

大數據與人工智慧服務

何信瑩 教授



數冠科技科技股份有限公司
國立交通大學衍生新創事業
AI大數據研究發展服務
創辦人：何信瑩

國立陽明交通大學
生物資訊及系統生物研究所

農業數位學堂 2021/4/6

何謂智慧農業?

以現行產業生產模式為基礎，因應消費市場需求進行產銷規劃，生產管理上輔以省工省力機械設備、輔具及感測元件的研發應用，並結合跨領域之資通訊技術(ICT)、物聯網(IoT)、大數據(Big Data)分析、區塊鏈(Block Chain)等前瞻技術導入，減輕農場作業負擔降低勞動力需求，提供農民更有效率的農場經營管理模式，生產符合消費者需求，安全、安心及可追溯的農產品。

AI技術應用

1. 數學建模
2. 模擬、預測
3. 最佳化設計

數據收集關鍵重點

1. 多參數
2. 多樣性
3. 代表性
4. 數量充足

↑
建立決策數學模型
(數據驅動型農業)

感測元件、資通訊技術、物聯網、共通資訊平台(大數據)、智慧機具、人機輔具

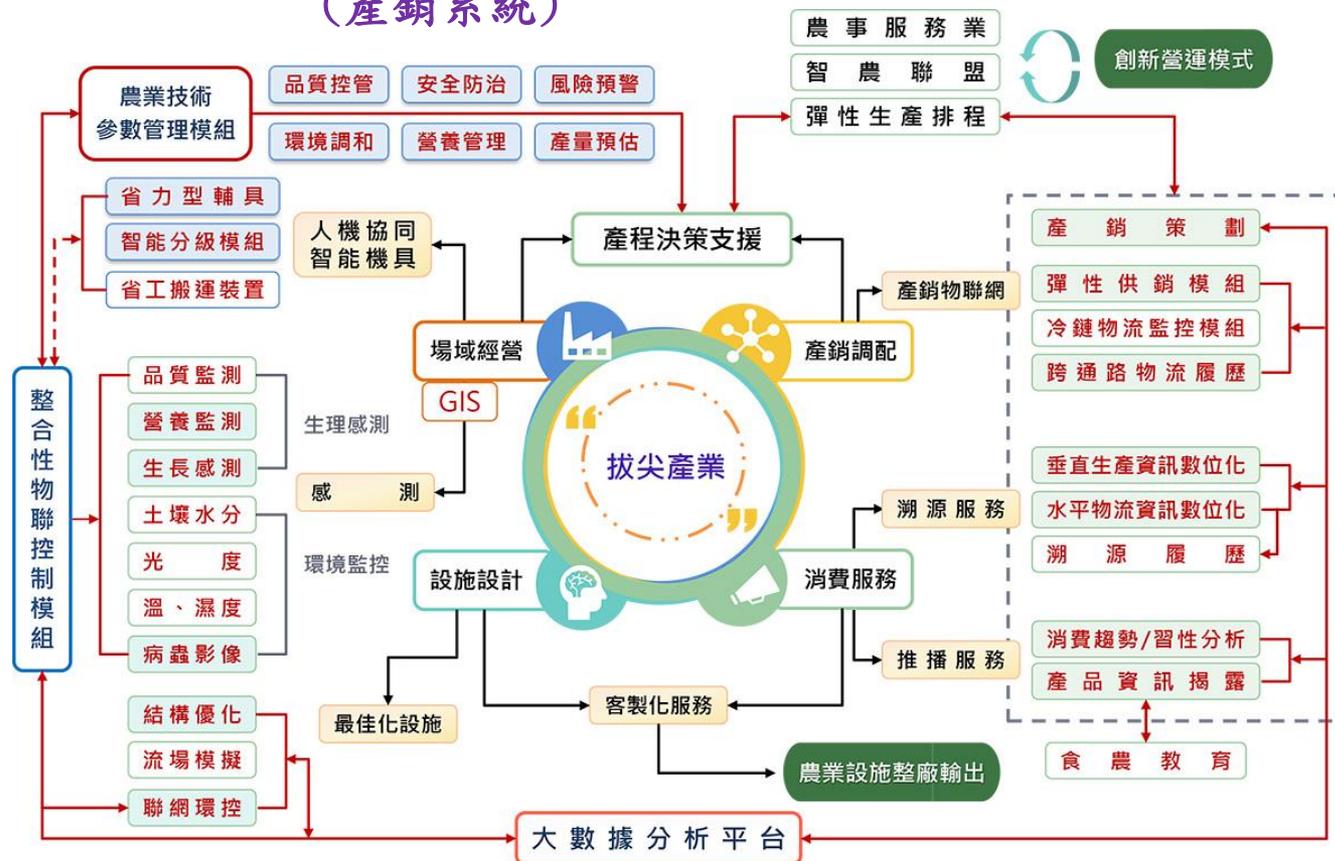


智慧農業 推動措施

建構跨域關鍵技術--智慧生產與數位服務產業應用架構

數位化資料 ➔ 自動化收集 ➔ 系統化分析 ➔ 建立數學模型 ➔ AI預測、設計 ➔ 決策建議

(生物環境) (產銷系統) (智慧化)



智慧農業的AI大數據挑戰

1. 短期不易獲取有用數據

農業數據收集是一個長週期過程。

2. 生長培育環境不同，需客製化模型

需要收集各自的數據才能建立專屬的模型。

3. 農業感測器數值的品質

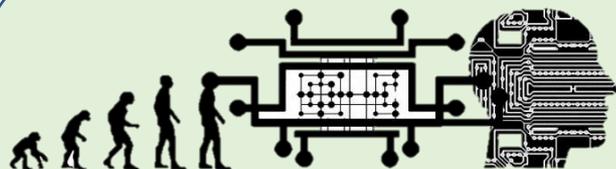
不精確的數據將導致所建立的模型不準確。

4. 樣本數量不充足，不易建模

可收集到感測器數據的農作物樣本數目充足，方能建立準確模型供預測分析。

人工智慧、機器學習、深度學習與演化學習

演化學習



Evolutionary Learning

以演化式演算法來優化機器學習的參數

Evolutionary Learning uses evolutionary algorithms to optimize model parameters in **machine learning**.

AI

計算機模仿人類思考能力與行為

machine learning

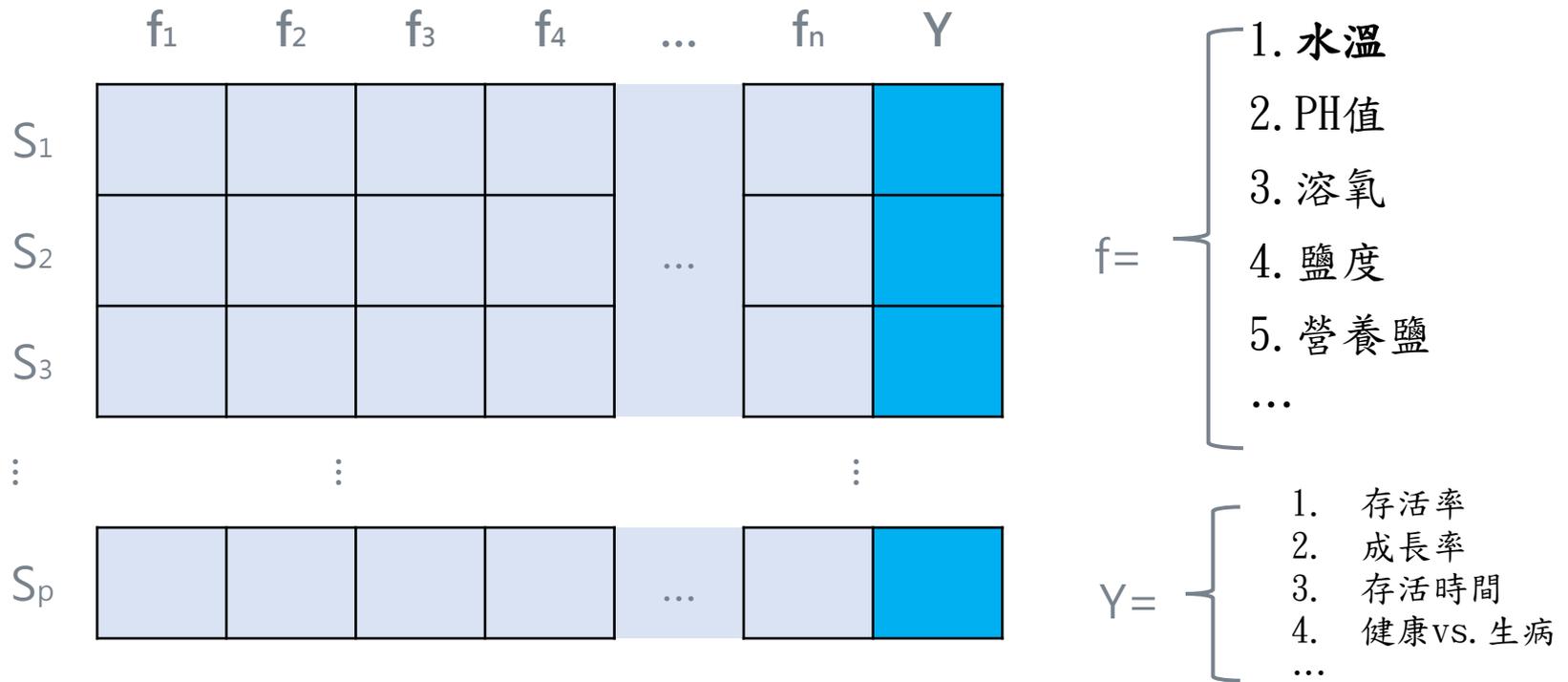
從資料中學習，並從中找出規律建立數學模型

deep learning

利用多層的非線性學習資料特徵

大數據可整合成為一張巨大表格 (f=特徵 S=樣本 Y=結果)

高效能數學建模，數據收集關鍵重點：1. 多參數 2. 多樣性 3. 代表性 4. 數量充足。



AI大數據現況

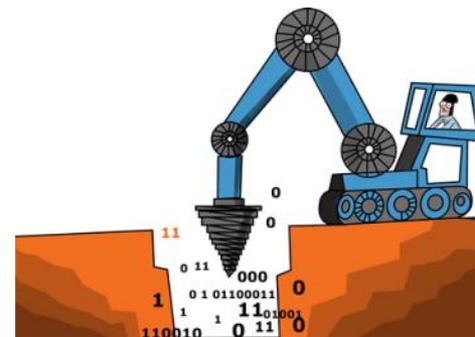
- 現今多使用統計分析或資料探勘作為選取和結果(Y)有關的因素(f_i)的方式^[1]。
- 要達到精準預測須利用**數學建模**，但礙於其候選特徵過多，導致建模預測仍待研究。



多



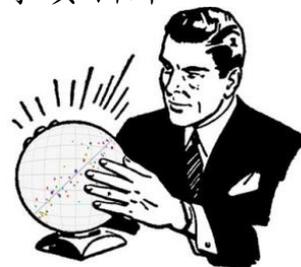
統計分析
醫學資料庫



資料探勘
EX: 序列比對



少

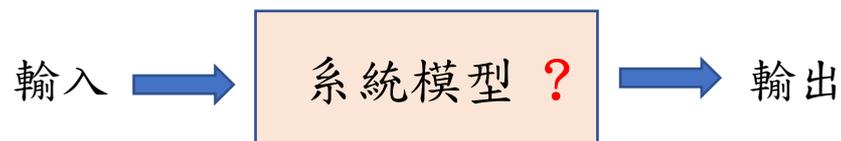
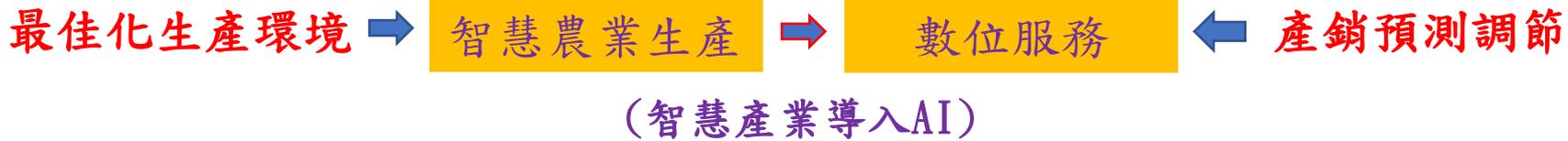


建模預測

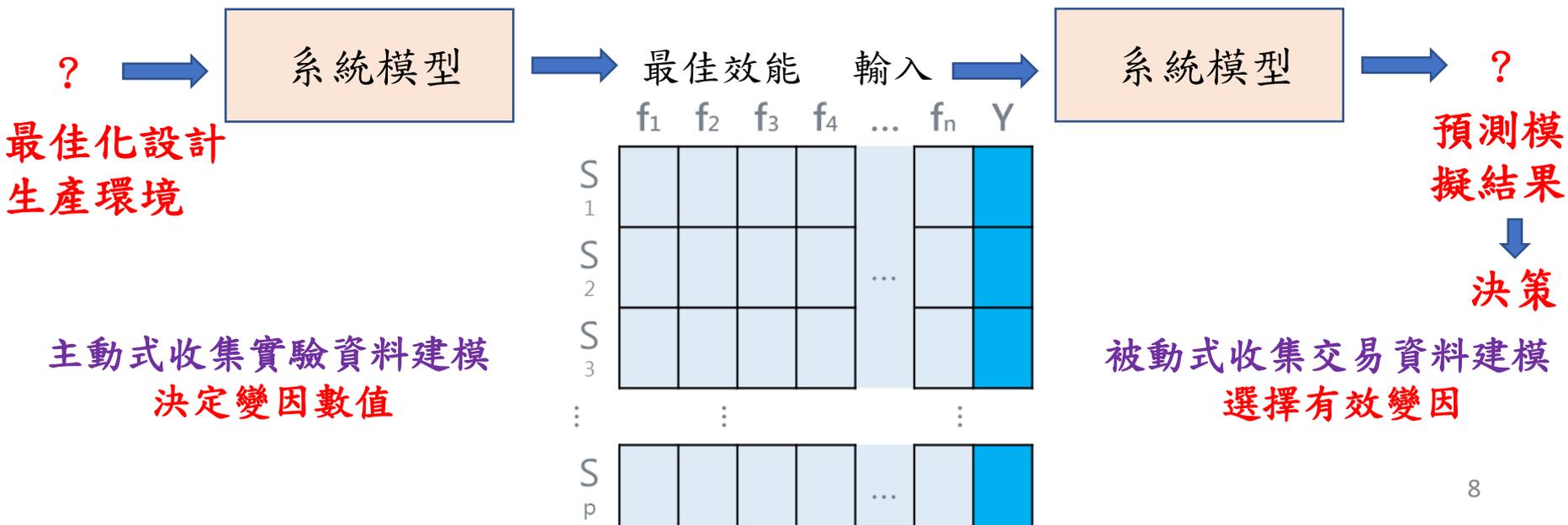
$Y=F(X)$
 $X=(f_1, f_2, \dots, f_m)$
 F: support vector machine

如何選取m特徵?

[1] Big data and big business: Should statisticians join in? D. Walker and K. Fung



建立決策數學模型
(數據驅動型農業)



統計分析與數學建模

- 許多企業皆重視**大數據分析**，但多數僅對“**農業大數據**”進行**統計分析**，較少能建立**數學預測模型**，難以做出預測與決策。
- 因應數據資料的多寡與資料的類型適合之**演算法**與**數學模型**也不同。
 - **深度學習**：可不需domain knowledge，但要非常大量的訓練資料。
 - **演化學習**：應用演化演算法來解決機器學習中的優化問題(要利用domain knowledge，降低訓練資料數量)



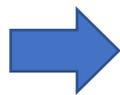
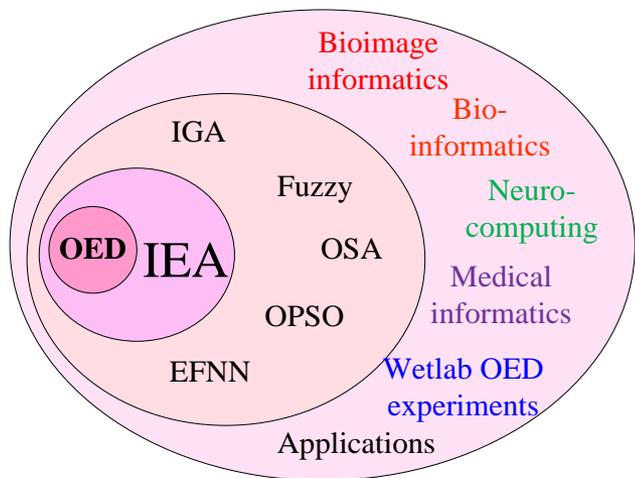
統計分析



數學建模

巨量資料分析-核心技術：智慧型演化演算法 (IEA)

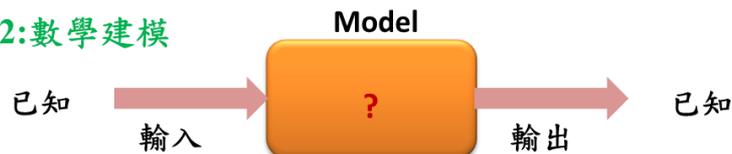
- Intelligent Evolutionary Algorithm (IEA)發表於IEEE Trans. Evolutionary Computation (過去五年平均IF=11.02, ranking=1/102), 為ISI高引用論文(Highly cited paper), 可以**有效地解決大型系統內大量參數之最佳化問題**。



Type 1:最佳化設計



Type 2:數學建模



Type 3:預測和模擬



GAs are first proposed by Professor John Holland in 1975.

(左圖為本計畫提出IEA的何信瑩教授。攝於1999年)

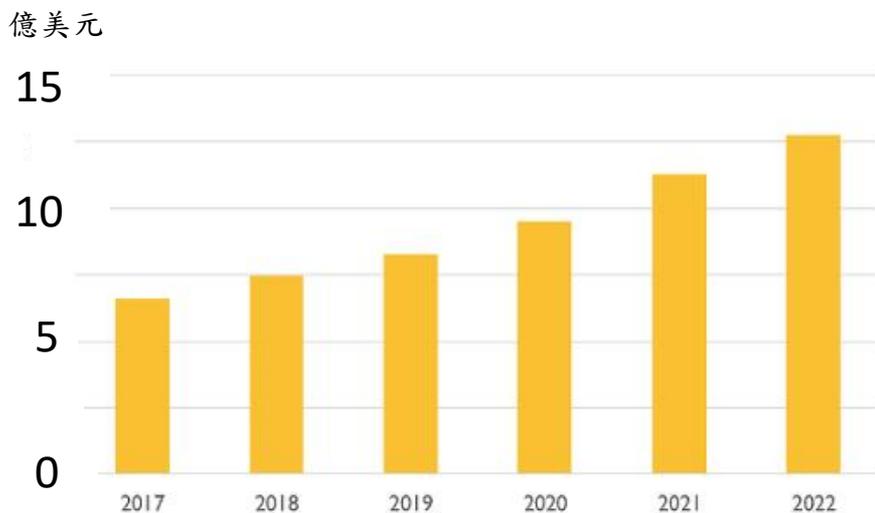


- 組合最佳化的一組解
- $C_m^n = n! / ((n-m)!m!)$
- 特徵選取：重要因子組合

智慧農業生產系統

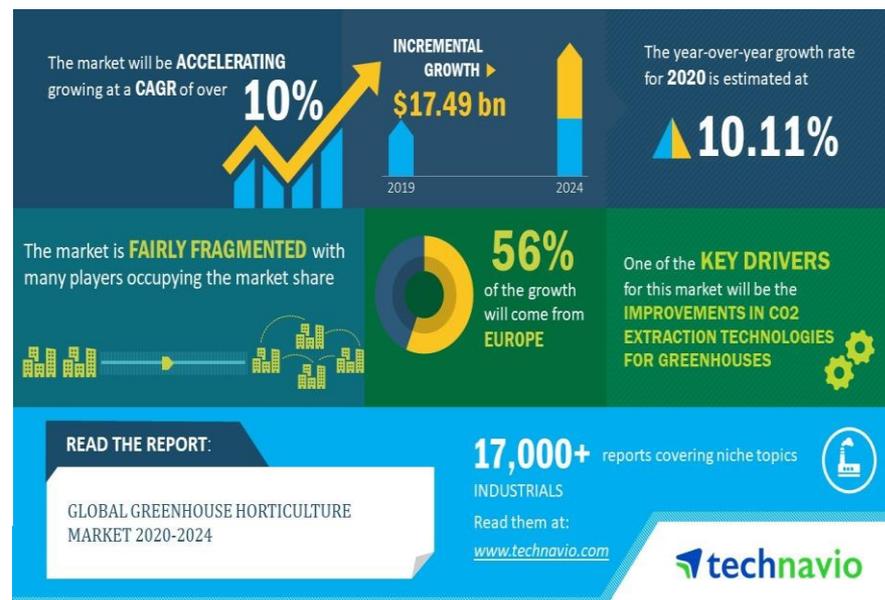
智慧型白蝦養殖系統

2016-2022年全球智慧養殖市場 CAGR=16%



蝴蝶蘭AI建模預測分析

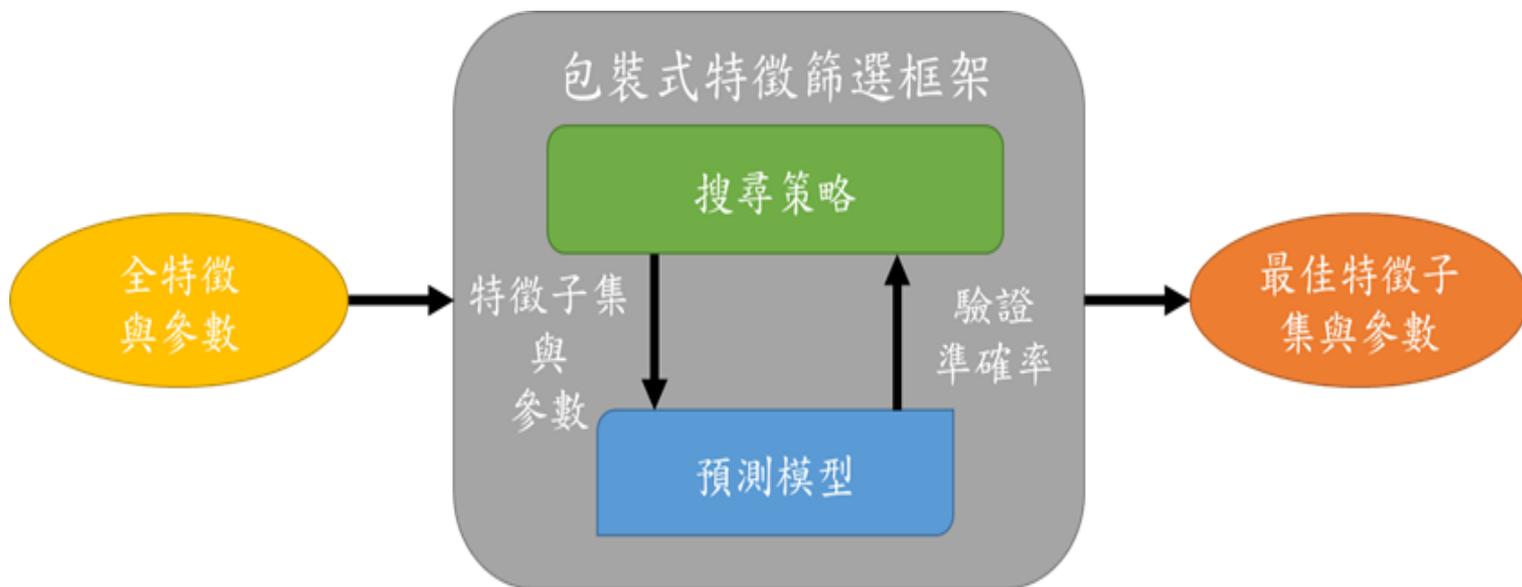
2020-2024年全球溫室園藝市場 CAGR=10%



蝴蝶蘭元素標準及葉片生育 AI建模預測分析與知識擷取

- 以花卉研究中心提供之蘭花植體元素含量及比例與對應生長品質的大量數據，建立植體元素與生長品質的數學函數，並分析各種植體元素對生長品質的重要性排名，便於日後的肥培管理，可根據數據分析出的植株需求，按照所需的比例給予植株最適合的成長環境。
- 以花卉研究中心提供之數個蘭花培育溫室的微氣候(溫度、濕度、照度等長時間紀錄數據)及生長品質數據(如葉片大小、產量等)，建立微氣候與生長品質的數學函數，用於預測模擬各種微氣候的生長品質，並推衍良好微氣候的管理要點知識。

本專案使用演化學習的一種方法，元啟發式演算法搭配直交實驗設計的機制，採分而治之 (Divide And Conquer) 的方式，將大問題拆成小問題解決，增強其效率，進而結合分類模型形成良好的動態特徵挑選與參數調整的方法，最終得到良好的回歸模型。



包裝式特徵篩選框架

蝴蝶蘭植體元素標準預測與分析

- 擷取各植體元素及葉面積、葉長度等資訊，整理資料後可用的蝴蝶蘭植體元素標準的資料集
- 供分析的資料集計有4個溫室(栽培業者)，3個花苗來源，11個營養元素(欄位特徵)，一個待預測結果是總增加葉面積，60個植體樣本。

蝴蝶蘭植體元素標準的可用資料集

植體編號	苗來源	栽培業者	K (%)	Ca (%)	Mg (%)	P (%)	Fe mg/L (ppm)	Mn mg/L (ppm)	Cu mg/L (ppm)	Zn mg/L (ppm)	N %	C %	C/N	總增加量
1A46L	台霖		1.465	3.4825	0.5475	0.241778	30.5	1380	4	33.625	1.477	44.492	30.12322	55.2
1A47L			1.59875	2.9275	0.63625	0.217729	58.125	1672.5	4	35.25	1.498	44.059	29.41188	50.7
1A48L			1.665	3.045	0.59625	0.150904	69.875	1778.75	2.875	46.75	1.37	44.084	32.1781	47.8
1A49L			2.445	2.705	0.5775	0.200893	58.5	900	2.75	45.875	1.546	42.509	27.49612	38.5
1A50L			1.695	3.0825	0.51125	0.221891	99.5	647.5	3.75	147.875	1.461	45.433	31.09719	49.2
1B45L			1.42375	2.5575	0.6175	0.399738	66.25	621.25	1.25	28.375	1.844	44.074	23.9013	54.4
1B46L			1.88625	2.57875	0.56125	0.504488	60.25	627.5	2.25	30.625	2.316	45.118	19.481	53.8
1B47L			2.32875	3.1425	0.56125	0.43816	49.75	546.25	2.5	35.125	2.011	44.616	22.18598	56
1B48L			2.0175	2.5575	0.4225	0.398381	96.625	431.75	2.5	32.375	1.854	45.086	24.31823	57.6
1B49L			1.35125	2.4275	0.53875	0.365046	53.625	450	2.25	33.375	1.895	44.927	23.70818	44.4
1C46L			1.89	6.41625	0.32375	0.221623	50.25	1190	3.25	21	1.559	42.773	27.43618	53.4
1C47L			2.11625	6.4125	0.4275	0.278173	42.75	1160	3.375	19.5	1.619	42.08	25.99135	41.4
1C48L			2.12	5.35875	0.455	0.2391	67	1061.25	3	19.625	1.686	42.716	25.33571	52
1C49L			2.2725	5.6925	0.55125	0.230624	60.25	776.25	2.875	20.75	1.389	42.199	30.38085	49.4
1C50L			1.4725	5.8125	0.515	0.235798	44.125	898.75	2.875	19.5	1.432	42.736	29.84358	46
1D46L			1.48375	2.475	0.59625	0.368531	54.5	645	2.75	27.625	1.915	44.794	23.39112	62.2
1D47L			2.415	2.76125	0.59625	0.331488	42.625	818.75	2.625	27	2.027	44.874	22.13814	47.5
1D48L			1.70625	2.94875	0.5675	0.373236	54.125	802.5	2.5	25.375	2.053	44.421	21.63712	58.4
1D49L			2.16125	2.745	0.47	0.317594	50	681.25	2.625	24.5	2.079	44.209	21.26455	54.3
1D50L			1.77375	2.9575	0.455	0.314501	44.875	598.75	2.125	25.625	2.099	44.73	21.31015	56.4
2A46L	龍鼎		1.41	3.465	0.505	0.264444	34.875	1011.25	4.5	31.5	1.544	44.56	28.8601	52.4
2A47L			1.7325	2.61625	0.39625	0.231923	34.125	952.5	4.25	56	1.406	45.472	32.34139	55.9
2A48L			1.555	3.49875	0.42875	0.207266	27.625	1066.25	2.875	37.875	1.654	44.949	27.17594	44.6
2A49L			1.765	3.32875	0.55	0.22167	24.25	992.5	2.25	83.375	1.582	44.471	28.11062	53
2A50L			1.74875	2.62875	0.60625	0.160216	49.625	1241.25	2.625	43.75	1.359	44.591	32.81163	45.8
2B46L			1.46	3.285	0.50625	0.400948	34	711.25	3.25	34.875	1.871	44.478	23.77231	59.1
2B47L			1.5375	3.44	0.4525	0.362649	46.625	688.75	1.625	30	1.921	44.773	23.30713	59.8
2B48L			1.58125	3.4225	0.46625	0.364856	34.75	900	2	25.125	1.822	44.879	24.63172	61.9
2B49L			1.40875	3.3175	0.49	0.334096	33.25	853.75	2.625	32.25	1.715	44.376	25.87522	55.2
2B50L			2.06875	2.6875	0.54875	0.370646	26.125	925	2.5	36.375	2.021	45.144	22.33746	46.2
2C46L			1.5175	5.77125	0.5125	0.23608	33.875	990	2.75	25.25	1.51	42.926	28.42781	56.4
2C47L			1.9075	5.55375	0.4325	0.223558	27.125	873.75	3.5	21.875	1.507	43.272	28.714	59.2
2C48L			1.68875	5.52	0.34125	0.283374	29.75	747.5	4.625	26.875	1.672	43.316	25.9067	154.3
2C49L			1.55875	5.56125	0.375	0.260219	28.875	907.5	2.5	29.875	1.674	42.78	25.5556	44.1
2C50L			1.8975	5.7375	0.43125	0.254709	27	957.5	3.375	28.5	1.562	42.984	27.51857	68.7

蘭花培育溫室的微氣候及生長品質分析

- 提供了四座蘭園長期由IOT收集的微氣候資訊，包含溫度、濕度、二氧化碳與有效光合輻射，在四座蘭園中皆有園中蘭花三種品種的五筆抽樣數據，並有幾種指標(葉面積、新增葉面積、葉長生長量與新增葉片數)，使用這些資料進行研究。
- 給定的樣本中含有一系列非常精細的微氣候變化資料，但因樣本珍貴且有限，太細微精確的數據還未能展現其價值，又因為資料的缺值問題在小樣本中較不易處置，因此本研究採用每日平均後的數據進行處理。

四種特徵的統計資訊

蘭園	溫度 (極大、極小)		濕度 (極大、極小)		二氧化碳 (極大、極小)		光合有效輻射 (極大、極小)	
	A	31.92	27.03	94.30	56.12	508.59	400.09	40.57
B	30.70	26.30	91.94	54.39	539.40	394.90	167.25	4.53
C	32.75	25.55	97.15	57.35	577.86	262.27	564.17	2.94
D	31.92	26.87	94.56	57.85	548.27	416.09	75.85	1.61
all	32.75	25.55	97.15	54.39	577.86	262.27	564.17	1.61

蘭園	溫度(平均、標準 差、亂度)			濕度(平均、標準 差、亂度)			二氧化碳(平均、 標準差、亂度)			光合有效輻射(平 均、標準差、亂 度)		
	A	29.07	0.91	1.97	80.55	6.59	2.24	441.8	19.83	1.43	24.13	6.38
B	28.63	0.90	2.00	77.96	6.44	2.24	456.0	26.19	1.67	37.93	12.00	0.70
C	28.32	0.87	1.88	84.93	6.37	2.14	472.8	44.03	2.01	45.41	66.66	0.90
D	28.97	0.80	1.89	83.21	5.26	2.02	471.0	24.77	1.64	38.90	11.05	0.74

智慧型白蝦養殖系統

克服蝦類養殖所面臨的問題，應從養殖池之環境改造、環境修復開始建立池塘中一個完整的生態系，並維持水中之生態平衡勢在必行。

建立客製化蝦類養殖系統專屬數學模型(決定可控參數最佳值)

1. 水溫：23-32°C都可養殖	3. pH：在8.0±0.3最佳
3. 鹽度：在5-40 ‰都可養殖，但最適鹽度為12-20 ‰	4. 溶氧：在4ppm以上，勿低於3ppm
5. 化學需氧量：在5-30ppm之間均可，但溶氧充足時，7-10ppm成長最佳	6. 透明度：在20-60公分都可，但在35±5公分最佳
7. 水色：維持綠色(在淡水地區)或紅棕色(在海水地區)	8. 營養鹽：磷酸鹽0.1-0.3毫克/公升。矽酸鹽在2.0毫克/公升左右。氨態氮在0.4毫克/公升以下。亞硝酸鹽在0.5毫克/公升以下(淡水)或1.0毫克/公升以下(海水)。
9. 總鹼度：最好能在180-240毫克/公升	10. 異營菌：在10 ³ -4 CFU/毫升之濃度較少發病
11. 放養密度：每公頃40-150萬尾	12. 每日餵食2-4次
13. 水池水量	14. 季節氣候

結果(預測) 成長率、存活率、生病.....

最佳養殖系統

1. 系統假設有9種可控制變數(參數)：水溫、pH、鹽度、溶氧、透明度、化學需氧量、總鹼度、放養密度、水池水量
2. 養殖系統要最大化白蝦的成長率/存活率
3. 若每一變數有3水準(23-32°C)，共有3的9次方=19,683種組合
4. 實驗數據收集：多參數(9變因)、多樣性(>3水準)、代表性(主動設計)
5. 搭配直交表實驗設計和演化學習建模，以最少次實驗(金額少)、最少代數(時間短)、選擇最少變因，達到最準確模型。
6. 利用數學模型，動態調節環境因素，保持最佳養殖環境。

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
A	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2
B	0	0	0	1	1	1	2	2	2	0	0	0	1	1	1	2	2	2	0	0	0	1	1	1	2	2	2
C	0	0	0	1	1	1	2	2	2	1	1	1	2	2	2	0	0	0	2	2	2	0	0	0	1	1	1
D	0	0	0	1	1	1	2	2	2	2	2	2	0	0	0	1	1	1	1	1	1	2	2	2	0	0	0
E	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2
F	0	1	2	0	1	2	0	1	2	1	2	0	1	2	0	1	2	0	2	0	1	2	0	1	2	0	1
G	0	1	2	0	1	2	0	1	2	2	0	1	2	0	1	2	0	1	1	2	0	1	2	0	1	2	0
H	0	1	2	1	2	0	2	0	1	0	1	2	1	2	0	2	0	1	0	1	2	1	2	0	2	0	1
I	0	1	2	2	0	1	1	2	0	0	1	2	2	0	1	1	2	0	0	1	2	2	0	1	1	2	0

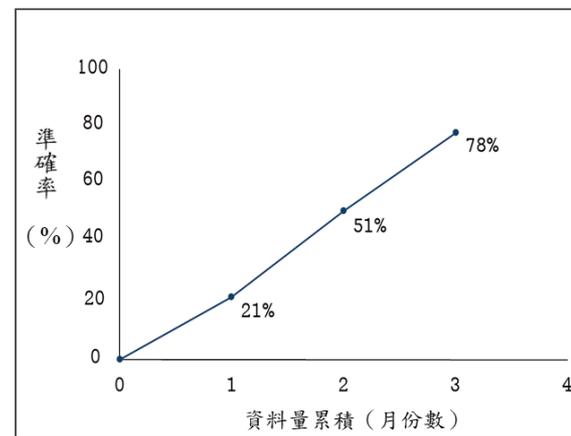
土壤菌相AI及智慧施肥系統

農譯科技

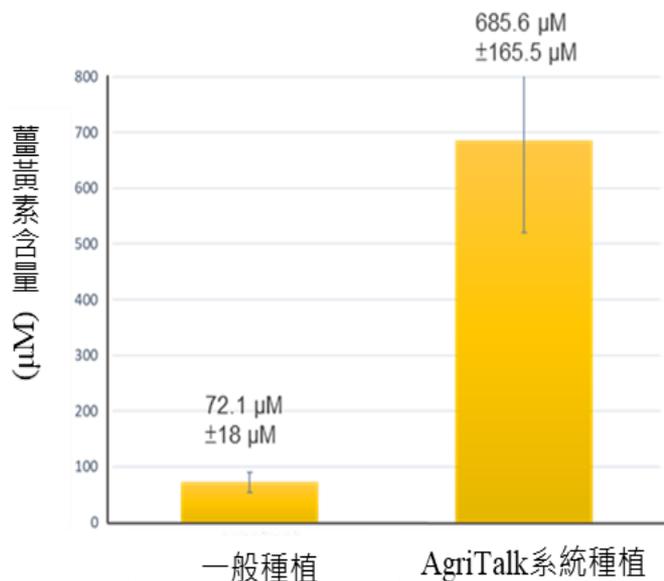
為特定作物、特定肥料客製化AI系統以精準施肥



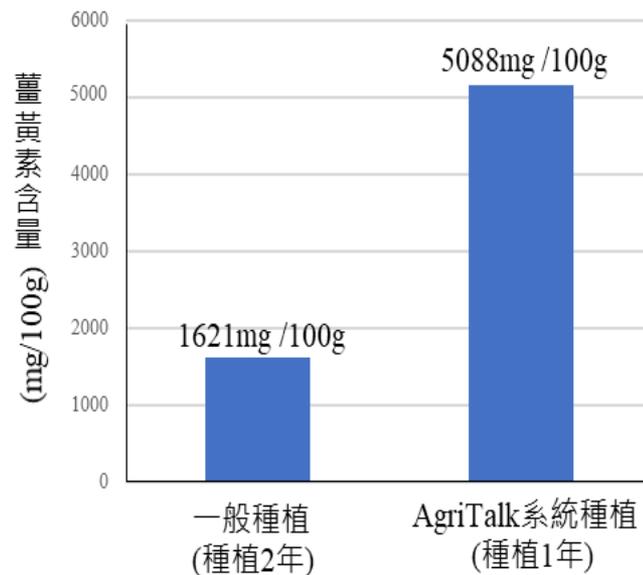
模型學習曲線的準確率



從作物功效性成分證明：土壤菌相AI預測系統之效益



AgriTalk 系統種植之葉片類黃素含量較一般種植高出9倍



AgriTalk 利用較短種植時間,達到較高的類黃素

核心技術

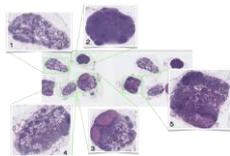
最佳化數學建模的演化學習平台-科技部台灣研究亮點 (2019/06)

Evolutionary Learning Platform for Biomedical Data (1/2)

Biomedical Data (Input)



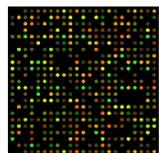
Pathological section whole scanner



HE Stain image



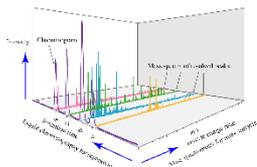
DNA sequencer & microarray



Expression data



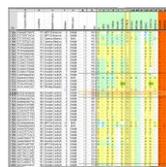
LC-MS/MS



3D spectrum



Biomedical databases
TCGA, GEO, NHRI ...



Excel Table

Problem dependent

Evolutionary Learning Platform

	f_1	f_2	f_3	f_4	...	f_n	Y
S_1					...		
S_2					...		
S_3					...		
\vdots							
S_p					...		

High dimension (n), insufficient samples (p), label uncertainty (Y) and strong interaction

- $f_i =$
- 1. Image features
 - 2. Gene expression values
 - 3. Spectrum values
 - 4. Biomedical values

- $F(X) = Y$
- 1. Cancer metastasis
 - 2. Survival time
 - 3. CKD occurrence
 - 4. Biomedical values

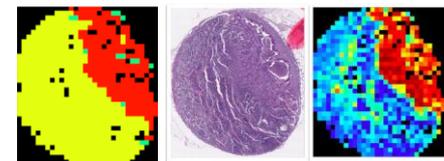
IEA
(Global Optimization Algorithm)
 C_m^n



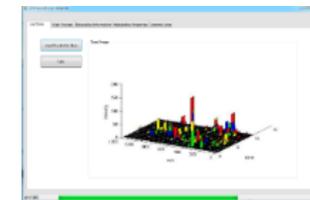
ISI 高引用論文

Problem independent

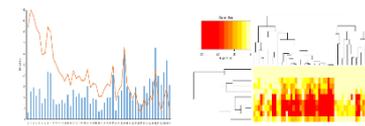
Prediction Model (output)



lymph node metastasis



CKD signature discovery



Survival time prediction and signature discovery

$$X = \{f_1, f_2, \dots, f_m\}$$

a set of biomarkers!



2017年科技部FITI創新創業激勵計畫以「醫療影像人工智慧電腦輔助分析系統」榮獲【創業潛力獎】

2019台北電腦展 發表： (2019-06-04)
「創新的演化學習平台於人工智慧醫療應用」



數冠科技科技股份有限公司
交通大學衍生新創事業
AI大數據研究發展服務

syho@drhow.com.tw syho@nctu.edu.tw 何信瑩 教授