

## 使用關聯規則挖掘埔里天氣因子與 PM2.5 規則

王珮如

國立台灣科技大學資訊管理學系

m11009113@gapps.ntust.edu.tw

林昱媛

中國科技大學資訊管理系

melody376@gmail.com

尹邦嚴

中國科技大學資訊工程系

pyyin97@cute.edu.tw

戴榮賦

國立暨南國際大學資訊管理學系

rfday@ncnu.edu.tw

### 摘要

空汙是人們關心的議題，其中 PM2.5 會對人造成健康危害，然而 PM2.5 沉積量會因地形、氣候而有所差異，故本研究找出埔里各級 PM2.5 與各級天氣因子隱藏規則，以觀察該地 PM2.5 之關聯樣態，幫助釐清 PM2.5 升降之原因。本研究將氣壓、溫度、濕度、風速、風向及 PM2.5 資料離散化，利用關聯規則學習設計挖掘流程，找出各天氣因子及 PM2.5 高中低等級之關聯規則，接著以統計評估雨量對於規則發生之影響。以生成規則相似度來說，介於 12~23 之 PM2.5 沉積量較無信心顯著的關聯規則，而小於 12 之 PM2.5 沉積量及大於 35 之 PM2.5 沉積量則有較顯著的關聯規則，可供後續學者進一步分析造成關聯規則的原因。

關鍵詞：PM2.5、天氣因子、資料探勘、關聯規則、樣態挖掘

## 壹、導論

### 一、研究背景

人類每天無時無刻都需要呼吸，而在呼吸的過程中，不可避免地會吸入空氣中的污染物，其中粒狀污染物相較於其他污染物能夠在空氣中停留更久，而直徑最小的粒狀污染物細懸浮微粒(PM2.5)，能夠進入人體更深處，沉積於人體氣管及肺部，Burnett et al. (2014)研究表明，暴露於 PM2.5 之中，有增加罹患心肺相關疾病之風險，亦有增加死亡率之可能，在 Elliott & Copes (2011)的研究中，更表明了即使是少量的 PM2.5，仍可能對於人體造成健康的威脅，故而對 PM2.5 進行相關研究，了解 PM2.5 及天氣因子樣態的關聯，有助於我們釐清 PM2.5 高低變化與天氣因子之間的關係。

### 二、研究目的與動機

目前對於 PM2.5 升高之原因有許多推斷與討論，Wei, Peng, Ma, Qiu & Wang (2019)以小波轉換研究北京四個街道的 PM2.5 與交通量的相關性及週期性，結果證明兩者週期有顯著的關聯，且 PM2.5 對於交通量有 0.3~0.9 分鐘的延遲，Tsai, Sopajaree, Kuo, & Yu (2015)則藉由分析台南在中秋節前中後的期間的化學成分，發現雖然烤肉活動對於 PM2.5 升高有所貢獻，但中國秋天農業廢棄物燃燒產生之氣溶膠藉由氣流輸送到台南才是 PM2.5 升高之主因，Kim, Chang & Kim (2021)使用回歸對人口、富裕度和技術的隨機影響(Stochastic Impacts by Regression on Population, Affluence and Technology, STIRPAT)模型分析六個國家 254 個城市的人均 GDP、人口密度、人口規模對 PM2.5 的影響，結果以人均 GDP 影響最大，每增加 1% 的收入，將會降低 0.074 的 PM2.5 濃度。這些研究結果皆能夠為 PM2.5 演變提供參考，然而 PM2.5 與天氣因子間之相關性會因地形、氣候、時間而有所差異，故對一特定地點進行研究，才能了解該地獨特的樣態。本研究以埔里為研究目標，希望藉由關聯規則挖掘找出隱藏的非直覺規則，協助人們了解埔里 PM2.5 及天氣因子間的樣態，從而能夠解析 PM2.5 升降之原因，同時亦可作為未來進行相關研究討論之依據。

## 貳、文獻探討

### 一、PM2.5 與天氣因子相關性

過去對於 PM2.5 及天氣因子之關係有過許多探討，Yang, Yuan, Li, Shen & Zhang (2017)以季節與空間進行多尺度相關分析中國 68 個城市 PM2.5 及天氣因子的相關性，說明地點與季節之差異會導致 PM2.5 與天氣因子有正負相關之差異，因此探討一地之天氣樣態有助於人們分析該地造成污染之原因，例如 Wang & Ogawa (2015)使用線性分析及斯皮爾曼等級相關係數探討了長崎 PM2.5 及天氣因子的相關性，並得出長崎主要 PM2.5 是由東亞飄散而來之結論，而因地區的不同，需使用不同的因子進行分析，Zhang, Zhang, Zhao & Lian (2020)對北京進行 PM2.5 預測，因其地處北方，相比緯度較低的地區，需多投入露點、降雪兩因子進行分析。

### 二、埔里地區樣態

本研究以埔里為目標，在過去亦已有研究對埔里進行 PM2.5 及天氣因子之探討，徐玉兒 (2021) 使用多尺度分析方法，發現埔里之 PM2.5 及天氣因子的相關性在一年四季之差異，PM2.5 與溫度在夏季為正相關，在其他季節則為負相關，林彥呈等人 (2019) 則是使用 KMeans 分群，依據逆溫現象，歸納出埔里 PM2.5 與天氣因子之關聯樣態，並考量外縣市污染因素，建立 PM2.5 預測模型進行預測。本研究希望藉由關聯規則挖掘出隱藏的規則，使人們得以對於埔里天氣因子及 PM2.5 樣態有更多的認識，並為未來建立預測模型或相關研究打下基礎。

### 參、研究方法

本文研究流程如圖 1，先將所有資料進行離散化，使用離散化的 PM2.5、氣壓、溫度、濕度、風速、風向資料依照設計挖掘出關聯規則，後將規則與各級雨量進行統計分析，評估雨量對於規則發生之影響。

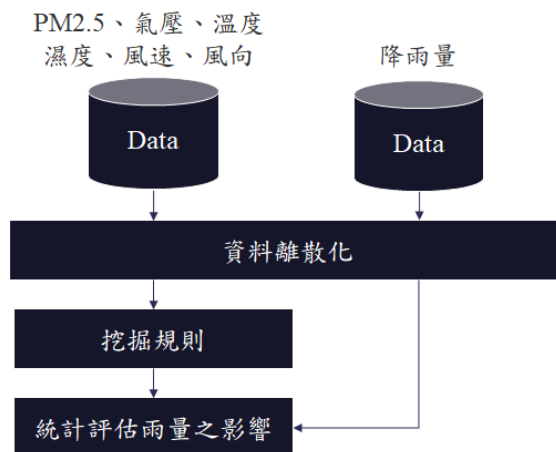


圖 1: 本研究之實驗流程

#### 一、資料來源

本研究以南投埔里為目標，採用 2016~2019 年環保署空氣品質監測網的 PM2.5 資料以及中央氣象局觀測資料查詢系統的氣壓、溫度、濕度、風速、降雨量資料進行關聯規則分析。

#### 二、前處理

PM2.5 及氣象因子資料皆為數值資料，故須將數值資料先進行離散化處理，根據關聯規則學習本身之特性，項集出現之頻率愈高愈有機會出現於規則中，因此資料均勻度會影響生成之規則，故在離散化資料時須注意離散化後資料各類別占比。觀察各資料分布情形如表 1，各圖中的 x 軸為資料數值，y 軸為資料筆數，在所有資料中，唯有雨量的筆數差距極端，為 0 的筆數在歷年皆占了八千筆左右，約九成的比例，因此不管如何切割雨量資料，皆會有一項出現次數占 90% 以上，若是將其一起進行關聯分析，占九成之雨量項容易出現於前排所有規則，導致無法挖掘出其他隱藏的規則，故在使用關聯規則生成規則的過程中，不投入雨量因子分析，而是在生成規則後，再以統計分析在各級雨量中該規則發生的機率。

在資料離散化的步驟，風向以八方位進行切割，PM2.5 以指標等級作為切割依據，資料均勻為目標進行分類，如圖 2，將第一級到第三級 PM2.5 分別切為 1、2、3 類，而第四到第十級則切分為 H 類，而氣壓、溫度、濕度、風速、雨量資料的切割，則使用 temper 進行切割，若是有數值筆數大於總筆數五分之一，則該數值自己為一類，並將剩下資料以數量均勻分割為兩等份，若是沒有數值筆數大於總筆數五分之一，直接以數量均勻分割為三等份。離散化結果如表 2，氣壓、溫度、濕度、風速皆切割得十分均勻，大約各自占三成，而風向以北風、西風、西北風、東南風為主，前兩項分別占了兩成，後兩項則分別占了一成左右，PM2.5 各項占比與天氣因子相較，以 2 類較多、1、3、H 類較少，在挖掘規則時，支援度需設定大於該項占比，否則將找不到含有該項之規則。

指標等級	1	2	3	4	5	6	7	8	9	10
PM2.5濃度	0~11	12~23	24~35	36~41	42~47	48~53	54~58	59~64	65~70	71~∞

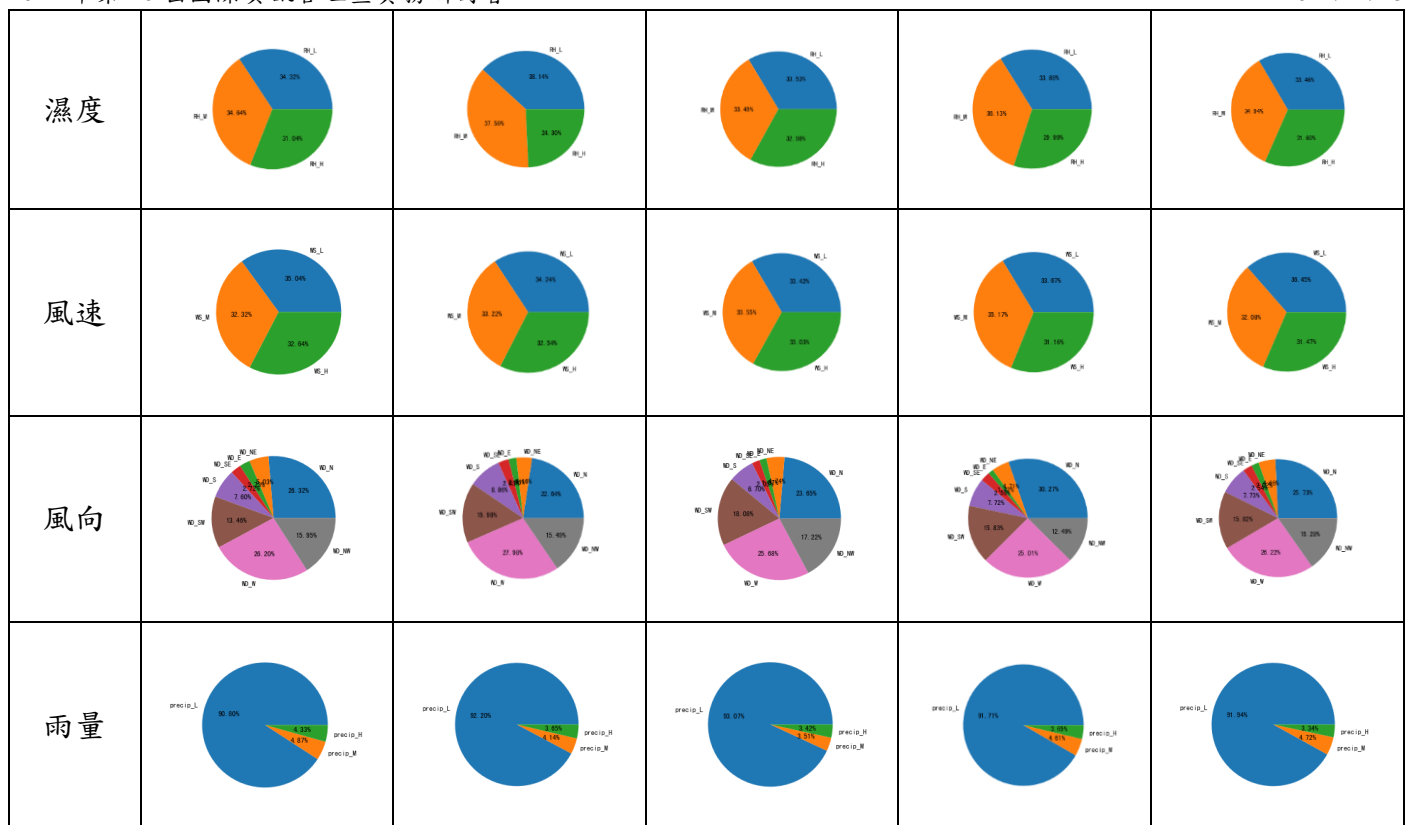
圖 2: PM2.5 資料離散化之切割邊界

表 1: PM2.5 與天氣因子之資料分布

因子	2016	2017	2018	2019	2016~2019
PM2.5					
氣壓					
溫度					
濕度					
風速					
風向					
雨量					

表 2: 資料離散化後各項資料占比

因子	2016	2017	2018	2019	2016~2019
PM2.5					
氣壓					
溫度					



### 三、挖掘規則

本研究使用關聯規則學習的概念，計算支援度、信賴度、提升度三指標數值，並設計步驟篩選規則，篩選出較有意義的規則，找出天氣因子與 PM2.5 高低之規律性，供與人們參考，以了解在同一時間內的 PM2.5 及天氣因子樣態。在本研究，以「因子\_等級」為項目命名，例如，PM2.5\_1 代表第一類 PM2.5，項集為整個專案集合的非空子集，{pressure\_H, temperature\_L}、{PM2.5\_1} 皆屬於項集，規則則為 A→C 的蘊含式，A 為先導(antecedent)，C 為後繼(consequent)，兩者皆是由項目組成的項集，例如：{pressure\_H, temperature\_L} → {PM2.5\_1}。計算支持度的方式如公式(1)，表示在整個專案集合中，A 事件及 C 事件同時發生之頻率，計算為信賴度的方式公式(2)，表示當 A 事件發生時，C 事件發生的機率，提升度的計算方式如公式(3)，表示在 A 事件發生時，C 事件發生之機率，與 C 事件發生之比率，A 對 C 是否有提升程度。

$$\text{support}(A \rightarrow C) = P(A \cup C) \quad (1)$$

$$\text{confidence}(A \rightarrow C) = P(C|A) \quad (2)$$

$$\text{lift}(A \rightarrow C) = \frac{P(C|A)}{P(C)} \quad (3)$$

本研究設計之挖掘步驟如下，在步驟 1 先過濾出支持度大於 0.2 且提升度大於 1 之規則，0.2 為本研究測試出較為適合之支持度，而提升度大於 1，則是為了留下先導對於後繼有提升效果的規則，為篩選有興趣之各級 PM2.5 規則，在步驟 2 將後繼含有各級 PM2.5 的結果分別過濾出來，認為信賴度較高之資料較有價值，因此在步驟 3 將規則以信賴度由大到小排序，在步驟 4，若出現規則之信賴度相等，則以支持度由大到小排序，在最後的步驟 5，取前 20 條規則，作為最有價值的規則。

步驟 1：過濾出支持度大於 0.2 且提升度大於 1 之規則

步驟 2：將後繼含有各級 PM2.5 的結果分別過濾出來

步驟 3：將規則以信賴度由大到小排序

步驟 4：若有幾條規則信賴度相等，則以支持度由大到小排序

步驟 5：取前 20 條規則，作為最有價值的規則

## 四、統計評估分析雨量之影響

為評估雨量對於規則之影響，本研究以公式(4)計算各分級雨量中發生規則的機率，意即在各級雨量中該規則發生的比例。

$$\text{confidence}(\text{precip\_x} \rightarrow \text{rule}) = P(\text{rule}|\text{precip\_x}) \quad (4)$$

## 肆、研究結果

經由本研究方法挖掘出的各級 PM2.5 規則如表 3，發現生成之歷年規則有極高的相似度及項目，如後繼含有 PM2.5\_1 之規則多由低氣壓、中溫、高濕度、低風速及北風幾個項目與 PM2.5\_1 組成，後繼含有 PM2.5\_2 之規則溫度多為高溫，濕度為低級及中級，風向北風、西北風、西風，氣壓、風速有較多變化，常出現在後繼含有 PM2.5\_3 之規則中的項目為高氣壓、低溫、中濕度、西南風，風速則是中級及低級皆有，常出現後繼含有 PM2.5\_H 規則之項目為高氣壓、低溫、中濕度，風速為中級及低級，各級 PM2.5 規則在歷年內的相似度的由高到低排序為 PM2.5\_1、PM2.5\_H、PM2.5\_3、PM2.5\_2，此因 PM2.5\_2 的天氣模式較為混亂，導致篩選出的規則中項目較不相似，相對而言，PM2.5\_1 及 PM2.5\_H 項目資料相似度較高，此因其天氣模式較為單一。探討各因子於歷年各級 PM2.5 規則出現頻率，氣壓與溫度歷年在各級 PM2.5 規則皆頻繁，濕度出現頻率稍低，風速及風向於 PM2.5\_H 極少出現，推論對於高 PM2.5 規則，風速及風向相對於其他因子，並非主要影響因子。

表 3: 歷年各級 PM2.5 前五名規則及其支持度、信賴度、提升度

含有 PM2.5_1 之規則				
年份	規則	支持度	信賴度	提升度
2016	pressure_L, temperature_M, RH_H→ PM2.5_1	0.0516	0.6231	2.2487
	WD_N, RH_H, WS_L, temperature_M, pressure_L→ PM2.5_1	0.0211	0.6229	2.2480
	pressure_L, temperature_M, WD_N, RH_H→ PM2.5_1	0.0236	0.6161	2.2233
	pressure_L, WD_N, RH_H, WS_L→ PM2.5_1	0.0217	0.6044	2.1813
	temperature_M, WD_N, RH_H, WS_L→ PM2.5_1	0.0326	0.6034	2.1775
2017	pressure_L, temperature_M, WD_N, RH_H→ PM2.5_1	0.0213	0.7333	4.0658
	pressure_L, WD_N, RH_H, WS_L→ PM2.5_1	0.0205	0.7200	3.9919
	pressure_L, WD_N, RH_H→ PM2.5_1	0.0235	0.7128	3.9520
	pressure_L, temperature_M, RH_H, WS_L→ PM2.5_1	0.0303	0.7029	3.8972
	pressure_L, temperature_M, RH_H→ PM2.5_1	0.0566	0.6748	3.7415
2018	pressure_L, temperature_M, RH_H→ PM2.5_1	0.0358	0.3597	2.6770
	pressure_L, temperature_M→ PM2.5_1	0.0421	0.3178	2.3655
	pressure_L, temperature_M, WS_L→ PM2.5_1	0.0223	0.3176	2.3637
	pressure_L, WD_N→ PM2.5_1	0.0224	0.2992	2.2271
	temperature_M, RH_H, WS_L→ PM2.5_1	0.0227	0.2979	2.2172
2019	WD_N, RH_H, WS_L, temperature_M, pressure_L→ PM2.5_1	0.0326	0.7352	2.6559
	pressure_L, temperature_M, RH_H, WS_L→ PM2.5_1	0.0427	0.7276	2.6285
	pressure_L, temperature_M, WD_N, RH_H→ PM2.5_1	0.0361	0.7264	2.6242
	pressure_L, temperature_M, RH_H→ PM2.5_1	0.0767	0.7241	2.6159
	WS_M, pressure_L, temperature_M, RH_H→ PM2.5_1	0.0252	0.7129	2.5753

2016~2019	WD_N, RH_H, WS_L, temperature_M, pressure_L→ PM2.5_1	0.0208	0.6207	2.8579
	pressure_L, temperature_M, WD_N, RH_H→ PM2.5_1	0.0233	0.6056	2.7883
	pressure_L, temperature_M, RH_H, WS_L→ PM2.5_1	0.0319	0.5838	2.6879
	pressure_L, temperature_M, RH_H→ PM2.5_1	0.0540	0.5794	2.6677
	pressure_L, WD_N, RH_H, WS_L→ PM2.5_1	0.0223	0.5762	2.6526
含有 PM2.5_2 之規則				
年份	規則	支持率	信賴度	提升度
2016	temperature_H, WD_NW, RH_L→ PM2.5_2	0.0204	0.4851	1.3781
	temperature_H, WS_L→ PM2.5_2	0.0221	0.4850	1.3778
	temperature_H, WD_NW→ PM2.5_2	0.0262	0.4713	1.3389
	RH_L, WD_NW→ PM2.5_2	0.0296	0.4635	1.3166
	temperature_H, pressure_M, RH_L→ PM2.5_2	0.0401	0.4589	1.3038
2017	pressure_L, RH_M, temperature_M→ PM2.5_2	0.0269	0.5871	1.6845
	pressure_L, RH_M, WS_L→ PM2.5_2	0.0235	0.5722	1.6419
	WS_M, temperature_H, RH_L→ PM2.5_2	0.0247	0.5640	1.6182
	pressure_L, temperature_H, WD_NW→ PM2.5_2	0.0204	0.5559	1.5951
	temperature_H, WD_N→ PM2.5_2	0.0231	0.5519	1.5836
2018	WS_M, pressure_L, RH_M→ PM2.5_2	0.0226	0.7645	1.5335
	pressure_L, RH_M, temperature_M→ PM2.5_2	0.0218	0.7579	1.5204
	pressure_L, RH_M→ PM2.5_2	0.0535	0.7042	1.4126
	pressure_L, temperature_H, WD_NW, WS_H→ PM2.5_2	0.0203	0.6794	1.3628
	pressure_L, RH_M, temperature_H→ PM2.5_2	0.0306	0.6700	1.3440
2019	WD_W, pressure_H, RH_L, WS_H→ PM2.5_2	0.0223	0.5669	1.3295
	WD_W, pressure_H, RH_L→ PM2.5_2	0.0264	0.5566	1.3055
	pressure_H, RH_L, WS_H→ PM2.5_2	0.0323	0.5421	1.2715
	WD_W, pressure_H, WS_H→ PM2.5_2	0.0248	0.5371	1.2598
	WD_W, temperature_H, WS_H, pressure_L, RH_L→ PM2.5_2	0.0269	0.5315	1.2466
2016~2019	temperature_H, WD_NW, RH_L→ PM2.5_2	0.0220	0.5251	1.2924
	WS_M, temperature_H, RH_L→ PM2.5_2	0.0209	0.5191	1.2776
	temperature_H, WD_NW→ PM2.5_2	0.0301	0.5156	1.2690
	temperature_H, WD_N→ PM2.5_2	0.0224	0.5061	1.2456
	WS_M, pressure_L, temperature_H→ PM2.5_2	0.0229	0.5038	1.2398
含有 PM2.5_3 之規則				
年份	規則	支持率	信賴度	提升度
2016	pressure_H, RH_M, WS_L→ PM2.5_3	0.0209	0.3841	1.7172
	temperature_L, RH_M, WS_L→ PM2.5_3	0.0213	0.3645	1.6295
	pressure_H, RH_M→ PM2.5_3	0.0458	0.3612	1.6146
	temperature_L, pressure_H, RH_M→ PM2.5_3	0.0370	0.3521	1.5740
	temperature_L, RH_M→ PM2.5_3	0.0454	0.3440	1.5376

2017	temperature_L, WD_SW→ PM2.5_3	0.0293	0.4092	1.5813
	pressure_H, WD_SW→ PM2.5_3	0.0260	0.3979	1.5376
	temperature_L, pressure_H, WD_SW→ PM2.5_3	0.0211	0.3911	1.5113
	temperature_L, pressure_H, RH_M, WS_M→ PM2.5_3	0.0220	0.3907	1.5097
	WS_M, pressure_H, RH_M→ PM2.5_3	0.0237	0.3873	1.4967
2018	WS_M, temperature_L, WD_SW→ PM2.5_3	0.0240	0.4918	1.7744
	WD_SW, pressure_M→ PM2.5_3	0.0255	0.4514	1.6287
	temperature_L, WD_SW→ PM2.5_3	0.0370	0.4500	1.6236
	temperature_L, pressure_H, WD_SW→ PM2.5_3	0.0234	0.4227	1.5250
	temperature_L, pressure_M, WS_L→ PM2.5_3	0.0204	0.4192	1.5124
2019	pressure_H, RH_M, WD_SW→ PM2.5_3	0.0201	0.4681	2.1312
	temperature_L, RH_M, WD_SW→ PM2.5_3	0.0209	0.4378	1.9933
	WS_M, temperature_L, WD_SW→ PM2.5_3	0.0219	0.4192	1.9087
	temperature_L, pressure_H, WD_SW→ PM2.5_3	0.0229	0.4179	1.9026
	WS_M, pressure_H, RH_M→ PM2.5_3	0.0292	0.4142	1.8860
2016~2019	temperature_L, WD_SW→ PM2.5_3	0.0284	0.4104	1.6763
	temperature_L, pressure_H, WD_SW→ PM2.5_3	0.0200	0.4028	1.6452
	pressure_H, WD_SW→ PM2.5_3	0.0249	0.3922	1.6020
	WS_M, pressure_H, RH_M→ PM2.5_3	0.0212	0.3824	1.5622
	WS_M, temperature_L, RH_M→ PM2.5_3	0.0224	0.3705	1.5132
含有 PM2.5_H 之規則				
年份	規則	支持率	信賴度	提升度
2016	temperature_L, RH_M→ PM2.5_H	0.0408	0.3086	2.2935
	pressure_H, RH_M→ PM2.5_H	0.0386	0.3046	2.2635
	temperature_L, pressure_H, RH_M→ PM2.5_H	0.0320	0.3044	2.2625
	pressure_H, RH_M→ PM2.5_H	0.0320	0.2525	3.8703
	temperature_L, RH_M→ PM2.5_H	0.0320	0.2422	3.6126
2017	temperature_L, RH_M, pressure_M→ PM2.5_H	0.0242	0.4629	2.3184
	temperature_L, pressure_H, WD_SW→ PM2.5_H	0.0247	0.4567	2.2872
	temperature_L, RH_L→ PM2.5_H	0.0316	0.4504	2.2559
	temperature_L, pressure_H, RH_M, WS_L→ PM2.5_H	0.0253	0.4476	2.2417
	temperature_L, RH_M, WS_L→ PM2.5_H	0.0370	0.4444	2.2260
2018	pressure_H, RH_M→ PM2.5_H	0.0204	0.1393	1.7533
	pressure_H→ PM2.5_H	0.0357	0.1099	1.3837
	temperature_L→ PM2.5_H	0.0354	0.1097	1.3811
	pressure_M→ PM2.5_H	0.0347	0.1079	1.3578
	temperature_L, pressure_H→ PM2.5_H	0.0239	0.1076	1.3545



2019	RH_M→ PM2.5_H	0.0329	0.0915	1.6399
	temperature_L→ PM2.5_H	0.0277	0.0832	1.4898
	pressure_H→ PM2.5_H	0.0259	0.0783	1.4022
	pressure_M→ PM2.5_H	0.0237	0.0726	1.3010
	WS_L→ PM2.5_H	0.0224	0.0669	1.1979
2016~2019	temperature_L, RH_M→ PM2.5_H	0.0382	0.2485	2.1167
	temperature_L, pressure_H, RH_M→ PM2.5_H	0.0278	0.2412	2.0545
	pressure_H, RH_M→ PM2.5_H	0.0332	0.2394	2.0391
	WS_M, temperature_L→ PM2.5_H	0.0253	0.2046	1.7427
	pressure_H, RH_M→ PM2.5_H	0.0278	0.2007	3.1614

評估在各級雨量中規則發生的機率如表 4，依 2016~2019 年挖掘出的規則順序，依序進行雨量之統計分析，若為左斜框線表示在該年該規則不屬於前 20 名。發現含有 PM2.5\_1 之規則在各級雨量發生之機率由大到小為 confidence(precip\_H→ rule)、confidence(precip\_M→ rule)、confidence(precip\_L→ rule)，意指 PM2.5\_1 的規則較容易出現於雨量高的時候，而 PM2.5\_2、PM2.5\_3、PM2.5\_H 之 confidence(precip\_L→ rule) 高於 confidence(precip\_H→ rule) 及 confidence(precip\_M→ rule)，則代表在低雨量時較容易發生 PM2.5\_2、PM2.5\_3、PM2.5\_H 之規則，合於人的主觀意識。除此之外，confidence(precip\_H→ rule)、confidence(precip\_M→ rule) 數值普遍相近，confidence(precip\_L→ rule) 與其二較有落差，此代表在應用雨量因子時，可粗略分為下雨及未下雨。

表 4: 統計分析雨量對於各級 PM2.5 之影響

含有 PM2.5_1 之規則				
規則集合	年份	confidence(precip_x→ rule)		
		precip_L	precip_M	precip_H
RH_H, pressure_L, temperature_M, PM2.5_1, WS_L, WD_N	2016	1.99	2.6	4.52
	2017			
	2018			
	2019	2.73	9.23	9.66
	2016~2019	1.77	5.43	6.64
RH_H, pressure_L, temperature_M, PM2.5_1, WD_N	2016	2.16	3.55	5.85
	2017	1.54	7.71	10.94
	2018			
	2019	2.96	10.47	11.84
	2016~2019	1.89	6.35	9.48
RH_H, pressure_L, temperature_M, PM2.5_1, WS_L, WD_N	2016	2.73	4.73	7.18
	2017	2.22	11.29	14.06
	2018			
	2019	3.56	11.22	14.02
	2016~2019	2.65	8.91	10.95

含有 PM2.5_2 之規則				
規則集合	年份	confidence(precip_x→ rule)		
		precip_L	precip_M	precip_H
WD_NW, temperature_H, PM2.5_2, RH_L	2016	2.24	0.47	0.0
	2017	2.39	0.83	0.0
	2018			
	2019			
	2016~2019	2.39	0.37	0.0
temperature_H, WS_M, PM2.5_2, RH_L	2016	2.26	0.0	0.0
	2017	2.64	0.83	0.0
	2018			
	2019	2.48	0.25	0.0
	2016~2019	2.28	0.24	0.0
WD_NW, temperature_H, PM2.5_2	2016	2.83	1.18	0.53
	2017	3.23	1.65	0.31
	2018			
	2019			
	2016~2019	3.22	1.34	0.17
含有 PM2.5_3 之規則				
規則集合	年份	confidence(precip_x→ rule)		
		precip_L	precip_M	precip_H
PM2.5_3, WD_SW, temperature_L	2016			
	2017	3.17	0.28	0.0
	2018	3.95	0.66	0.68
	2019	3.22	0.25	0.31
	2016~2019	3.05	0.48	0.17
PM2.5_3, pressure_H, WD_SW, temperature_L	2016			
	2017	2.28	0.28	0.0
	2018	2.53	0.33	0.34
	2019	2.51	0.0	0.31
	2016~2019	2.16	0.24	0.09
PM2.5_3, pressure_H, WD_SW	2016			
	2017	2.81	0.28	0.0
	2018	3.1	0.33	0.34
	2019	3.03	0.0	0.31
	2016~2019	2.69	0.24	0.09

含有 PM2.5_H 之規則				
規則集合	年份	confidence(precip_x→ rule)		
		precip_L	precip_M	precip_H
RH_M, PM2.5_H, temperature_L	2016	4.44	1.18	0.0
	2017	8.36	5.23	1.25
	2018			
	2019			
	2016~2019	4.14	0.73	0.0
RH_M, PM2.5_H, pressure_H, temperature_L	2016	3.47	0.95	0.0
	2017	5.79	3.86	0.94
	2018			
	2019			
	2016~2019	3.01	0.61	0.0
PM2.5_H, pressure_H, RH_M	2016	4.21	0.95	0.0
	2017	6.25	3.86	0.94
	2018	2.21	0.0	0.0
	2019			
	2016~2019	3.59	0.67	0.0

### 伍、結論與未來研究

本研究致力於挖掘天氣因子及 PM2.5 隱藏之關聯規則，除了能藉以了解 PM2.5 及天氣樣態外，亦能提供未來研究分析 PM2.5 之基礎。在挖掘出的規則中，由各級 PM2.5 規則在歷年內的相似度，發現因 PM2.5\_2 的天氣模式較為混亂，導致篩選出的規則項目較不相似，PM2.5\_1 及 PM2.5\_H 則因天氣模式較為單一，其項目的相似度較高，在 PM2.5\_H 的規則中，風速及風向較少出現，推論對於高 PM2.5 而言，氣壓、溫度及濕度的影響遠高於風速及風向。根據統計分析，小於 12 的第一類 PM2.5 易發生在有降雨的時候，其他類的 PM2.5 則易發生於無降雨的時候。未來，擬加入其他因子進行規則挖掘，例如，加入人類活動，可分析人類活動產生的規則樣態，推論人類行為及 PM2.5 之關聯。

### 致謝

本研究部份成果為科技部計畫(MOST 107-2410-H-260-015-MY3、MOST 110-2420-H-163 -001 -MY2 之產出。

### 參考文獻

1. 林彥呈, 尹邦嚴, 戴榮賦, & 陳秉鈞 (2019)。以逆溫現象及測站間關聯分析預測全日每小時 PM2.5 沉積值。TANET2019 臺灣網際網路研討會, 326-332。
2. 徐玉兒 (2021)。多因子多尺規 PM2.5 資料分析—以埔里為例。台灣博碩士論文知識加值系統, 國立暨南國際大學資訊管理研究所, 南投縣。
3. Burnett, R. T., Pope III, C. A., Ezzati, M., Olives, C., Lim, S. S., Mehta, S., Shin, H. H., Singh, G., Hubbell, B., Brauer, M., Anderson, H.R., Smith, K.R., Balmes, J.R., Bruce, N.G., Kan, H., Laden, F., Prüss-Ustün, A., Turner, M.C., Gapstur, S.M., Diver, W.R. & Cohen, A. (2014). An integrated risk

- function for estimating the global burden of disease attributable to ambient fine particulate matter exposure. *Environmental health perspectives*, 122(4), 397-403.
4. Elliott, C. T., & Copes, R. (2011). Burden of mortality due to ambient fine particulate air pollution (PM<sub>2.5</sub>) in interior and Northern BC. *Canadian journal of public health*, 102(5), 390-393.
  5. Kim, M. J., Chang, Y. S., & Kim, S. M. (2021). Impact of Income, Density, and Population Size on PM<sub>2.5</sub> Pollutions: A Scaling Analysis of 254 Large Cities in Six Developed Countries. *International Journal of Environmental Research and Public Health*, 18(17), 9019.
  6. Tsai, Y.I., Sopajaree, K., Kuo, S., & Yu, S. (2015). Potential PM<sub>2.5</sub> impacts of festival-related burning and other inputs on air quality in an urban area of southern Taiwan. *The Science of the total environment*, 527-528, 65-79 .
  7. Wang, J., & Ogawa, S. (2015). Effects of meteorological conditions on PM<sub>2.5</sub> concentrations in Nagasaki, Japan. *International journal of environmental research and public health*, 12(8), 9089-9101.
  8. Wei, Z., Peng, J., Ma, X., Qiu, S., & Wang, S. (2019). Toward Periodicity Correlation of Roadside PM<sub>2.5</sub> Concentration and Traffic Volume: A Wavelet Perspective. *IEEE Transactions on Vehicular Technology*, 68(11), 10439-10452.
  9. Yang, Q., Yuan, Q., Li, T., Shen, H., & Zhang, L. (2017). The relationships between PM<sub>2.5</sub> and meteorological factors in China: seasonal and regional variations. *International journal of environmental research and public health*, 14(12), 1510.
  10. Zhang, B., Zhang, H., Zhao, G., & Lian, J. (2020). Constructing a PM<sub>2.5</sub> concentration prediction model by combining auto-encoder with Bi-LSTM neural networks. *Environmental Modelling & Software*, 124, 104600.

## **Mining Association Rules with Weather Variables and PM<sub>2.5</sub> in Puli**

Pei-Ru Wang

Department of Information Management, National Taiwan University of Science and Technology  
m11009113@gapps.ntust.edu.tw

Yu-Yuan Lin

Department of Information Management, China University of Technology  
melody376@gmail.com

Peng-Yeng Yin

Department of Computer Science and Information Engineering, China University of Technology  
pyyin97@cute.edu.tw

Rong-Fuh Day

Department of Information Management, National Chi Nan University  
rfday@ncnu.edu.tw

### **Abstract**

Air pollution is a topic that people much concern. PM<sub>2.5</sub> is easy to enter human body because its small particle diameter, and it may be harmful to human health. However, the concentrations of PM<sub>2.5</sub> depend on the location and climate, so this research tries to find out association rules between PM<sub>2.5</sub> and weather variables in Puli. The result can be used to observe the relationship between PM<sub>2.5</sub> in the area and help clarify the reasons for the fluctuation of PM<sub>2.5</sub>. This research discretizes the data of air pressure, temperature, relative humidity, wind speed, wind direction and PM<sub>2.5</sub> concentrations, and applies association-rule learning to design the mining process to find out significant rules, and use statistical analysis to realize the impact of rainfall on the occurrence of rules. The rules for PM<sub>2.5</sub> concentration between 12 and 23 is not significant, while the rule of PM<sub>2.5</sub> concentration less than 12 and PM<sub>2.5</sub> concentration greater than 35 are both salient.

Keyword: PM<sub>2.5</sub> 、 Weather Variables 、 Data Mining 、 Association Rule 、 Pattern Discovery