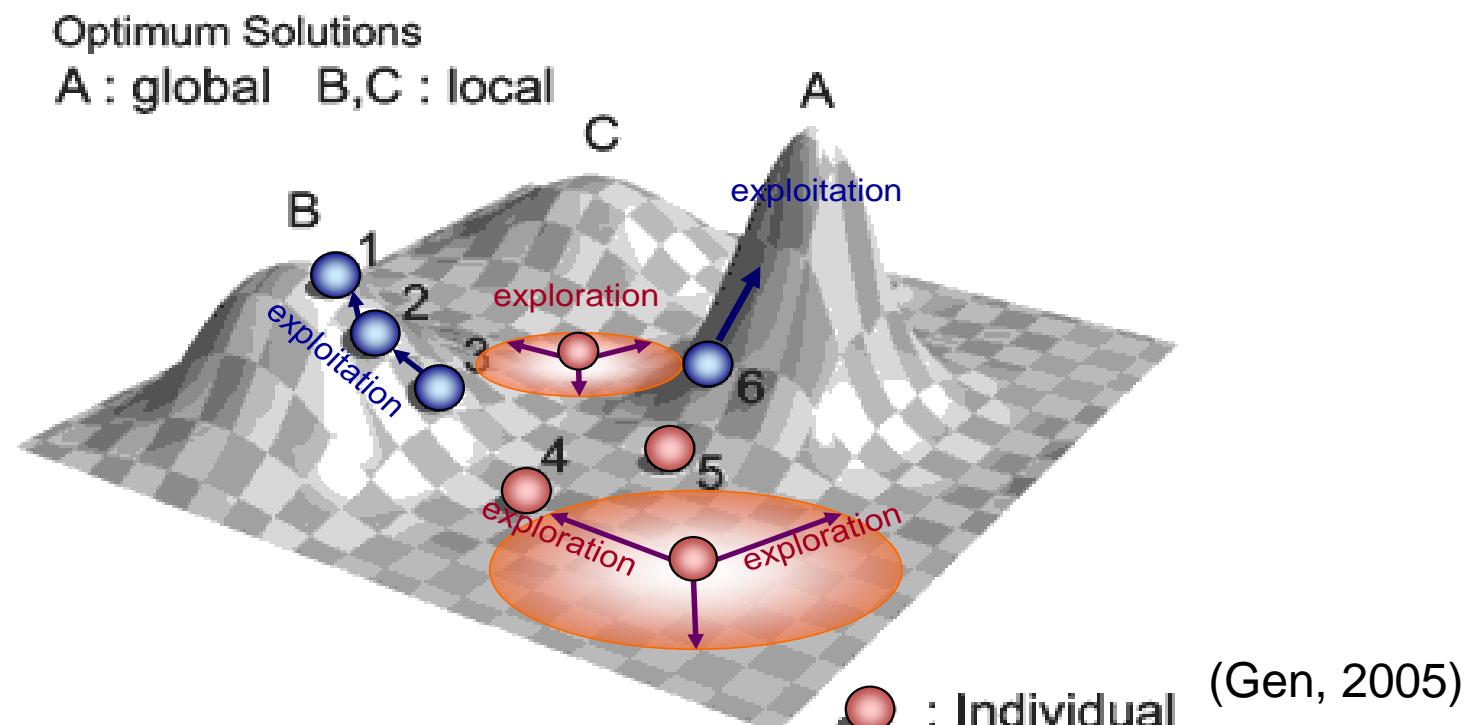
- Iteration 5: neighborhood search at (4,1,2,3).
 - Schedules obtained by pairwise interchanges are:
(1,4,2,3), (4,2,1,3) **TABU!**, and (4,1,3,2)
 - $T(1,4,2,3) = 408$, $T(4,2,1,3) = 460$ **TABU!**, and $T(4,1,3,2) = 586$
 - (1,4,2,3) is the best from the non-tabu results
 - Update Tabu list, TABU = < (4,1), (2,1) >

□ Stochastic (Random) Search



□ Intensification and diversification

- Intensification: a form of **exploitation** (開採)
 - Based on modifying choice rules to encourage good move combinations and solution attributes, and it may lead to return to attractive regions
 - Examines neighbors of prerecorded elite solutions
- Diversification: a form of **exploration** (探索)
 - Examines unvisited regions, generates different solutions



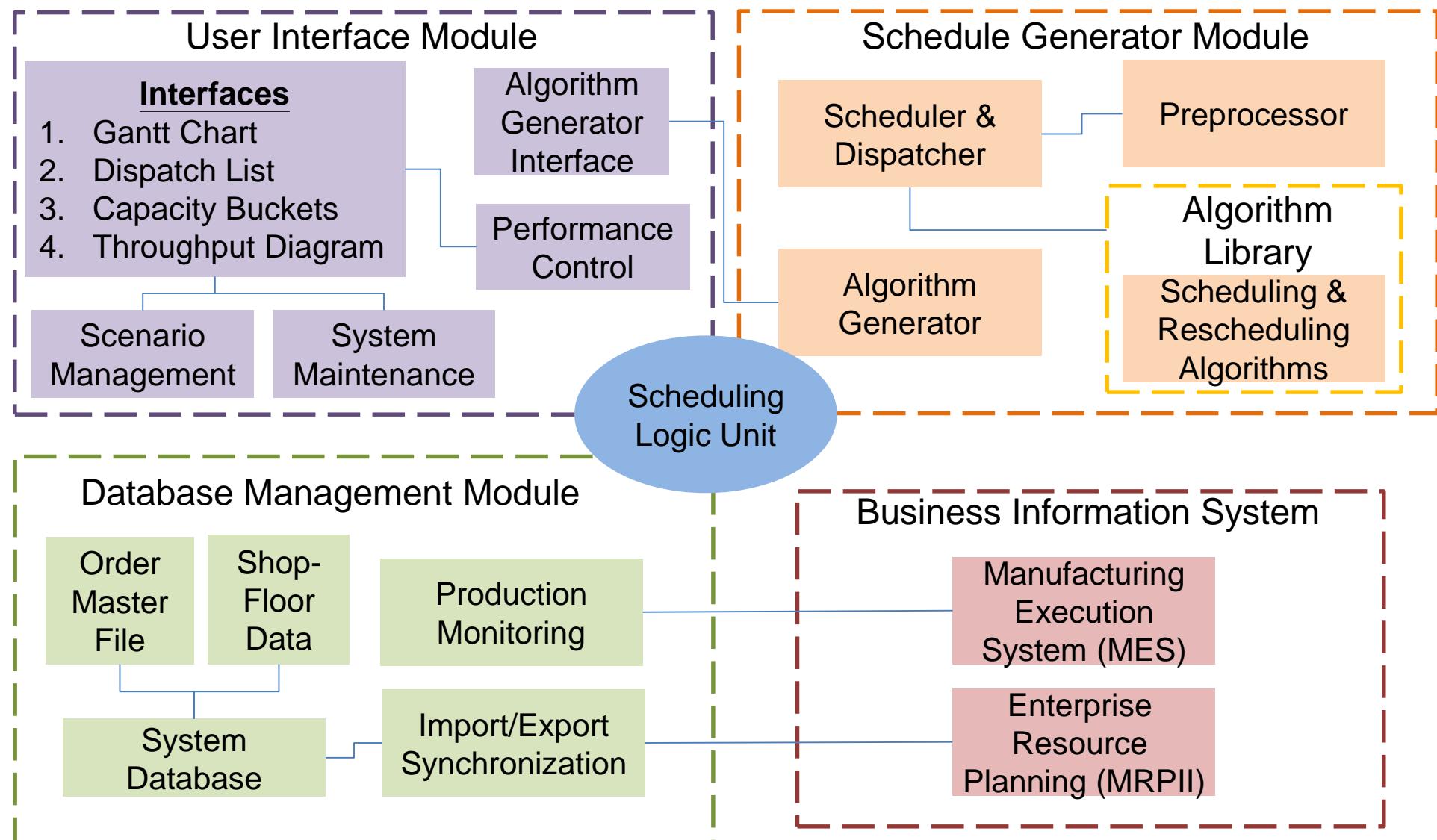
If you ...



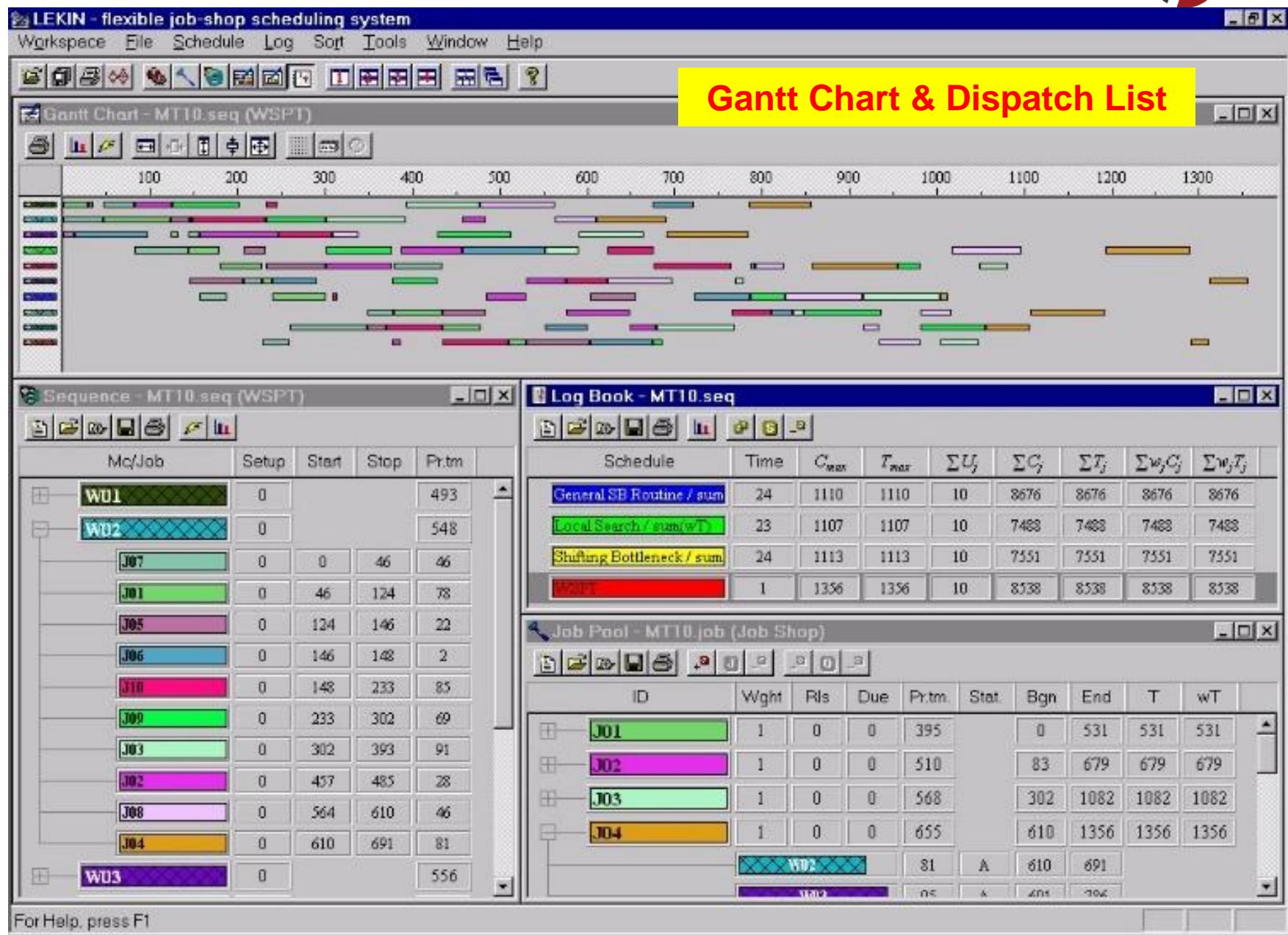
If you change the size of TABU list, what's the result?

A trade-off between exploitation and exploration!!!

Module Architecture of MFG Scheduling Systems



revised from Framinan and Ruiz (2010)



Multi-Objective Scheduling

Schedule	Time	C_{max}	T_{max}	$\sum U_j$	$\sum C_j$	$\sum T_j$	$\sum w_j C_j$	$\sum w_j T_j$
General SB Routine / sum(wT)	24	1110	1110	10	8676	8676	8676	8676
Local Search / sum(wT)	23	1107	1107	10	7488	7488	7488	7488
Shifting Bottleneck / sum(wT)	24	1113	1113	10	7551	7551	7551	7551
WSPT	1	1356	1356	10	8538	8538	8538	8538

Objective Chart - MT10.seq

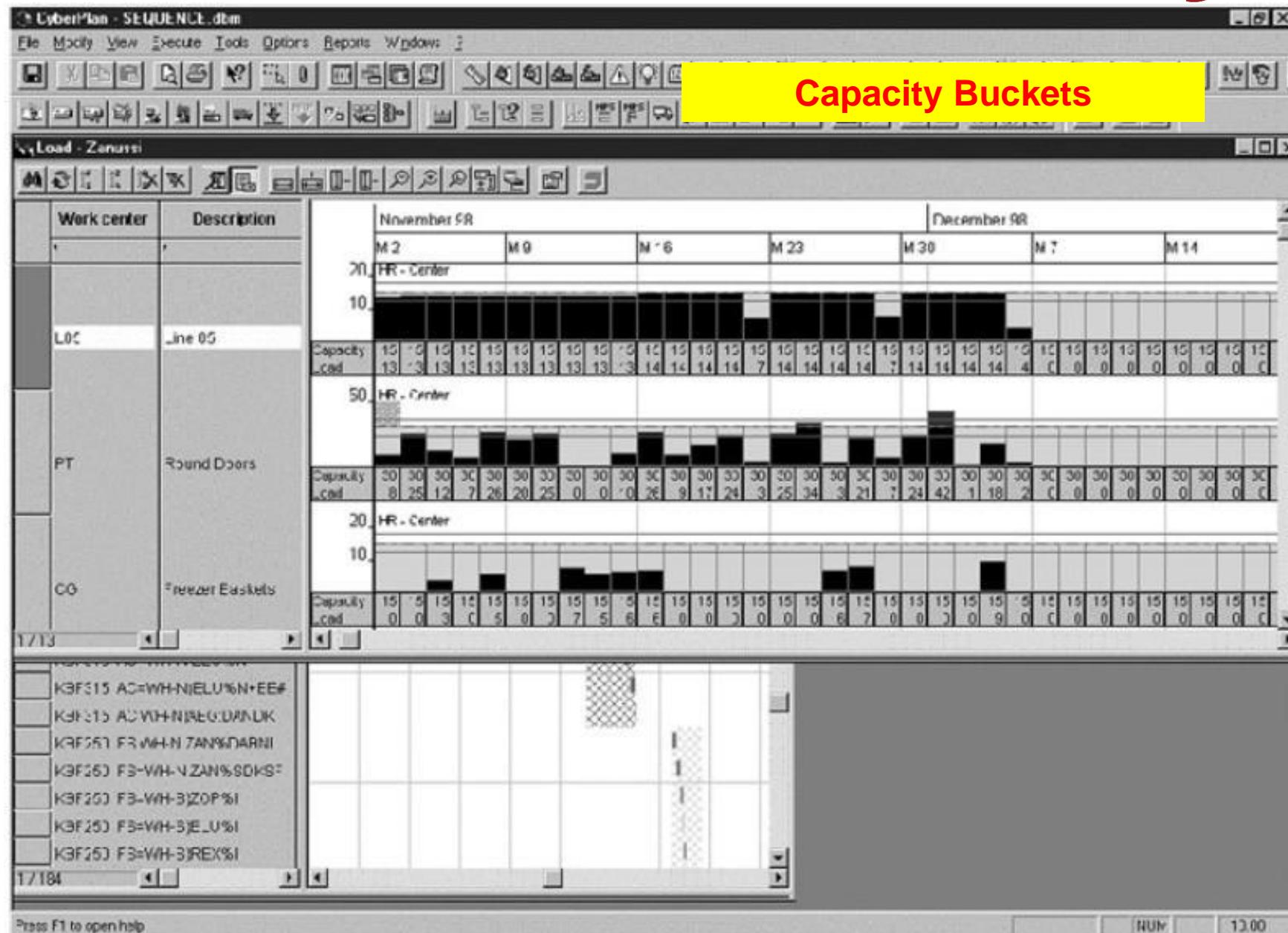
$\sum w_j T_j$

C_{max}

- General SB Routine / sum(wT)
- Local Search / sum(wT)
- Shifting Bottleneck / sum(wT)
- WSPT

Job Pool Sequence Gantt Cha

For Help, press F1



□ 排程的好處多多

- 先從某站點A開始做(可能是瓶頸)
- 做完上下游的問題自然會浮出來
 - 上游變卡了，WHY? 下游變卡了，WHY?
 - 改善的機會!!
 - 但並不是把A產出極大化→上游慢，則A的稼動率會很低..下游慢，則會堆貨

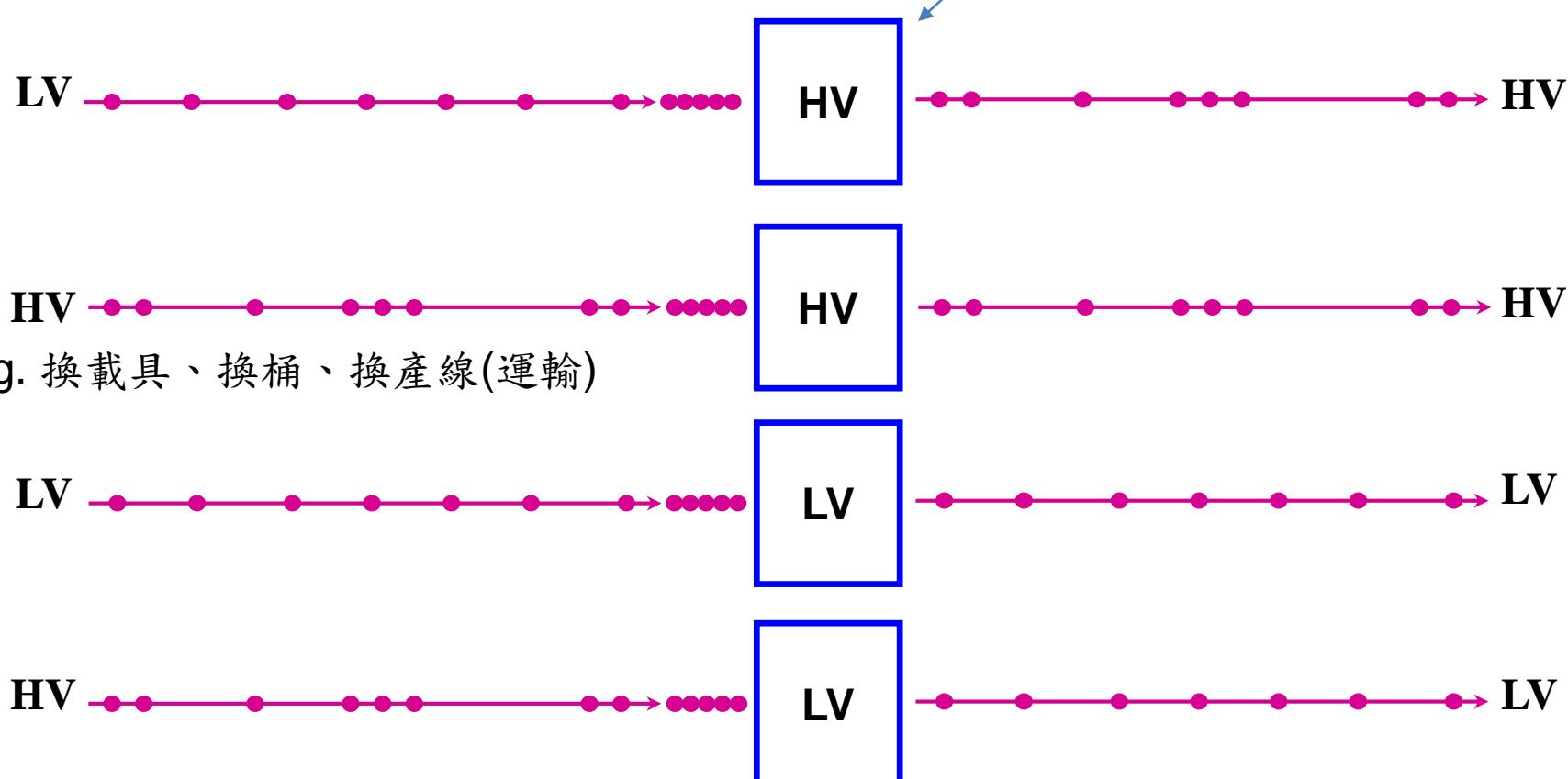
□ 講究的是synergy

- 排程需整合人、機、料、法、環等資源
 - 人員技能、機台能力、模具原料庫存、批次工作站、Queue time limit
- 部份手動排程調整rescheduling大部分有其必要 → 半自動化
- Web-based scheduling system
- 排程系統長期發展應與其他系統整合
 - 庫存管理、人員排班、維修保養排程、產能規劃與定價(?)
 - 供應鏈上下游的系統整合 (資訊透明化)

□ 瓶頸站的重要性- Flow Variability

- Bottleneck (high utilization station)

Must check the **variance** of processing time



Conclusion: flow variability out of a high utilization station is determined primarily by process variability at that station.

Hopp, W. and Spearman, M., 2001. Factory physics. 2nd ed. Boston, MA: Irwin McGraw-Hill.

□ 靜態n個工作排程最佳化

- 實務上是動態的，因此當新的工作(job)來，必須要rescheduling
- Note1: 設定一個未來時段排程不可變動，eg. 未來6hrs內的排程固定
- Note2: 插入空隙(slack time)於排程中，以防止急單或機台當機狀況

□ 複雜的現場環境

- 機台能力(e.g. 機差)、製程限制(e.g. 繩機)、訂單相關(e.g. 優先權)、工序相關(e.g. 前置時間)、時間相關(e.g. 等候限制)、機台狀況(e.g. 當機、維修保養)

□ 工作優先權的動態調整

- 學理上會固定工作/工單的優先權，然而實務上權重會動態調整，甚至最低優先權的工作突然變成最高優先權的工作

□ 目標函數/懲罰函數通常不是"線性"的

- 線性：Penalty function = $\max\{0, c_j - d_j\}$
- 非線性：Penalty function = $\max\{0, (c_j - d_j)^2\}$

□ 多個目標函數，且每個目標函數的權重會隨時間動態調整

- A. 最小化 sum of the sequence dependent setup times → increase throughput
- B. 最小化 total weighted tardiness → maintain quality of service
- If workload is relatively heavy, then A; otherwise, B.

□ 排程常跟人員排班與加班有密切關係

□ 加工時間的隨機性不易建模

- 98%是固定值(eg. 10mins); 2%會加上平均數相當大的指數分配時間
- 自動化製造與組裝現場尤其常發生
- 機台或機器人平常加工很順；但突發狀況會立即造成非常糟的加工時間
- 加工時間也常受到人員訓練(經驗曲線)以及模(刀)具耗損的影響

□ 排程系統開發過一陣子就...

- Out-of-fashion database without updated order, upgraded-version machine (throughput change), labor (learning curve), ...
- Rescheduling

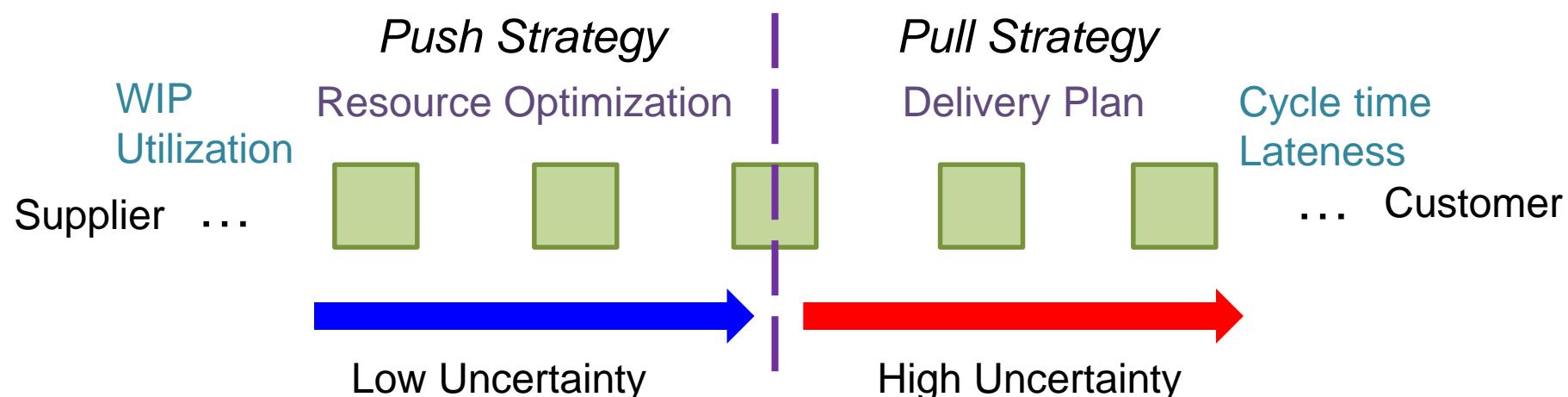
由於眾多因素的複雜關係，排程基本上都是”客製化”

□ 如果現場的"剩餘"產能是足夠的

- 某工作可以在這機台處理，也可再另外一個機台 (有選擇性)
- 先考慮 Setup time 與 due date → high priority
- 再考慮 flexibility (有多少機台可以做這 job) → low priority

□ 如果現場有兩個瓶頸，**下游的瓶頸排程先處理!** (WHY?)

- Hint: WIP
- 但下游若是偏向組裝線 (i.e. delayed differentiation)，人為調整因素較多，自動化導入也就愈加困難
- 推拉界限(push-pull boundary)視角



□ 排程最強的地方在於..一翻兩瞪眼，立即見效

- Cycle Time Reduction
 - Time-to-market
 - Inventory = Lead Time + Uncertainty
- 問題的浮現
 - 降低bottleneck的variability → 專線, 綁機, 有經驗的師傅, etc.
 - Bottleneck上下游擴散開
- 促進衛星工廠改善

□ 教育訓練上

- 排程是組合最佳化(Combinatorial optimization)的問題，通常為NP-hard
- 一旦學會解這類題目，應用範圍就變廣了
 - 運輸最短路徑問題 (eg. travelling salesman problem, TSP)
 - 網路結構最佳化問題 (eg. minimum spanning tree, MST)
 - 財務投資組合問題/ 人員排班問題/ 資源優化的問題/ 設施規劃/廠址選擇
 - 產能規劃/庫存管理/ 電信基地台架設 選址問題/ 車輛配送路徑問題等

- 1. 開發資訊系統連結資料庫、ERP、MES串起來，導入人工智慧AI演算法優化(eg. 基因演算法GA)，排程就”成功”了？
- 2. 排程有沒有一般化(generalized)的模組？
- 3. 排程系統真正關鍵在交接、維護、與持續更新
 - 教育訓練
 - 現場動態調整

- 到這裡..可以發現...
- 排程...好像不是要預測些什麼!?
- FCFS, SPT, EDD, CR 哪一個好?
- 製造資源的最佳化 → 處方性分析

□ 再過一陣子..

“業界缺的就不是Data Scientist，而是如何將Model產生的Pattern或預測結果，與公司、現場、內部、外部的資源整合，以進行Prescriptive Analysis”
(范治民博士，富士康大數據解決方案處，2017)

□ 預測但不做決策..預測就可惜了..

- 然而，決策是根據“資源”來做判斷

RESOURCES



在製造業中
自動化程度愈高...
排程需求就愈重要!

而排程就是一種
Prescriptive Analysis (處方性分析)
根據資源”人機料法環”達成的綜效synergy

2020

Intelligent Manufacturing and Industry Practice Workshop

智慧製造與產業實務研習會

研習會議程

09:00~09:20 註冊與報到

09:20~09:30 開幕致詞 – 國立成功大學 蘇慧貞 校長
電機資訊學院 許渭州 院長

09:30~10:20 講者：漢翔航空工業 吳天勝 研發長
人工智能與作業研究在航空製造業的應用

10:20~10:40 大合照、茶敘與交流 (Break)

10:40~11:30 講者：旺宏電子 陳瑞坤 副總經理
智慧製造在半導體業的應用

11:30~12:20 講者：新漢科技 林弘洲 總經理
智造賦能、工業4.0關鍵技術與對標應用

12:20~13:30 中餐時間

13:30~14:20 講者：成功大學 陳裕民 特聘教授
智能化企業再造

14:20~15:10 講者：友達光電 王德弘 處長
AI驅動新智造

15:10~15:30 茶敘與交流 (Break)

15:30~16:20 講者：富強鑫精密工業 林宗彥 部經理
挖掘射出成型工藝數據的力量

16:20~16:50 論壇與討論
引言人：國立成功大學 李家岩 教授
與談人：講者群

16:50 閉幕、交流討論與賦歸

主辦單位： 國立成功大學製造資訊與系統研究所  國立成功大學資訊工程學系

協辦單位： 科技部 科技部 工業工程與管理學門  AIMS 人工智慧製造系統研究中心  台灣作業研究學會

GLORIA NCKU 成功大學國際產學聯盟

2019 Intelligent Production Scheduling System Workshop

智慧生產排程系統研習會

2019

Intelligent Production Scheduling System Workshop

智慧生產排程系統研習會

研習會議程

08:45~09:05 註冊與報到 (領取講義)

09:05~09:10 開幕致詞

台灣作業研究學會理事長 林義貴 講座教授

逢甲大學工業工程與系統管理學系主任 王逸琦 教授

09:10~10:30 生產排程學理基礎

講者：逢甲大學工業工程與系統管理學系 王宏錨 教授

10:30~10:50 大合照、茶敘與交流 (Break)

10:50~12:10 生產排程數學規劃

講者：逢甲大學工業工程與系統管理學系 王宏錨 教授

12:10~13:30 中餐時間

13:30~15:00 生產排程AI演算法

講者：亞洲大學經營管理學系 鄭家年 教授

15:00~15:20 茶敘與交流 (Break)

15:20~16:40 智慧生產排程實務個案

講者：亞洲大學經營管理學系 鄭家年 教授

16:40 閉幕、交流討論與賦歸

逢甲大學工學館504

DATE

2019/06/15 (六) 09:05~16:40

報名費用：2500元 (含茶飲、中餐、與講義)

對象學員：大專院校師生、業界工程師

聯絡電話：06-2757575 轉34223 蘆先生

聯絡信箱：n07051305@mail.ncku.edu.tw

報名連結：



台灣作業研究學會



逢甲大學工業工程與系統管理學系



AIMS 人工智慧製造系統研究中心



亞洲大學大數據研究中心

主辦單位：



□ 主管對現場認知的落差

- PPT製作、KPI達成、報喜不報憂
- Sol: 總經理辦公室、特助出來協調

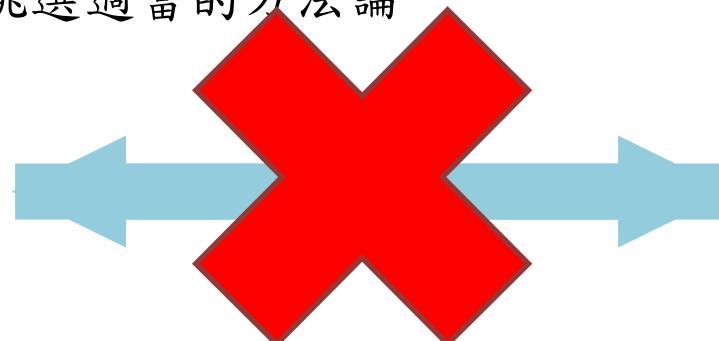
□ 數據、資料與IT基礎建設

- 進去找問題點後，發現沒有數據（現場都是有問題後才開始收集數據）
- Sol: 分階段進行、初步試驗性質的問題診斷

□ “解”問題過程

- 為AI而AI？為自動化而自動化？
- Sol: 應根據現場挑選適當的方法論

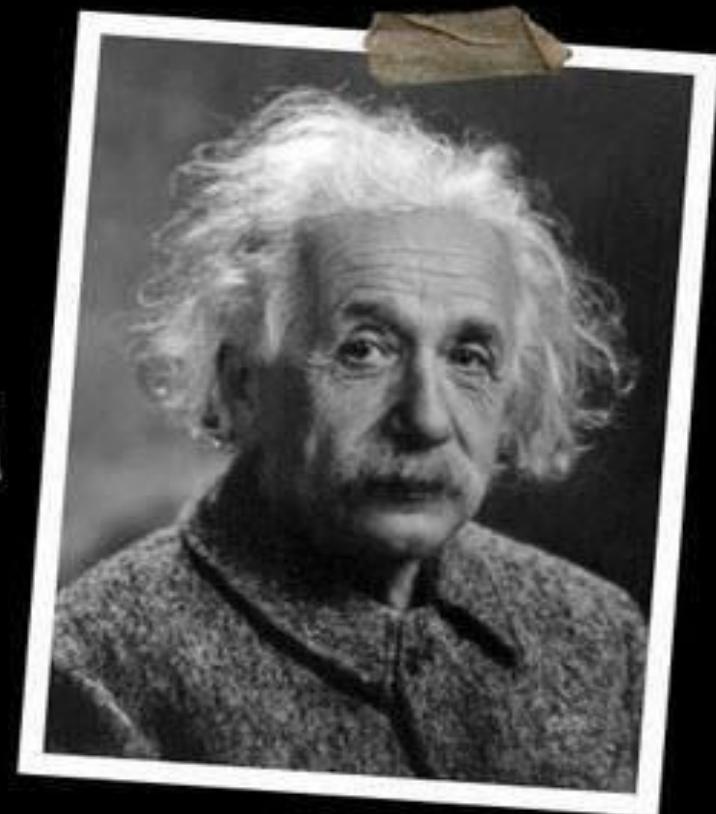
方法導入
-Methodology-based



問題現場
-Experience-based

**"Education is not
the learning of
facts, but the
training of the mind
to think."**

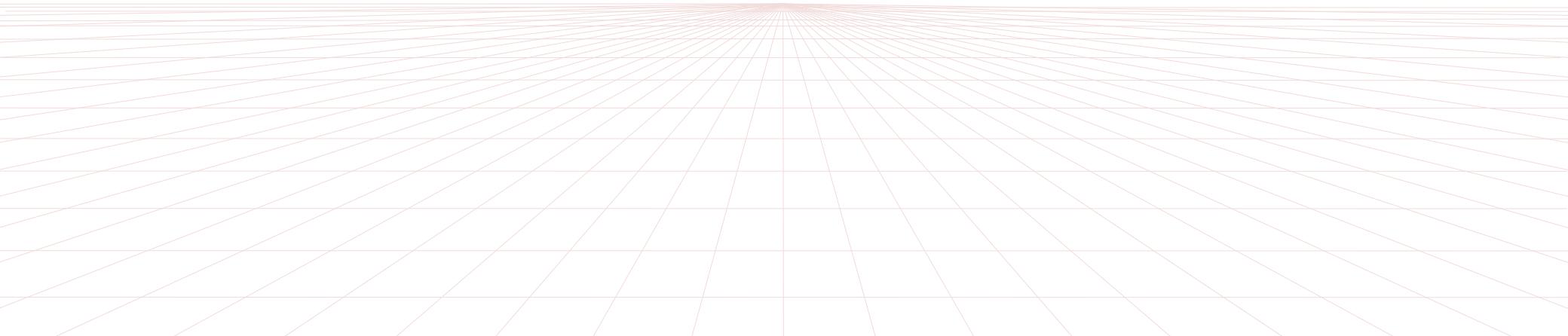
-Albert Einstein



More @ QuotesDump.com

從預測性思維到處方性決策

From Predictive to Prescriptive



預測要愈準愈好？

- 這是當然的..
- 但另一方面，預測的本質是...

- 找出關鍵因子
- 建構因果關係
- 把正或負的影響(趨勢)抓到
 - eg. 線性回歸中的獨立變數的係數
- 協助決策制定，以趨吉避凶



為什麼要預測?



預測的目的是什麼?



下一步為何?

預測像...



Record Japan.net

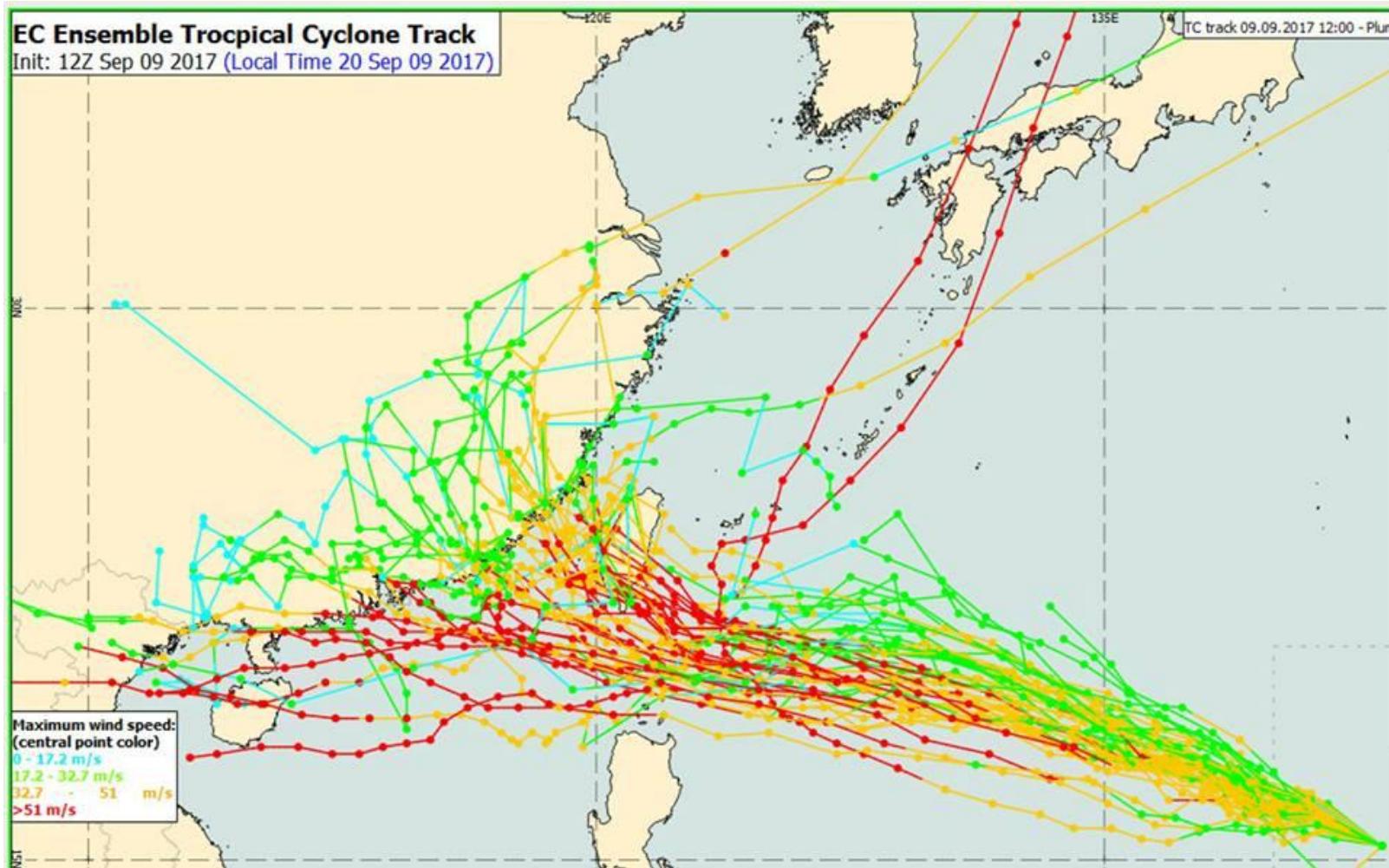


<http://www.zou1.com/news/show-97765.html>

<http://thewhen.pixnet.net/blog/post/42532195-%E5%A1%94%E7%BE%85%E5%8D%A0%E5%8D%9C%3A%E6%88%91%E8%A9%B2%E5%A6%82%E4%BD%95%E5%A2%9E%E5%8A%A0%E8%B2%A1%E9%81%8B%3F>

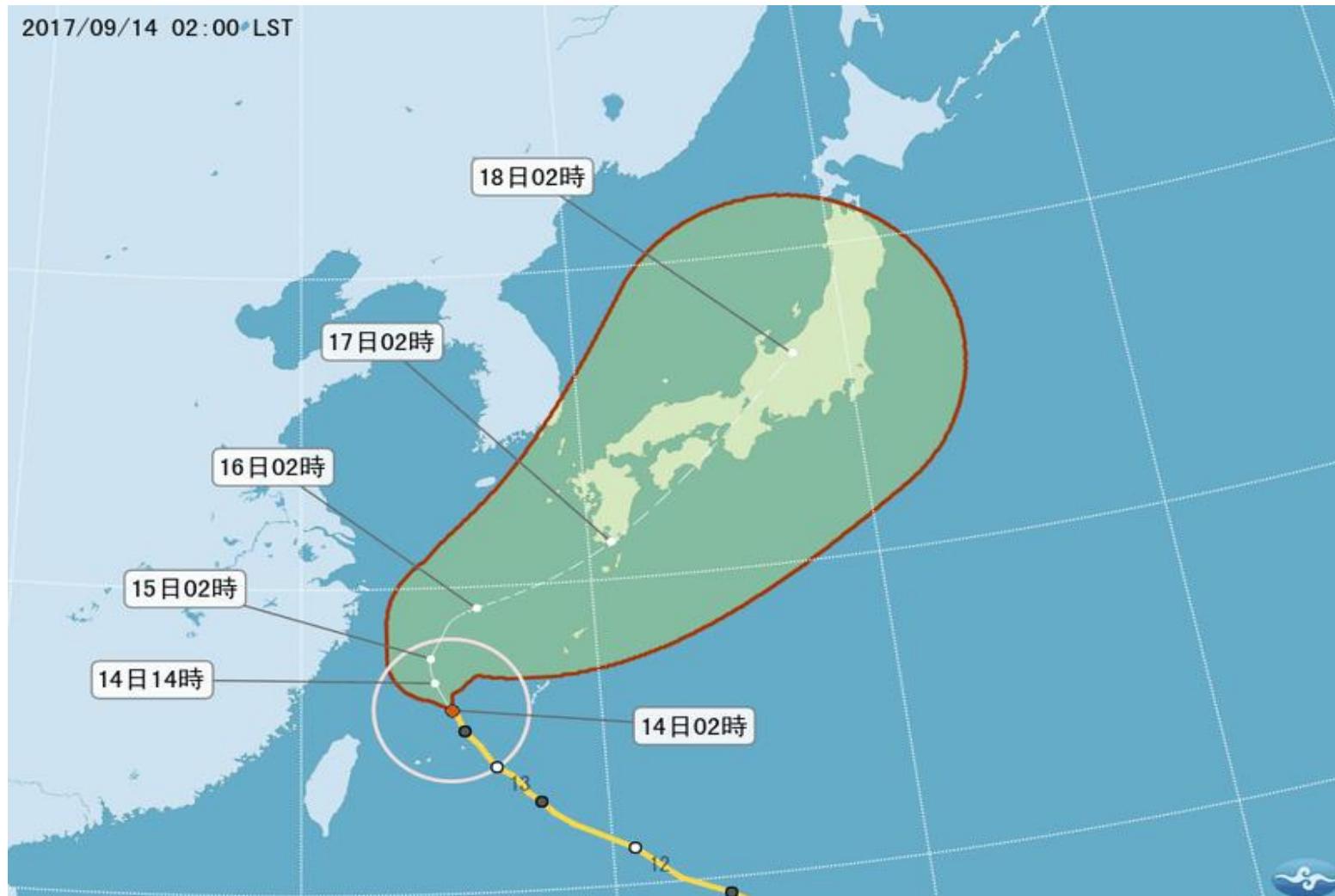
但...其實是...

2017/09/09 歐洲模式中心的五十幾種氣象模式的颱風路徑預測結果，泰利侵台機會頗高。



<http://news.ebc.net.tw/news.php?nid=77898>

2017/09/14 中颱泰利「髮夾彎」 北部、東北部風雨較大



<http://www.cna.com.tw/news/firstnews/201709140012-1.aspx>

所以”決策”很.....難做....

瞭解這就...

不要太為難你的長官 XD

因此...說穿了...

”推論(預測)”就是
透過分析資訊來降低未來不確定性
的一種方法
以利決策的進行!

其實...

重點不在於”預測”本身

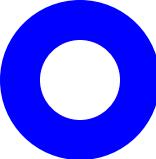
重點在於我們開始”動起來”

透過”收集資訊”來降低所面對的不確定性

- 根據“資訊”收集的狀況，決策的類型可以簡單分成下面三種
- 確定性下決策 (Decision under Certainty)
- 風險下決策 (Decision under Risk)
- 完全不確定下決策 (Decision under Strict Uncertainty)

□ 天氣與決策

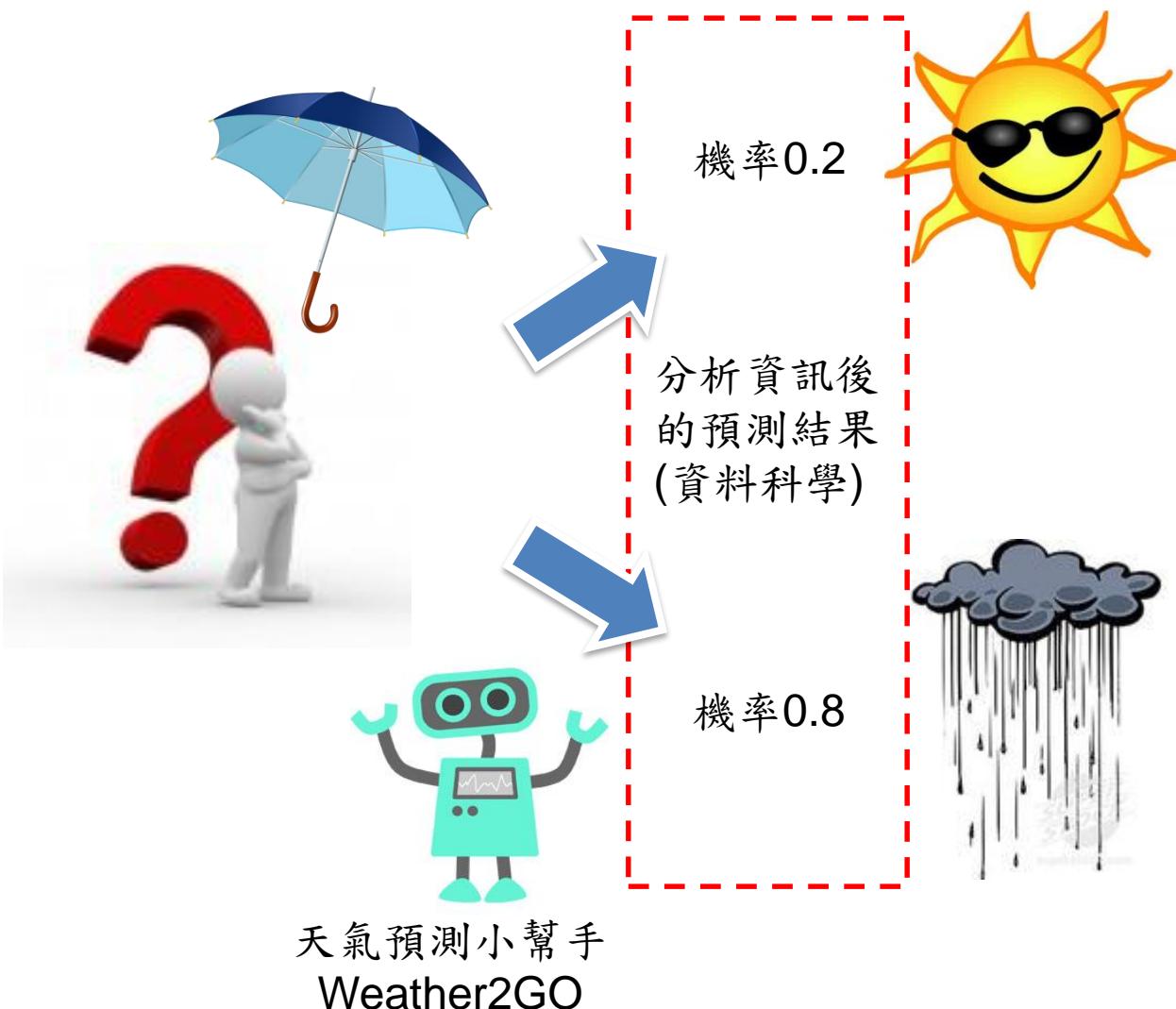


	不帶傘	帶傘
晴天		 1. 笨重 (不方便) 2. 弄丟 (再買一支)
雨天		 1. 落湯雞 (感冒) 2. 搭計程車 (支出) 3. 拿別人的傘 !! (良心不安)

□ 確定性下決策 (Decision under Certainty)



□ 風險下決策 (Decision under Risk)



- 方法一：期望值

$0.2 \times \text{太陽} + 0.8 \times \text{下雨} (!?)$

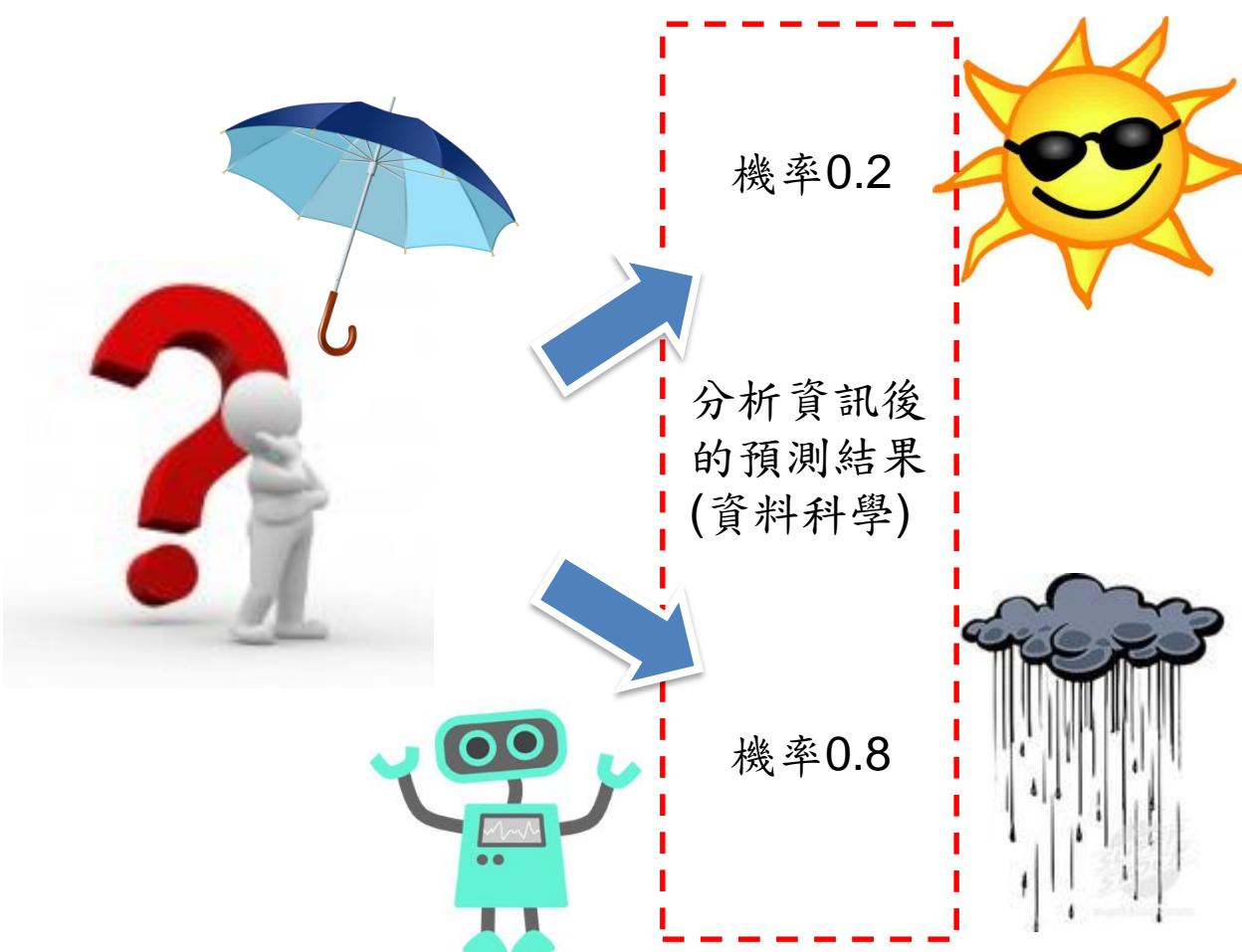
迷思- 殺子六個面

每一面 $1/6$, 期望值為 3.5

賭殺子應該出現 3.5 (!?)

- 方法二：決策風險

□ 風險下決策 (Decision under Risk)



天氣預測小幫手
Weather2GO

- 方法二：決策風險
- 處方性的分析

風險成本矩陣		
	不帶傘	帶傘
晴天	成本 \$0	成本 \$100 (弄丟)
雨天	成本 \$200 (感冒)	成本 \$0
期望成本	0.2×0 $+ 0.8 \times 200$ $= 160$	0.2×100 $+ 0.8 \times 0$ $= 20$
決策：帶傘		

□ 風險下決策 (Decision under Risk)

切記...

不是因降雨機率0.8，所以我們才帶傘，主要是因為**決策風險(成本)**的考量後，我們才決定帶傘。

又或...

我們可以想想這件事，若弄丟後買一把傘**成本為100,000元**，我們還會決定帶傘嗎？

- 方法二：決策風險
 - 處方性的分析

		風險成本矩陣	
		不帶傘	帶傘
晴天	成本 \$0	成本 \$100 (弄丟)	
	雨天	成本 \$200 (感冒)	
期望成本	0.2x0 +0.8x200 =160	0.2x100 +0.8x0 =20	
決策：帶傘			

由此可知...

預測(降雨機率)很重要，
決策風險(帶不帶傘的後果)也很重要

互為表裡、相輔相成

Lee (2019)

□ 完全不確定下決策 (Decision under Strict Uncertainty)



方案1：沒有帶傘的習慣耶

方案2：就每天帶把小傘吧~

Based on your
Preference
Structure

□ 公司規劃投資產能擴充，希望明年銷量愈高愈好

- 目前公司年度營業利益約25億元
- 根據明年景氣狀況，營業利益預估如下

方案	明年景氣變化情境		
	景氣好	景氣持平	景氣差
併購公司A	42	21	8
購置機台	35	28	13
外包	28	25	23

你會選哪一個方案？

Revised from Chien (2005)

□ 悲觀準則 (The Maximin Payoff Criterion)

- 小中取大報酬準則
- 作最壞的打算，再由各最壞的報酬中選擇最好的結果

方案	天氣情境			Min Return
	景氣好	景氣持平	景氣差	
併購公司A	42	21	8	8
購置機台	35	28	13	13
外包	28	25	23	23

□ 樂觀準則 (The Maximax Payoff Criterion)

- 又稱大中取大報酬準則
- 作最好的打算，再由各最好的報酬中選擇最好的結果

方案	天氣情境			Max Return
	景氣好	景氣持平	景氣差	
併購公司A	42	21	8	42
購置機台	35	28	13	35
外包	28	25	23	28

□ 薩維基準則 (The Minimax Regret Criterion) (Savage, 1951)

- 又稱最小最大悔惜準則
- 悔惜定義為「一決策方案之報酬與相同自然狀態下可以產生最大報酬之決策方案，兩者報酬的差額」
- Step 1：建立悔惜償付表
- Step 2：在悔惜償付表中採大中取小準則(minimax)

		天氣情境			悔惜矩陣			
方案		景氣好	景氣持平	景氣差	景氣好	景氣持平	景氣差	Max Regret
併購公司A	42	21	8	0	7	15	15	15
	35	28	13	7	0	10	10	10
	28	25	23	14	3	0	14	14

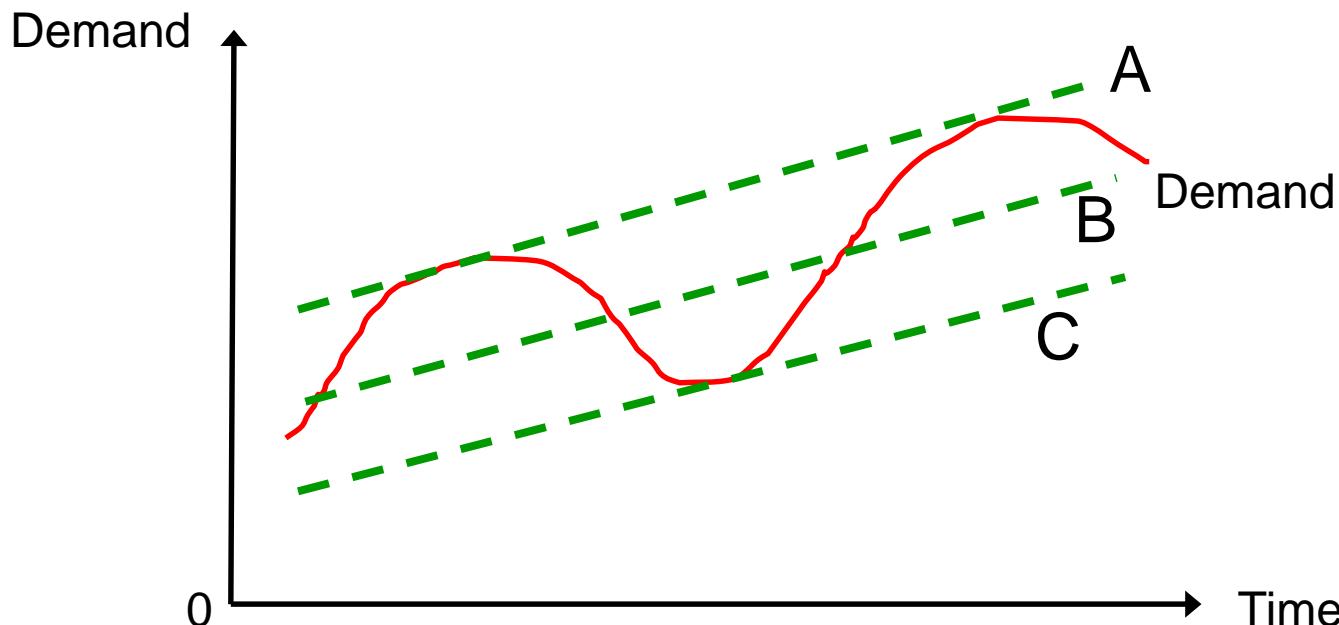
□ 公司規劃投資產能擴充，希望明年銷量愈高愈好

- 無資訊情況下，理性決策並沒有對錯的差別
- 只是根據每個人的”喜好結構 (preference structure)”所呈現決策結果

方案	天氣情境		
	景氣好	景氣持平	景氣差
併購公司A	42	21	8
購置機台	35	28	13
外包	28	25	23

你會選哪一個方案？

□ Capacity Planning Problem (Newsboy problem over time)



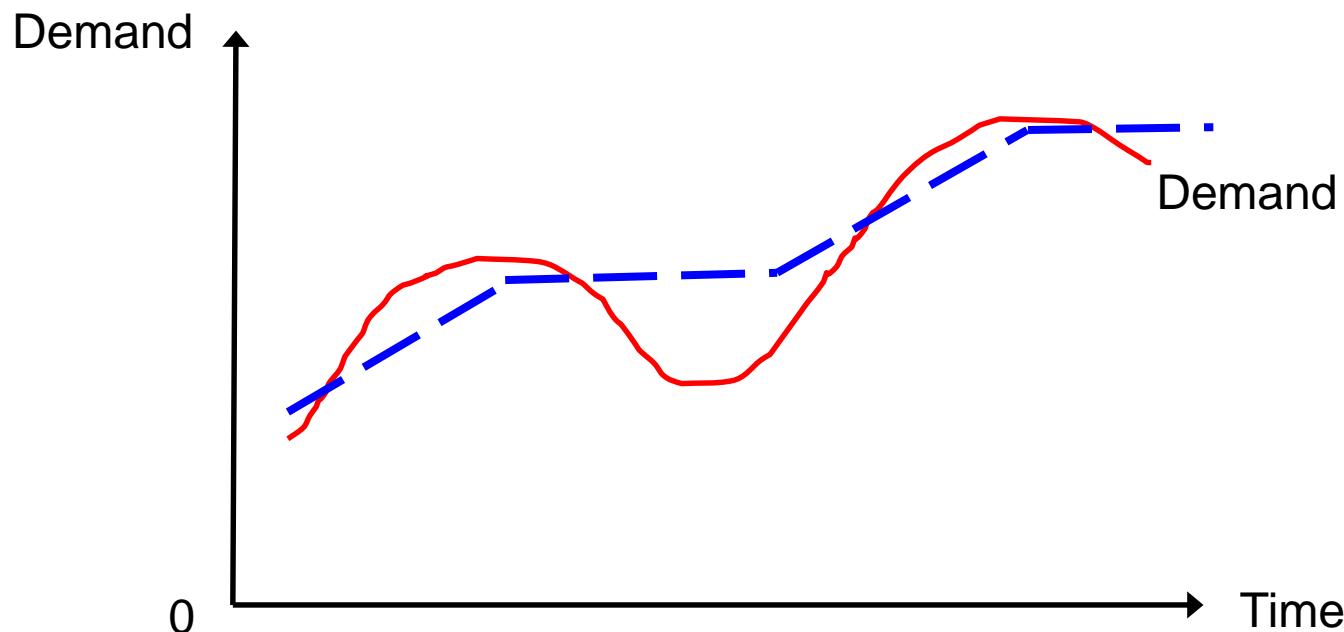
□ Capacity Plan: A, B, C or D (demand)?

□ Two Types of Risk

- Capacity Shortage 產能短缺: loss of sales, loss of market share
- Capacity Surplus 產能過剩: machine idleness, inventory, holding cost

□ Capacity Planning Problem (Newsboy problem over time)

- 其實，產能規劃經常是...



追高殺低!!??

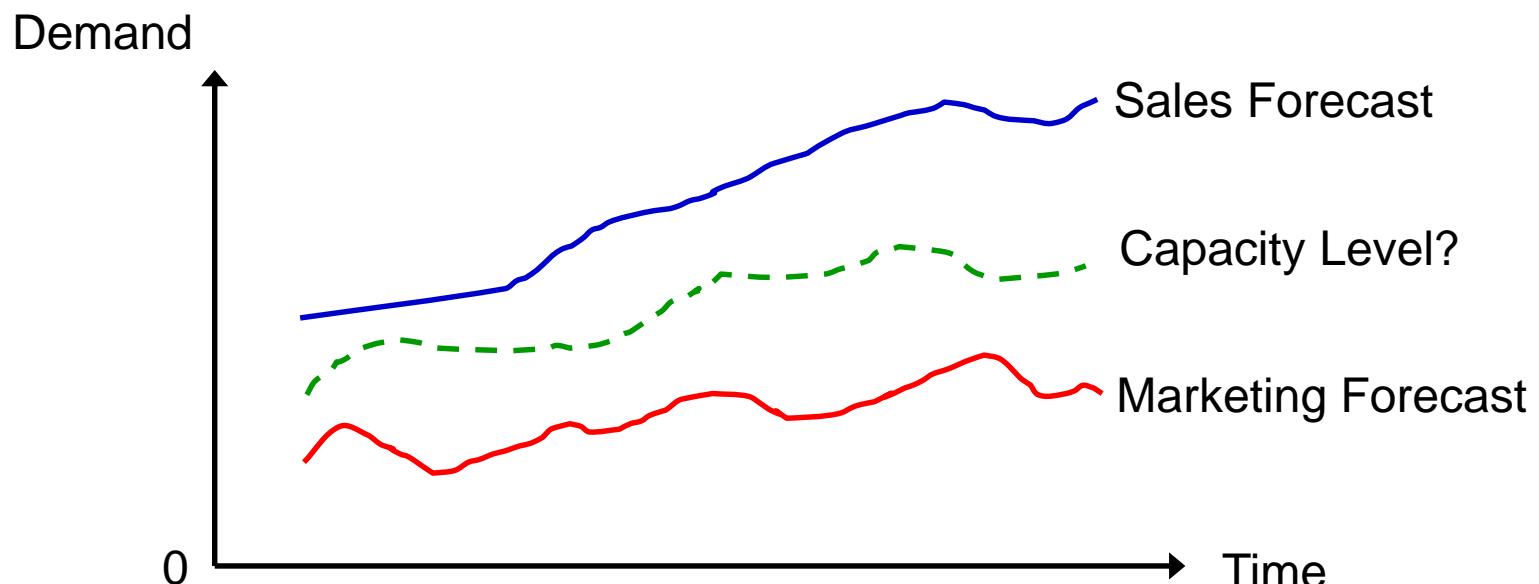
□ Newsboy problem 報童問題 (典型的作業研究OR問題)

- 每天一大早出門到批發商訂報紙，送報賺錢
- 當日買報紙的需求是不確定的
- 報紙若當天沒賣出去，則殘值為0 (由於時效性，隔天就不能賣了)
- 決策議題：要訂多少報紙呢？



□ Capacity Planning Problem (more demand scenarios)

- How about **more than two** demand scenarios?
- Two demand forecasts: **Marketing vs. Sales**
 - Should capacity level follow marketing, or sales, or ...?



- Scenario Simulator

- Generate capacity plans (>10 versions) **manually**...
 - KPIs: **CapEX per k, ROIC, payback...**

□ Empirical Study- Taiwan TFT-LCD Manufacturer

- We introduce a **two-phase framework** to decide the best capacity level reducing the losses (i.e., **risks**) of **capacity surplus** and **capacity shortage**.

□ 1st stage: Demand Forecast (預測模型)

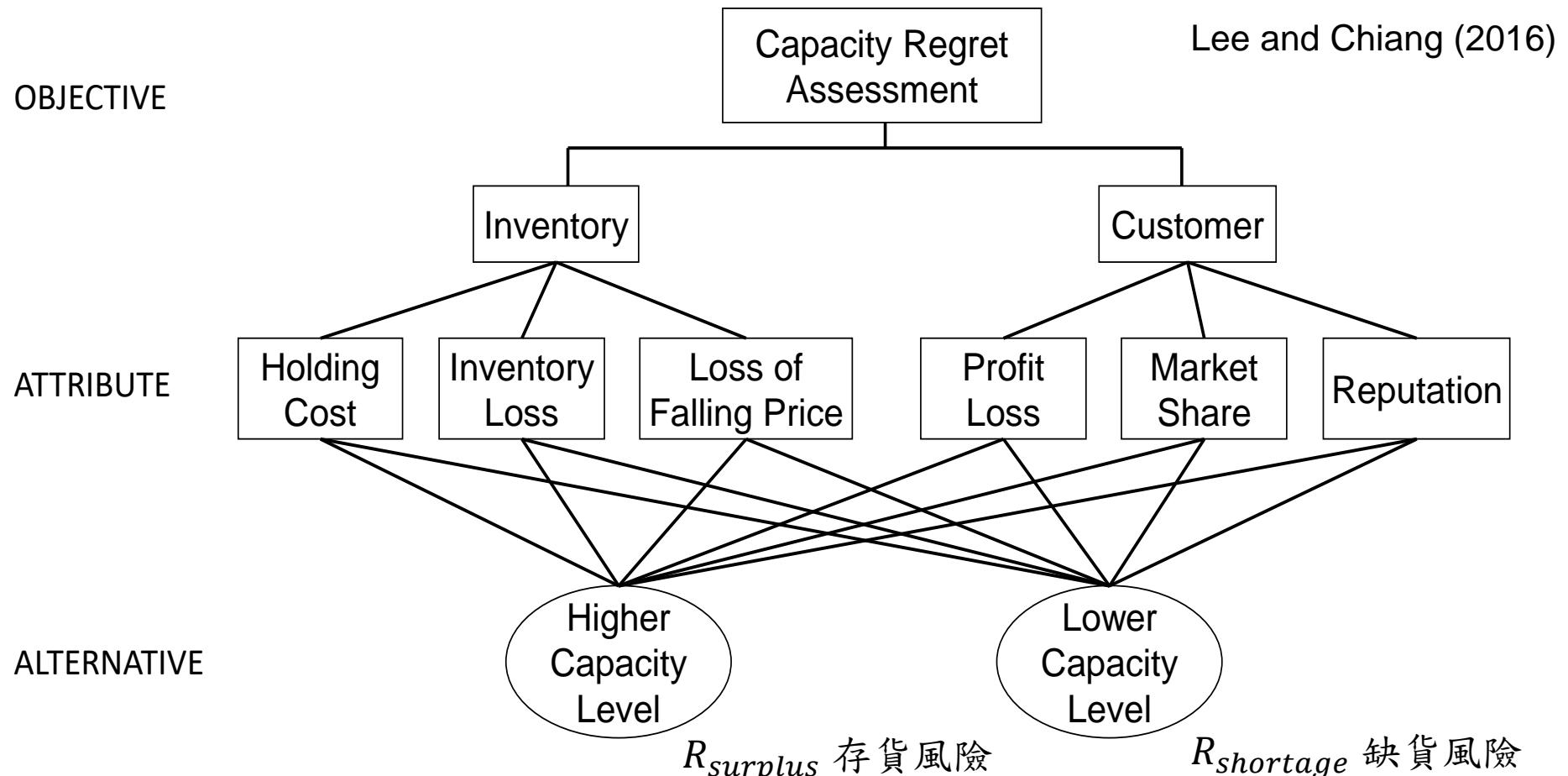
- Linear Regression
- Autoregression
- Neural Network

□ 2nd stage: Capacity Decision (最佳化模型)

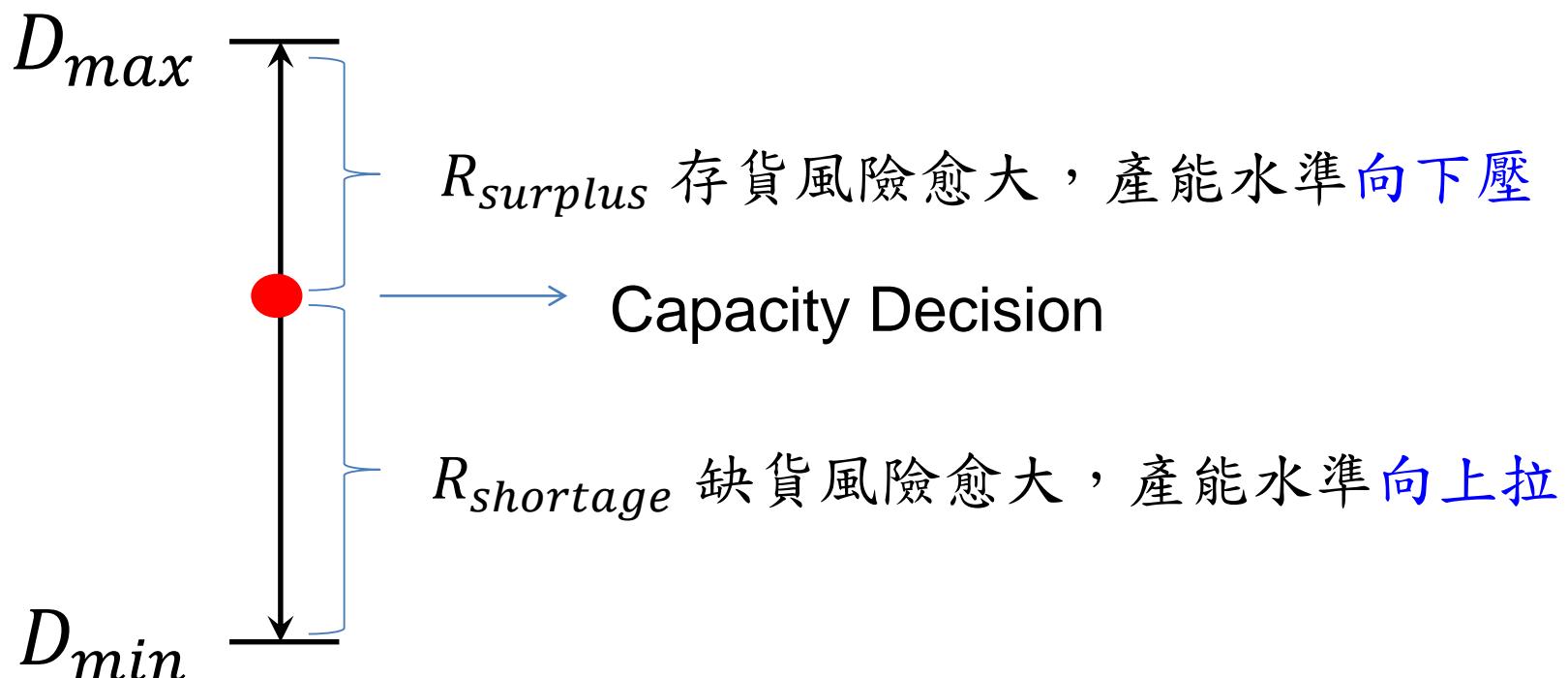
- Expected Value (EV, 期望值)
- Minimax Regret (MMR, 薩維基準則)
- Stochastic Programming (SP, 隨機規劃)

□ Capacity Regret Assessment

- 透過問卷調查以系統化的方法來收集決策者的風險偏好
- Analytic Hierarchical Process (分析層級程序法, AHP) (Satty, 1980)

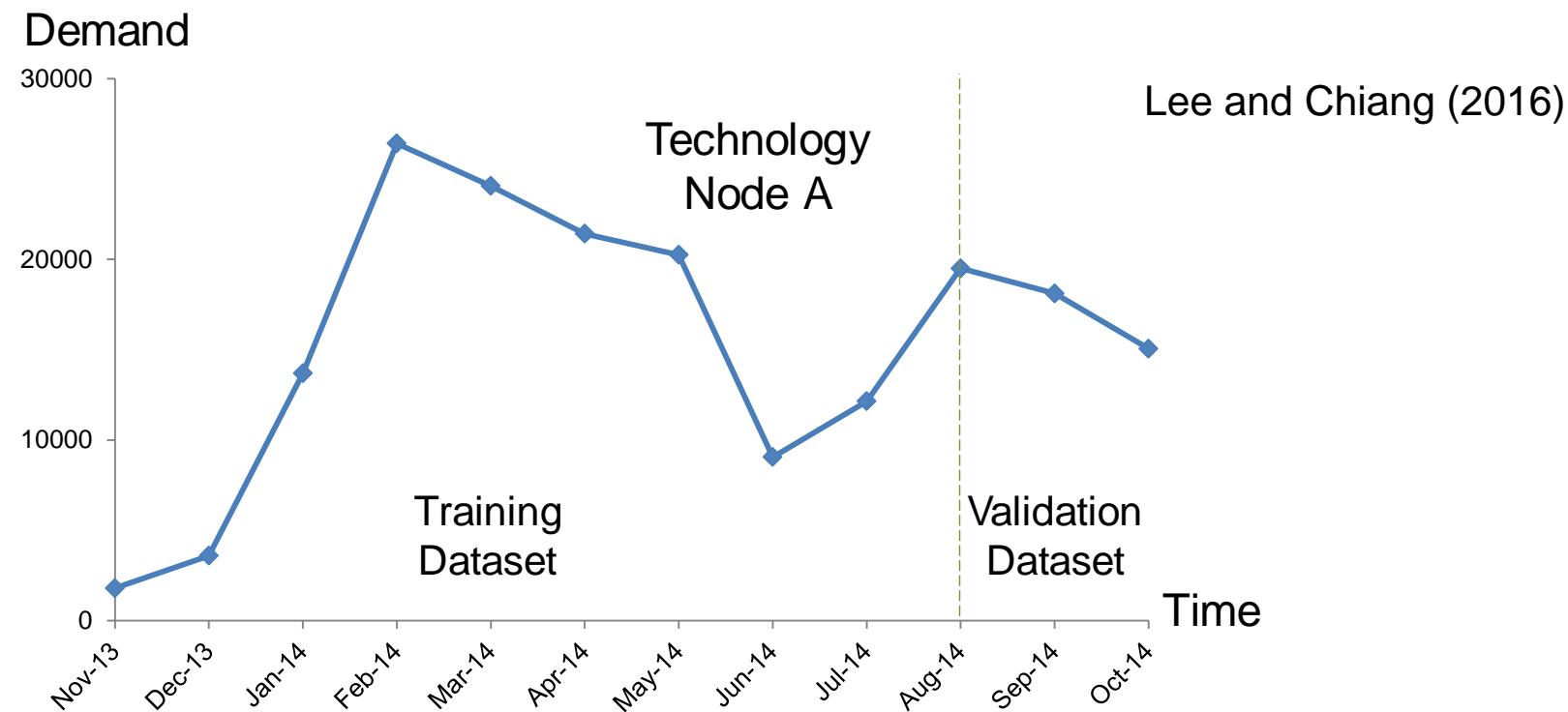


□ A Tradeoff Between Two Risks



□ Data Source

- Nov. 2013 to Oct. 2014
- Aggregate demand for a specific **technology node**
- Dataset: training data (9 months) and validation data (3 months)
- Objective: determine the **capacity level** for the future 3 months (Aug. 2014 to Oct. 2014) based on historical demand until July 2014.

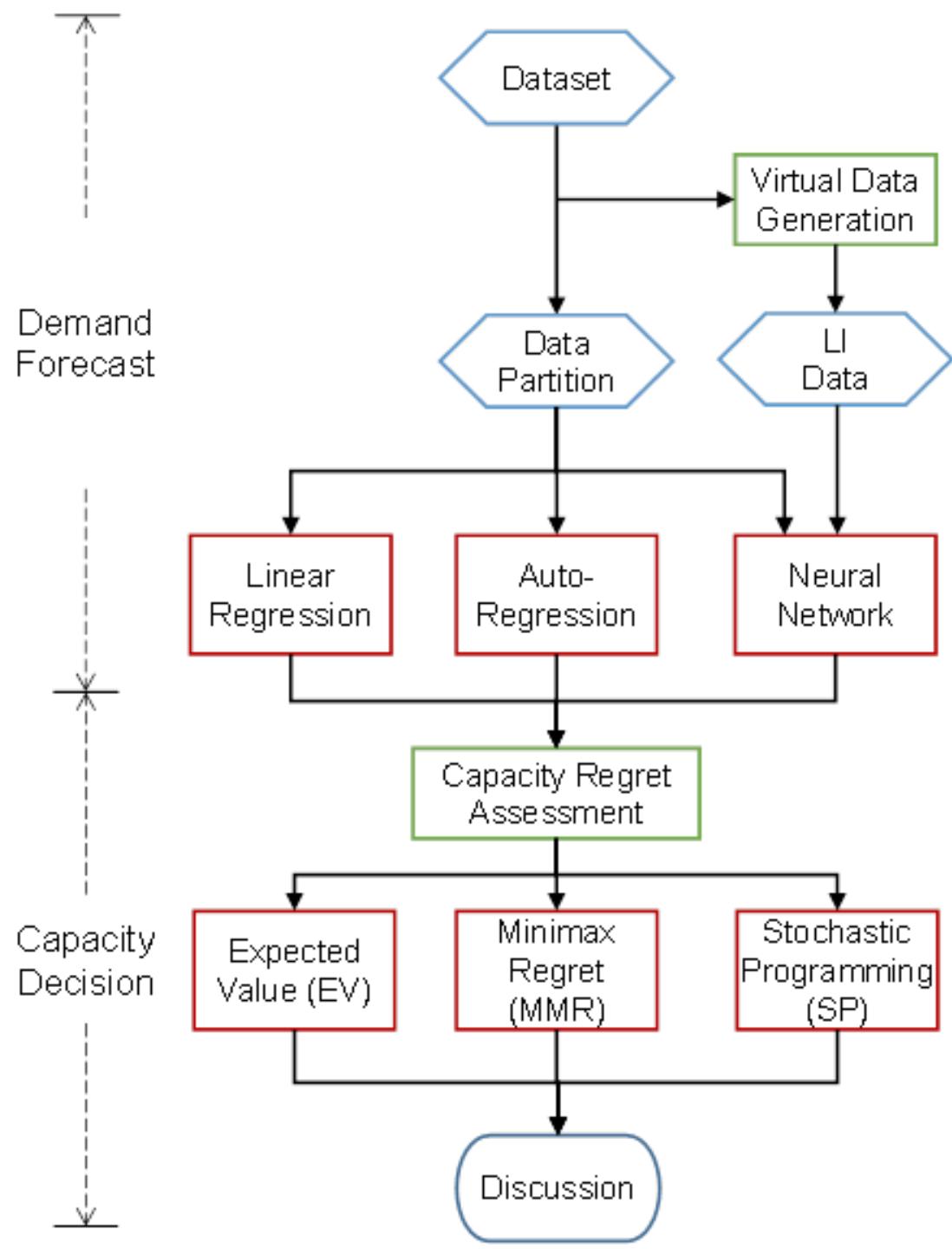


Empirical Study

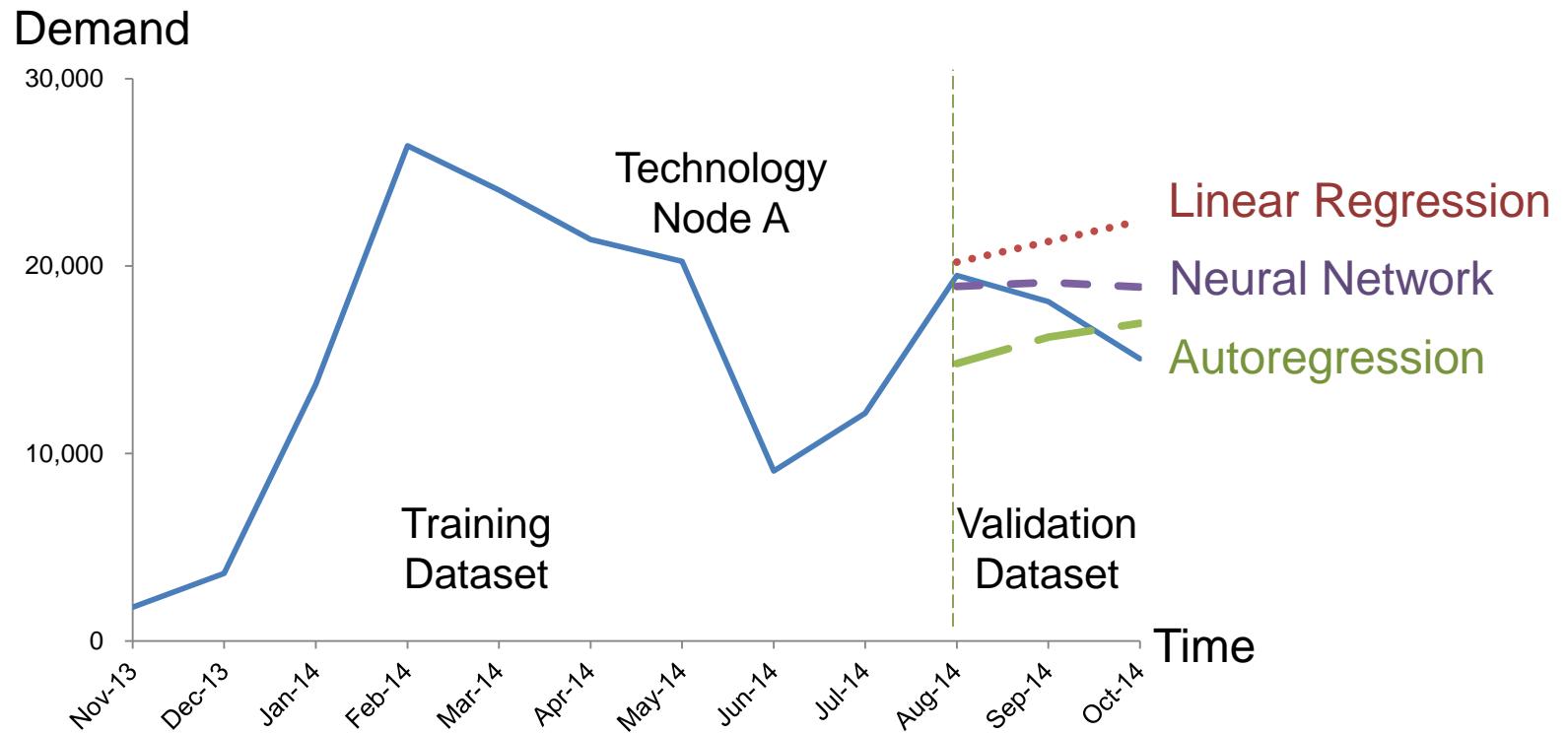
- 2-stage approach
 - Demand Forecast
 - Capacity Decision

- VDGP
 - LI: latent information
(Chang et al., 2014)

Lee and Chiang (2016)



- 1st phase: Demand Forecast
 - linear regression (OLS), autoregression (AR), neural network (NN)
 - Criterion: Mean Squared Error (MSE): NN < AR < OLS



□ 2nd phase: Capacity Decision

- AHP gives **pair-wise comparisons** of the capacity surplus and shortage, and calculate the **eigenvector** (特徵向量) as weights of two risks.
- 31 individuals with TFT-LCD industrial backgrounds
 - 2 for 0-2 yrs, 9 for 2-5 yrs, 13 for 6-9 yrs, 7 for more than 10 yrs of service
 - one third have undergraduate degrees and the others for graduate degrees
- $R_{shortage} = 0.521$ and $R_{surplus} = 0.479$.

Month	Actual Demand	Capacity Decision			
		Current Policy	EV	MMR	SP
Aug. 2014	19499	20207	16941	17157	18922
Sep. 2014	18110	21305	17921	18306	19135
Oct. 2014	15059	22404	18536	19291	18885
Standard Deviation	2271.3	1098.2	804.7	1068.1	134.8
Average Regret	-	1796.0	1032.2	1113.8	874.8

*Current policy builds up the capacity level based on the maximal forecast demand.

**Avg. Reg.: the weighted sum of squared error (WSSE) over the number of validation periods

- 隨機規劃(stochastic programming, SP)提供一穩健決策(robust decision)

數學規劃目標與限制

$$\text{Min } Cx + E[f(x, \tilde{D})]$$

$$\text{s.t. } x \geq 0$$

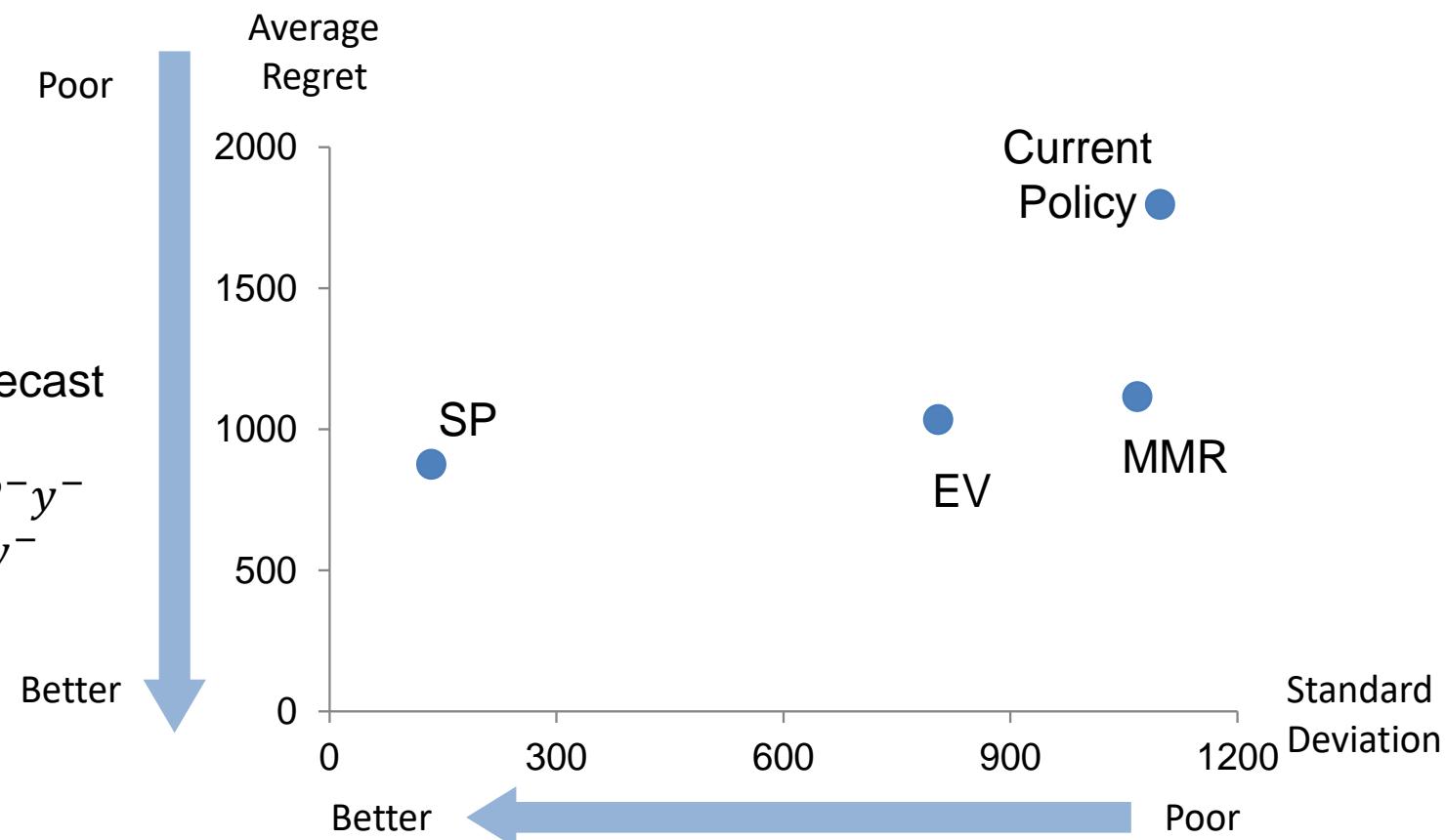
where for a demand forecast

$$d \in \tilde{D}$$

$$f(x, d) = \text{Min } P^+y^+ + P^-y^-$$

$$\text{s.t. } d - x = y^+ - y^-$$

$$y^+, y^- \geq 0$$



Lee, Chia-Yen & Ming-Chien Chiang (2016). Aggregate Demand Forecast with Small Data and Robust Capacity Decision in TFT-LCD Manufacturing. Computers & Industrial Engineering, 99, 415-422.

- 我們可以透過模型來預測或模擬許多情境(scenarios)
 - BPNN, SVM, CART, Boosting, PLS,
- Then...Which one is correct? What is the next step
- 預測...既然不可能100%準，那決策上的損失風險，就應該納入考量
- 讓我們想個簡單例子：
 - 模型A準確度95%，但不準時的決策會造成"龐大損失"
 - 模型B準確度90%，但不準時的決策只會造成"些許損失"
 - 我們要選模型 A還是模型B呢？
- 預測性的思維 → 處方性的決策

□ 精度/良率預測

- Classification → 探討Confusion Matrix中的trade-off → Cost-sensitive
- 權衡誤放(type-I)與誤宰(type II)風險的成本 → 處方性分析

Model A		預測	
		FAIL	PASS
實際	FAIL	61	7
	PASS	29	31

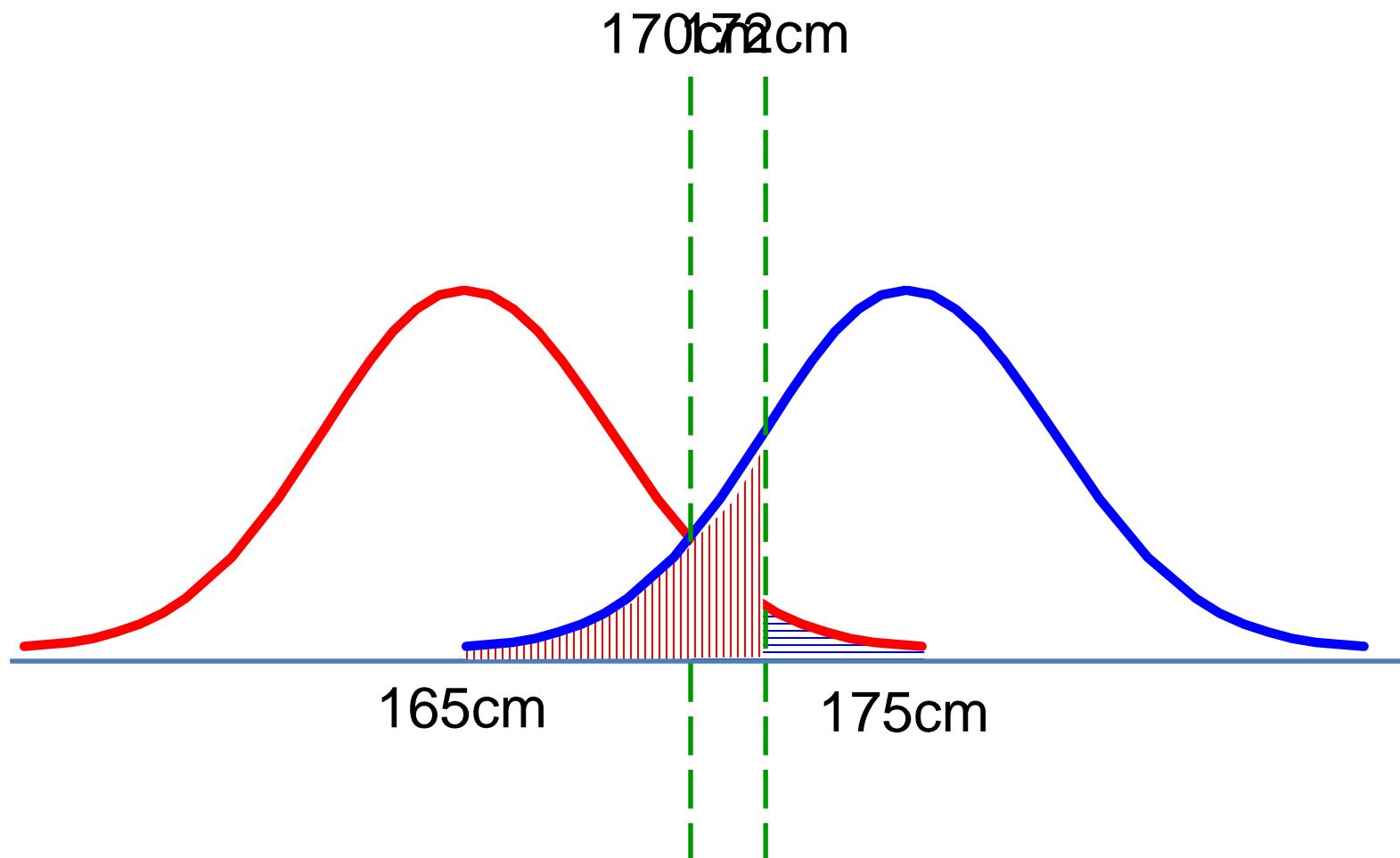
Model B		預測	
		FAIL	PASS
實際	FAIL	47	21
	PASS	7	53

		Testing	
		Accuracy	AUC
Model A		71.9%	70.2%
		78.1%	78.9%

AUC: Area under the Curve of ROC

Lee (2019)

可能同時降低型一與型二誤差？



讓我們整理一下.....

- 為了降低對未來的不確定性 (建構關鍵因子推論的預測模型)
 - 資料科學其實並非單純用數學模型~ (即使有時候讓我們方便許多)
 - 記得 Manufacturing → “Hand Make” → 黑手~
 - 站起來，走到現場去。透過觀察與討論，找出“推論用的關鍵因子”
 - 切記，一旦收集到“資料”，便已成歷史
 - 如何透過歷史來探索未來 (鑑往知來)，資料科學扮演關鍵角色。

- 為了提升決策品質 (建構風險權衡最佳化的決策模型)
 - 由於預測不一定準，且常有多個預測情境(scenarios)
 - 決策風險的量測是必要的
 - 站起來，走到現場去。透過觀察與討論，找出“權衡用的風險因子”
 - 切記，“魚與熊掌不可兼得”，因此需要權衡得失 (風險)
 - 如何透過有限資源來規劃未來 (見識謀斷)，作業研究扮演關鍵角色

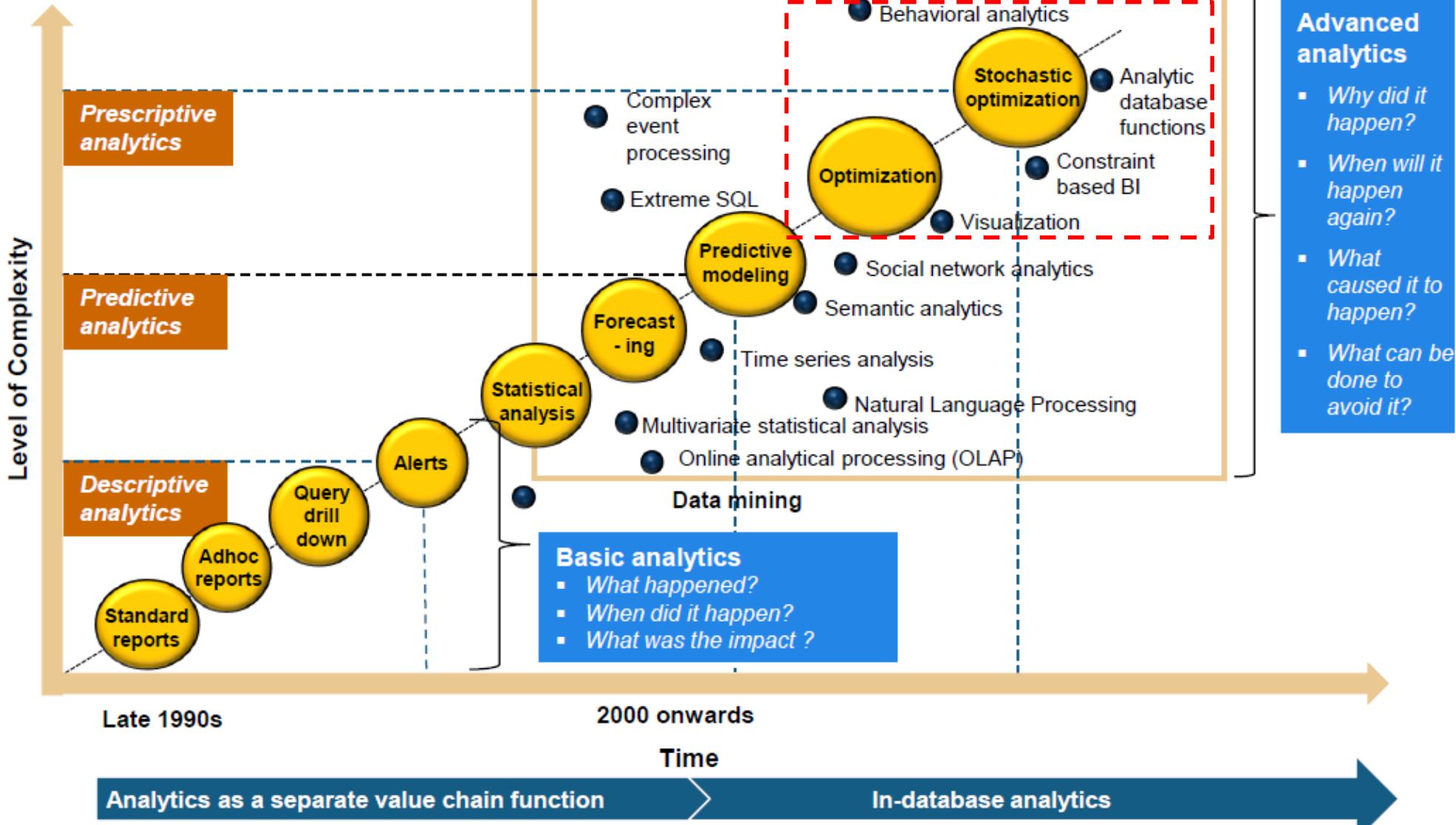
資料科學與大數據分析是萬靈藥(panacea)?

方法複雜度與附加價值

Operations Research 作業研究 (多角化、投資組合、排程、產能規劃)

POLab
(製程監控)

Evolution of analytics



Source: CRISIL GR&A analysis

□ Hillier and Lieberman (2010)

- “Analytics is the scientific process of transforming data into insight for making better **decisions**”

□ Descriptive analytics (敘述性分析)

- Using innovative techniques to locate the relevant data and identify the interesting patterns in order to better **describe and understand** what is going on now.

□ Predictive analytics (預測性分析)

- Using the data to predict what will happen in the **future**.

□ Prescriptive analytics (處方性分析)

- Using the data to prescribe what should be done in the future. The powerful optimization techniques of **operations research** are what are used here.

什麼是處方性分析？

- The Wyndor Glass Co. products high-quality glass products, including windows and glass doors. (Hillier and Lieberman, 2010)



- It has three plants.
 - Plant 1 makes aluminum frames and hardware
 - Plant 2 makes wood frames
 - Plant 3 produces the glass and assembles the products

- Top management has decided to revamp the product line.

- Production capacity are released to launch two new products having large sales potential:
 - Product 1: An 8-foot glass door with aluminum framing
 - Product 2: A 4×6 foot double-hung wood-framed window



http://img.archieexpo.com/images_ae/photo-g/55456-4334799.jpg
<https://www.andersenwindows.com/>

什麼是處方性分析？

- This problem can be recognized as LP problem of the classic product mix (產品組合) type.

- Linear Programming (線性規劃, LP)

- Formulation as a Linear Programming Problem
 - x_1 : number of batches of product 1 produced per week
 - x_2 : number of batches of product 2 produced per week
 - Z : total profit per week from producing these two products

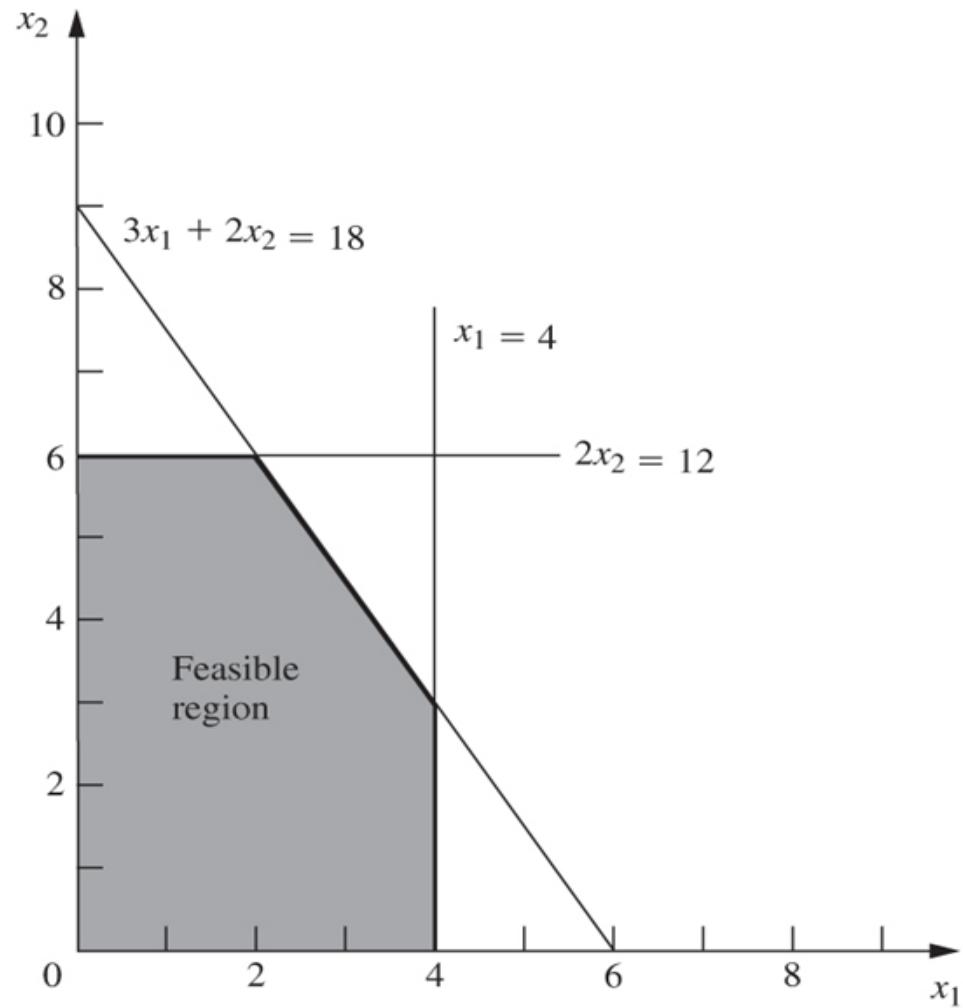
Plant	Production Time		Available Time
	Product 1	Product 2	
1	1	0	4
2	0	2	12
3	3	2	18
Profit (\$1,000)	\$3	\$5	

$$\begin{aligned}
 Z &= 3x_1 + 5x_2 \\
 \text{s.t.} \quad x_1 &\leq 4 \\
 2x_2 &\leq 12 \\
 3x_1 + 2x_2 &\leq 18 \\
 x_1, x_2 &\geq 0
 \end{aligned}$$

什麼是處方性分析?

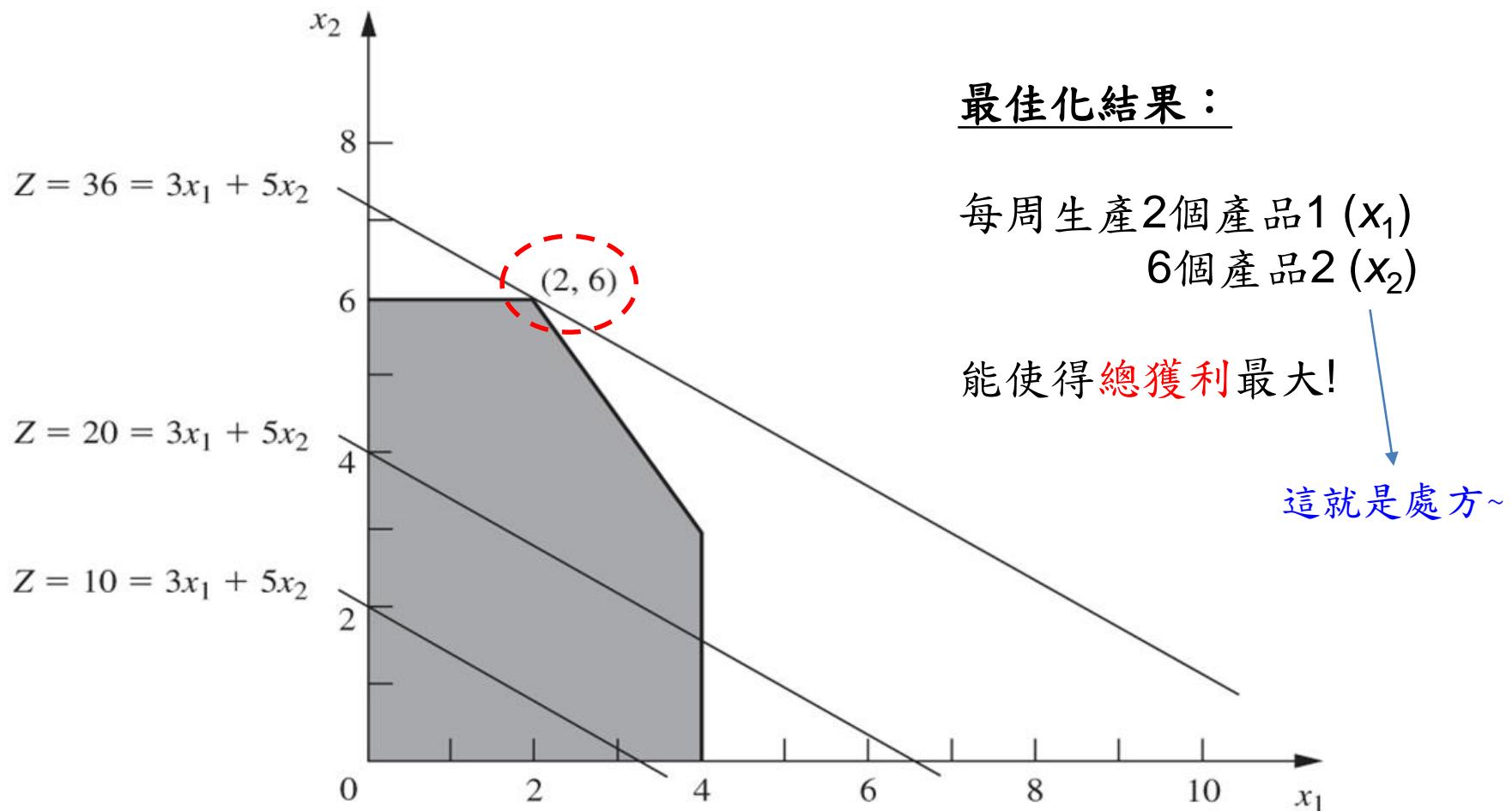
- Graphical Solution
- The resulting region of permissible values of (x_1, x_2) , called the feasible region.

$$\begin{aligned} Z &= 3x_1 + 5x_2 \\ \text{s.t. } x_1 &\leq 4 \\ 2x_2 &\leq 12 \\ 3x_1 + 2x_2 &\leq 18 \\ x_1, x_2 &\geq 0 \end{aligned}$$



什麼是處方性分析?

- Graphical Solution
- Move the objective function with fixed slope through the feasible region in the direction of improving Z .



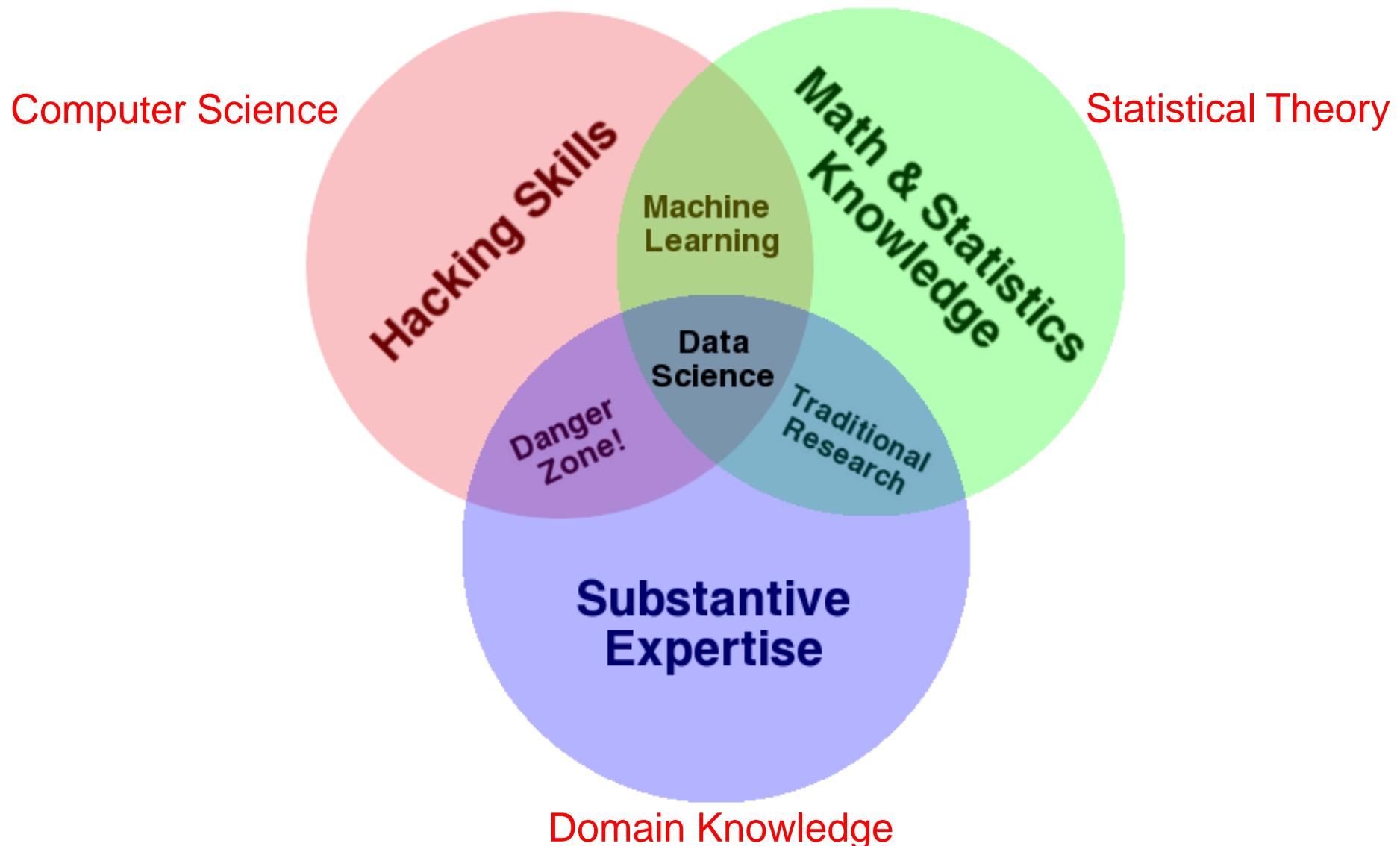
遇到問題

給定資源限制條件（限制式）

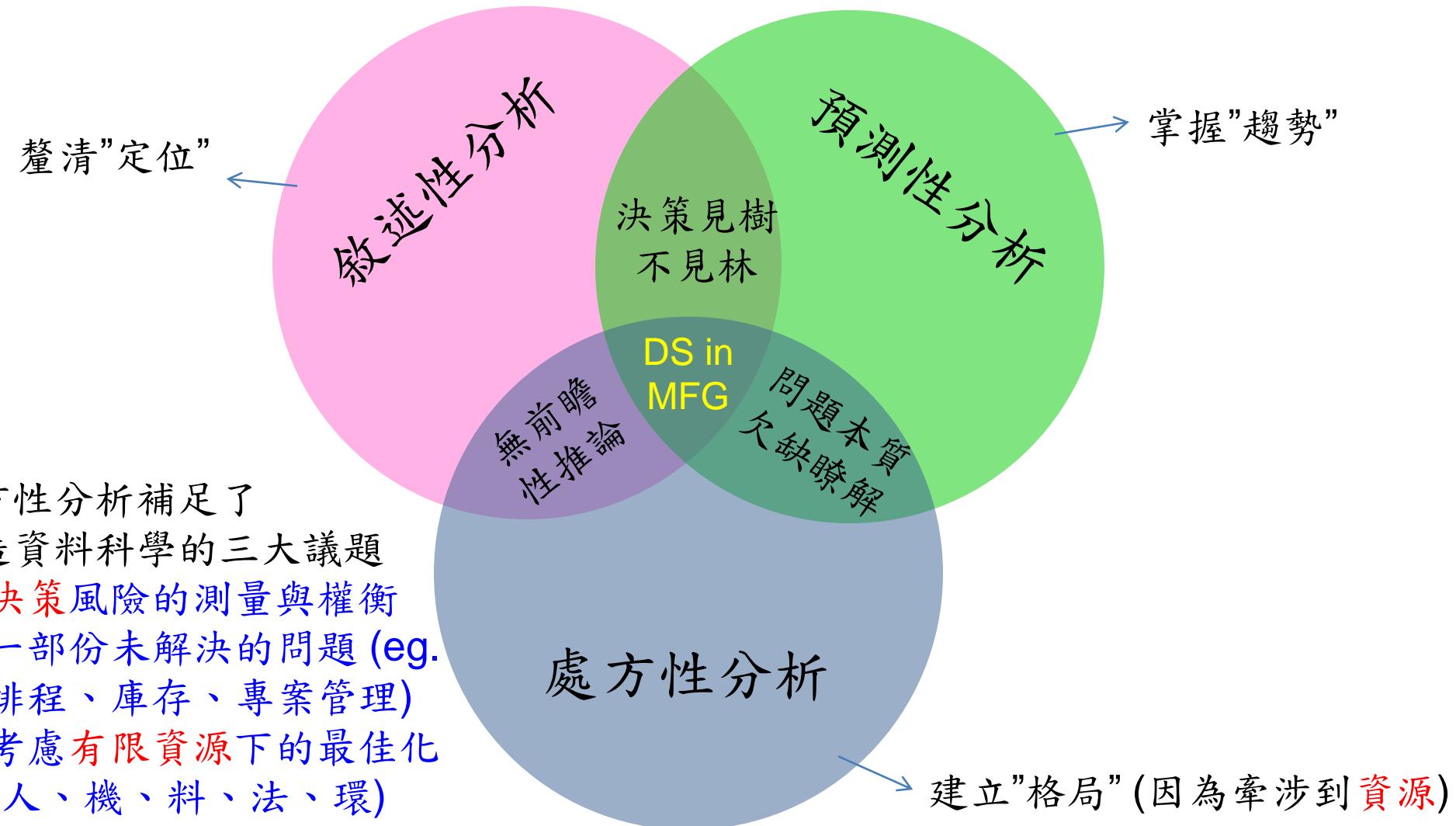
給定期望改善的KPI（目標函數）

提供solution

這就是一種處方

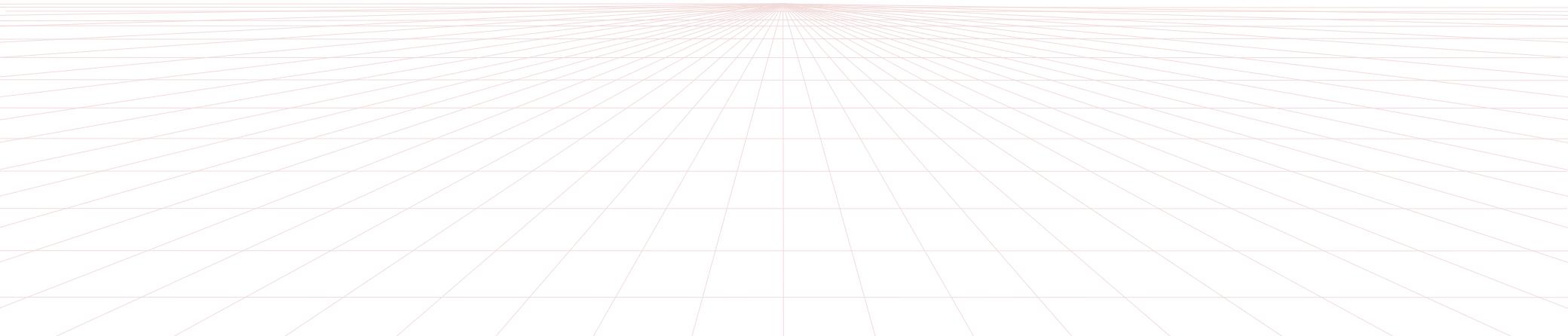


事實上，製造資料科學是...



結語

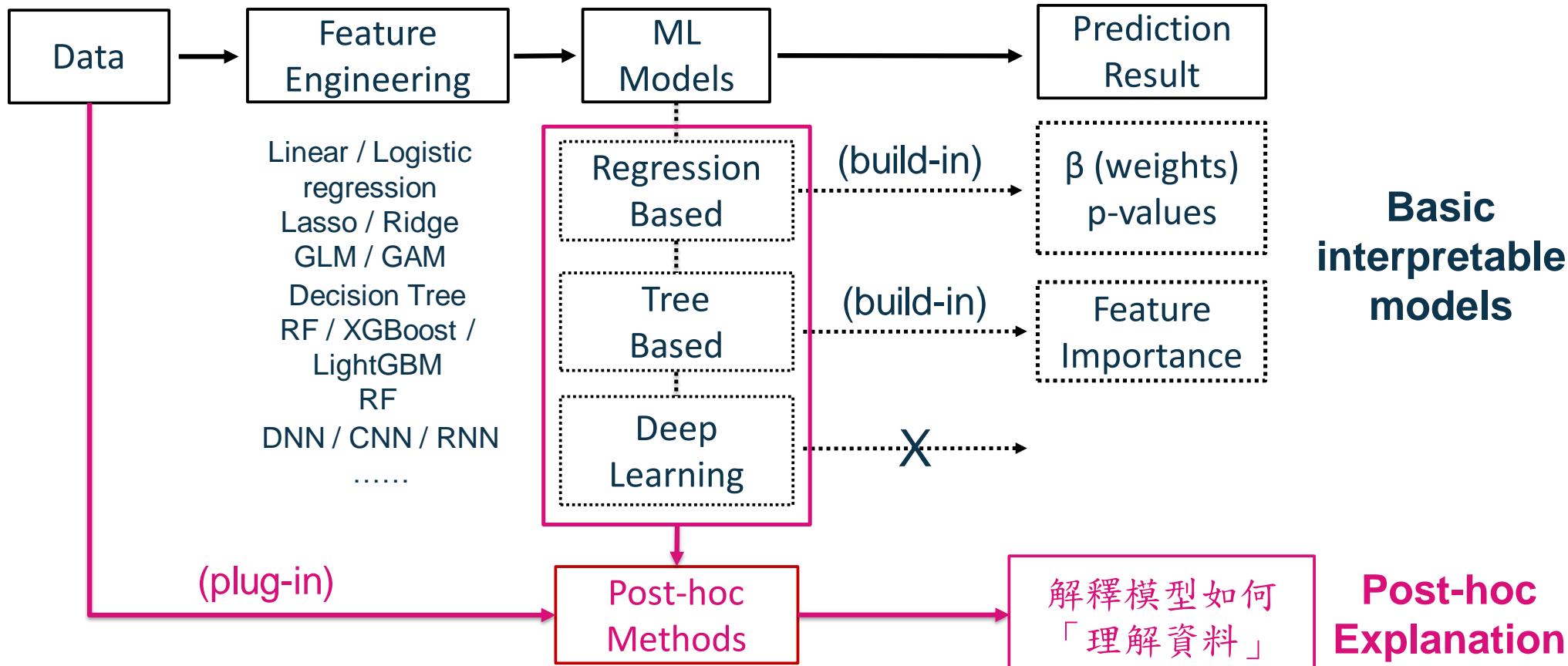
Conclusion



Big Data 資料量愈多愈好?
找到的Pattern也是愈多愈好?

- 對，但也不完全對
- 統計是以樣本推論母體，因此樣本愈多，對母體推論愈精確
- 「證據愈多，說話愈大聲」
- 並非單純資料量或資訊量的多寡
- 重要的是資訊的「異質性(Heterogeneity)」
- 猶太人的第十人理論
- 如何得到異質性資訊!?
 - 找一些想法或背景不一樣的”人”加入團隊
 - Sensor不是裝在自家、競爭者、客戶，而是End User身上

□ AI is black box? Explainable AI (XAI)?



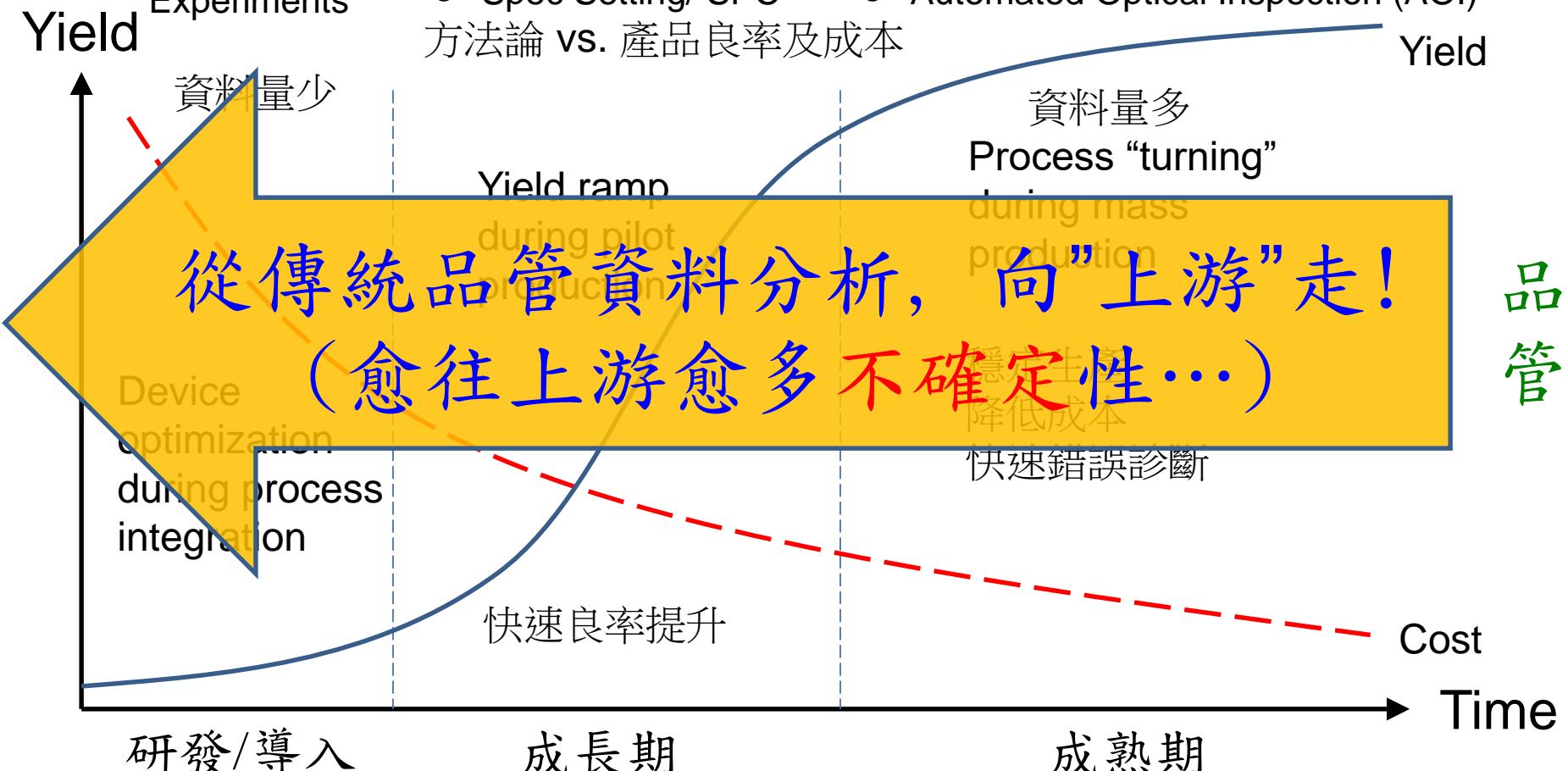
洪紹嚴, 2019. XAI (Explainable AI)-Introduction.

<https://speakerdeck.com/skydome20/xai-explainable-ai>

洪紹嚴 (2019)

跨越問題與方法的鴻溝

- Engineering Parameter Optimization
- Design of Experiments
- Tool Matching
- Benchmarking
- Troubleshooting
- Scheduling
- Spec Setting/ SPC
- Prognostics & Health Monitoring (PHM)
- Virtual Metrology (VM)
- Run-to-Run (R2R) Control/ FDC
- Preventive/Predictive Maintenance(PM/PdM)
- Automated Optical Inspection (AOI)

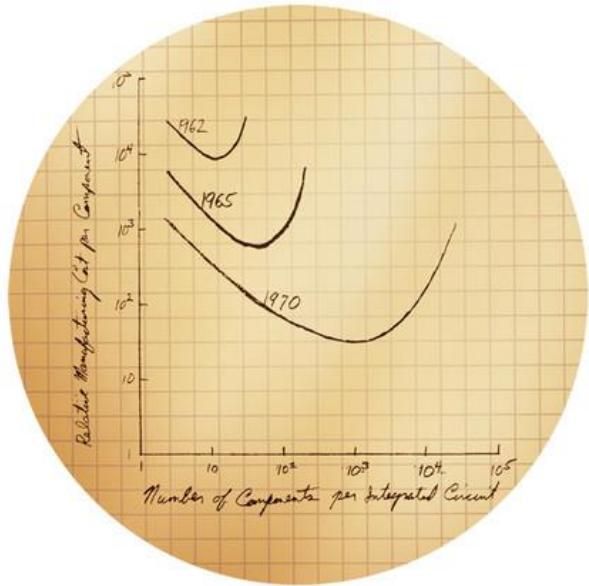


Lee (2019)

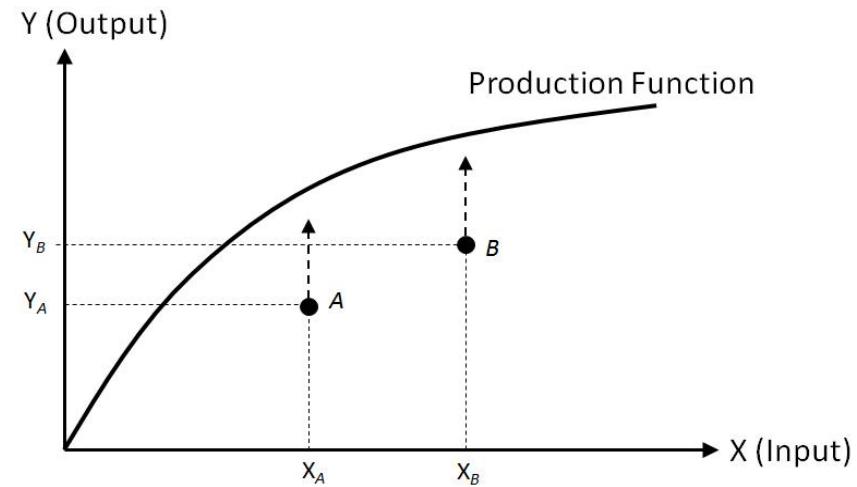
過去台灣製造業這幾年的歷史拉開...
我們的生產力與品質持續提升...
但為何獲利下滑？

可能是市場、競爭者、政府政策等，
但這些都不是我們能控制的因子...

Market Structure and Game Theory (external analysis外部分析)

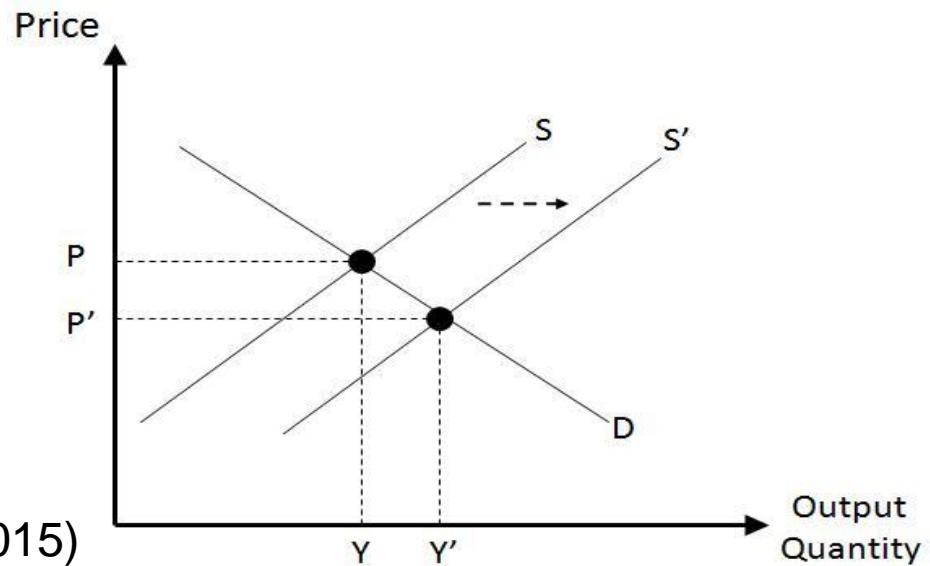


列強的競爭…



Curse of Productivity!?

Productivity driving → Price reduction (!?)



Lee and Johnson (2015)

智慧製造與工業3.5

“「工業3.5」是工業3.0和工業4.0之間的混合策略...
當務之急，先發展能讓智慧製造系統發揮效能的大數據分析和彈性決策能力...”

□ 德國工業4.0的願景

- 物聯網蒐集和分析所有生產過程相關的大數據
- 虛實整合系統以整合價值鏈的所有生產區塊
- 「大規模個人化生產」(mass personalization)
 - 延遲差異化/模組化

□ 工業4.0

- 大量個人化、彈性決策、聰明生產
- 有需求才生產、少量多樣高附加價值
- 結合終端使用者的物聯網與虛實整合，使各環節上的浪費得以消除
- 人體神經系統，偵測變化和即時決策
- 平台化，提升對供應鏈上下游廠商的資訊穿透和掌控能力，建置專屬製造平台，透過網路定期更新收費服務。



簡禎富 (2017, 哈佛商業評論)

□ 工業3.0

- 有限的數據和資訊來預測產品的需求
- 建置產能，然後想方法銷售
- 提升產能利用率和規模報酬
- 但由於生產數據與終端市場並沒有全面連結，最後一定會因為供需落差，而出現存貨和缺貨同時存在的現象。

□ 產業與企業層次：四大策略關鍵

● 策略關鍵1：現有製造優勢與管理經驗系統化與數位化

- 不只是靠低成本競爭，在競爭紅海中存活，是賺「**管理財**」
- 很大比率早年技職體系培養，年齡通常在50到65歲之間...

● 策略關鍵2：產品生命週期與營收管理

- 讓製造系統兼顧前台客製化和後台規模化的效能，需要串連設計、製造、銷售、物流、服務與保固的管理流程和大數據，縮短市場端到製造端的落差
- 台廠高度分工，個別企業難以掌握整個供應鏈的供需數據和產品生命週期各個階段的需求，...有些公司內部甚至不同資訊系統各自為政沒有整合...

● 策略關鍵3：軟硬體設備和分析能力垂直整合

- 台灣製造業大多中小規模，專業分工，受限於企業資源和決策者視野的局限，以致重視硬體甚於軟體，不重視數據分析和決策的軟實力，和人才培養。

● 策略關鍵4：永續發展和綠色供應鏈

- 歐盟國家未來更可能以此做為防守其他國家製造業的戰略武器...
- 特別是台灣在資源回收和循環經濟也有隱形冠軍...整合到廠務與綠色供應鏈，促進產業共生和資源循環利用



簡禎富(2019), 工業3.5：台灣企業邁向智慧
製造與數位決策的戰略，天下文化。

□ 馬雲 (2017/04)

- 第一次技術革命的能源是煤，第二次技術革命的能源是石油，第三次技術革命的能源主要是數據。所以社會的變革會越來越大，每一次技術的變革都是就業的變革…
- 一切業務不是基於數據的，忘掉，別做了，所有數據必須業務化，一切數據必須業務化，只有這樣才能迎接這個時代，
- 我們要讓機器做人做不好的事情，人創造不了的事情，這是我認為在 Machine intelligence 上面…人類有責任和擔當讓機器成為人的合夥人，成為人的合作夥伴，而不是讓機器來取代，這是我們這一代的人。
- 未來三十年已經不是力量的競爭、肌肉的競爭，甚至不是知識的競爭，而是服務別人能力的競爭，而是體驗的競爭，所以女人在未來的三十年將會蓬勃起來。
- 我們不能一方面說轉型升級，一方面還要對落後產業進行保護，這是不靠譜的。
- 小型企業創新是靠產品，中型企業創新靠技術，大型企業創新靠制度
- 智慧的競爭，我們一定要想明白讓孩子們必須去學習音樂、體育、運動、美術。音樂讓人有智慧通靈，體育運動讓人懂得團隊精神，我們讓孩子學會畫畫，有想像力，我們讓孩子懂得很多知識以外的事情。

□ The source of complexity in manufacturing is “variability (變異)”.

- Automation (自動化) is a powerful way to reduce variability.
- Data Science facilitates automation and real-time decision.

□ However ...

- 1. 資料科學是為了“追求真理”，而非單純的”商業管理或IT”問題
- 2. 資料科學的最終目的在於”決策”，而”預測”只是過程
- 3. 如果真的要以data分析，務必收集第一手資料 (田野調查訪談)
- 4. 跟”變異”一起共處吧~
- 5. 實事實上 ... 凡是能”量化”的可都已經不有趣了 ...

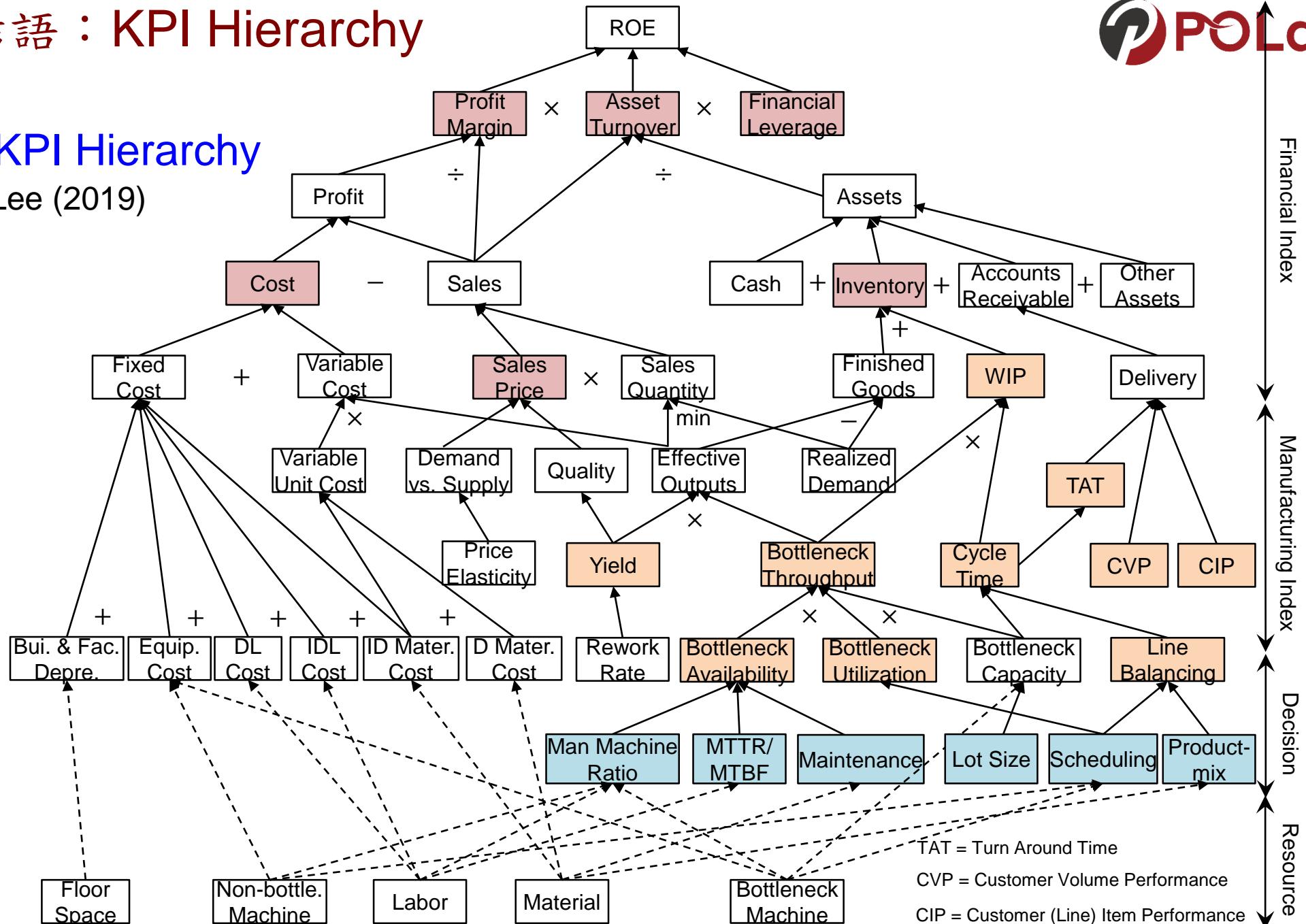
□ 預測性思維 → 處方性決策

□ 製造資料科學要掌握製造現場(過去、現在)並協助企業策略發展(未來).

- 掌握每個細節，製造現場的決策，如何影響企業的財務指標

KPI Hierarchy

Lee (2019)





放眼世界，我認為「藍海」根本不存在...

王品若能成為市場龍頭，不是因為有什麼了不起的創新，而是所有同仁都老老實實的捲起袖子打拚。

讓企業或個人成功的原因，絕對不是什麼藍海策略，而是那個**做到極致、絕不放棄**的堅持精神。

戴勝益 (2014)

感謝

成功大學 資訊工程系暨

製造資訊與系統研究所

有”羈絆”朋友們~

以及

參與這次課程的學員與

一起為製造業打拼的朋友們~

謝謝大家~

- 李家岩，2015，資料探勘，課程講義，國立成功大學工學院工程管理碩士在職專班，台南，台灣。
- 簡禎富、許嘉裕，2014，資料挖礦與大數據分析，前程文化。
- Dong, Z.-H. and C.-Y. Lee, 2017. Equipment Health Monitoring of Analytic Heretical Process Index in Semiconductor Manufacturing. 27th International Conference on Flexible Automation and Intelligent Manufacturing (FAIM2017), 27-30 June 2017, Modena, Italy.
- Guyon, I. and A. Elisseeff, 2003. An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, 3, 1157–1182.
- Han, J., M. Kamber, and J. Pei, 2011. *Data Mining: Concepts and Techniques*, 3rd edition, Morgan Kaufmann.
- Hillier, F. S. and G. J. Lieberman, 2010. *Introduction to Operations Research*, 9th ed., McGraw-Hill, New York.
- Hsu, S. and C. Chien, 2007. Hybrid Data Mining Approach for Pattern Extraction from Wafer Bin Map to Improve Yield in Semiconductor Manufacturing. *International Journal of Production Economics*, 107, 88–103.
- Hung, S.-Y., Y. L. Lin, C.-Y. Lee, 2017. Data Mining for Delamination Diagnosis in the Semiconductor Assembly Process. 27th International Conference on Flexible Automation and Intelligent Manufacturing (FAIM2017), 27-30 June 2017, Modena, Italy.
- Lee, C.-Y. and M.-C. Chiang, 2016. Aggregate Demand Forecast with Small Data and Robust Capacity Decision in TFT-LCD Manufacturing. *Computers & Industrial Engineering*, 99, 415–422.
- Lee, C.-Y. and A. L. Johnson, 2013. "Operational Efficiency", book chapter edited in: Badiru, A. B. (Editor), *Handbook of Industrial and Systems Engineering*, 2nd Edition, 17-44, CRC Press.
- Lee, C.-Y. and A. L. Johnson, 2015. Measuring Efficiency in Imperfectly Competitive Markets: An Example of Rational Inefficiency. *Journal of Optimization Theory and Applications*, 164 (2), 702–722.
- Lee, C.-Y., and T.-L. Tsai, 2019. Data Science Framework for Variable Selection, Metrology Prediction, and Process Control in TFT-LCD Manufacturing. *Robotics and Computer-Integrated Manufacturing*, 55, 76-87.
- Lee, C.-Y., 2019. Pitfalls and Protocols in Practice of Manufacturing Data Science. arXiv: <https://arxiv.org/abs/1906.04025>
- Liu, C.-W. and C.-F. Chien, 2013. An intelligent system for wafer bin map defect diagnosis: An empirical study for semiconductor manufacturing. *Engineering Applications of Artificial Intelligence*, 26, 1479–1486.
- Savage, L. J. 1951. "The theory of statistical decision," *Journal of the American Statistical Association*, 46, 55–67
- Wu, M.-Y. and C.-Y. Lee, 2015. Sampling-based NSGA-II for Stochastic Scheduling in Auto Parts Manufacturer. 2015 Chinese Institute of Industrial Engineers (CIIE) Conference & Annual Meeting, Taichung, Taiwan.

感謝大家的支持跟參與 還請多多指教



Contact Information:

name: 李家岩 (Chia-Yen Lee)

phone: 06-2757575 分機34223

email: cylee@mail.ncku.edu.tw

web: <https://polab.imis.ncku.edu.tw/>

台灣人工智慧學校
智慧製造與生產線上的資料科學
http://polab.imis.ncku.edu.tw/Talk/Data_Science_in_Manufacturing.pdf