

機器學習於梅雨季南部強降雨判識研究

江宙君 朱容練 王潔如 徐理寰 王安翔 于宜強
國家災害防救科技中心

摘要

本研究利用支持向量機(Support Vector Machine, 簡稱SVM)為基底的機器學習模組進行台灣5、6月南部地區強降雨判識研究,其中台南、高雄以及屏東等三個縣市為主要研究區域。透過上述三縣市的測站雨量資料,搭配美國國家環境預測中心(National Centers for Environmental Prediction, 簡稱NCEP)所產製的氣候預報系統重分析資料(Climat Forecast System Re-Analysis, 簡稱CFSR),進行不同延時之強降雨判識模組建置與校驗工作。從降雨分析顯示,日雨量達40mm以上的區域多好發於南部山區。因此模組的建置,也以此區域為主。結果顯示,較短延時之預報模組,其技術得分累計有最好的結果。而在綜合表現上,則是以前一天12Z觀測資料(12HR)和前兩日00Z觀測資料(48HR)所建立之訓練模組有較好的表現。此外,從2018年5、6月梅雨季的測試結果發現,12HR與48HR模組平均正確率為0.76與0.68,平均陽性預測率為0.66與0.52,平均可偵測機率則為0.60與0.62。透過上述初步測試結果可知,以機器學習方式所建構之模組,於台灣南部地區梅雨季強降雨判識上,具有一定的掌握能力。

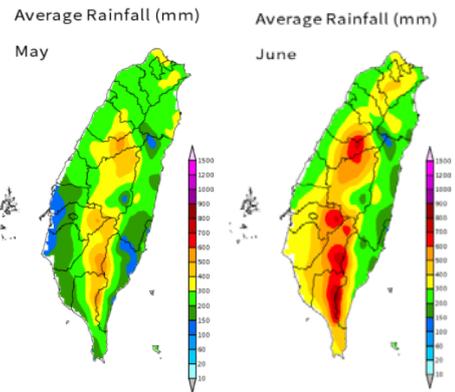
關鍵字：SVM、機器學習、梅雨

一、前言

每年五、六月於台灣附近所形成的梅雨鋒面,是台灣高衝擊的天氣系統,也是國內學者極為關注的研究課題(陳, 2007; 陳等, 2007)。梅雨鋒面移動過程所激發之中尺度對流系統,伴隨低層西南氣流,往往造成臺灣極大的災害衝擊,台灣西部低窪地區常因短延時強降雨而有積淹水事件的發生,中南部山區則容易因長延時強降雨而有發生坡地災害的可能;期間所造成的農漁業等經濟損失則有數億至數十億元(蘇等, 2017; 陳等, 2020)。

從五、六月雨量氣候值(圖一)可發現,台灣五、六月雨量較大區域集中於中南部地區。雖然數值模擬對梅雨鋒面已有一定的掌握能力,但是在鋒面強度、移行速度及地形效應的表現上,仍有進一步研究的必要(Ma et al., 2019)。近年來,隨著電腦軟硬體的迅速發展,加上高品質氣象觀測資料與高解析度模式資料的產製,使得人工智慧(Artificial Intelligence, 簡稱AI)技術於極端天氣的應用上變得可行(McGovern et al., 2017),為梅雨課題提供了另一個研究的方向。

有鑑於此,本研究將利用機器學習方法,進行南部地區(台南、高雄以及屏東)強降雨判識模組的建立與分析。本文分為四大部分,第一部分為前言,第二部分說明使用的資料與研究方法,第三部分為判識結果分析,最後為結論。



圖一、五、六月月雨量氣候值

二、資料與研究方法

本研究使用美國國家環境預報中心(National Centers for Environmental Prediction, NCEP)之全球氣候預報系統再分析資料(Climat Forecast System Reanalysis, CFSR)(Saha et al. 2010; Saha et al. 2014),輸入訓練資料於機器學習方法中,進而求得台灣南部地區梅雨季強降雨判識結果。

研究資料CFSR資料每日有四筆,分別為00Z, 06Z, 12Z以及18Z。垂直上共有37個標準層(1000, 975, 950, 925, 900, 875, 850, 825, 800, 775, 750, 700, 650, 600, 550, 500, 450, 400, 350, 300, 250, 225, 200, 175, 150, 125, 100, 70, 50, 30, 20, 10, 7, 5, 3, 2, 1(hpa)),每層提供五種變數,分別為重力位高度(HGT)、溫度(T)、濕度(RH)、東西風場(U)、南北風場(V),資料解析度為0.5度。本研究使用資料長度從

1989年至2017年，共29年資料。除此之外，本研究也搭配中央氣象局每日觀測測站降雨進行分析。

本研究使機器學習方法為支援向量機(Support Vector Machine, SVM)，SVM屬於機器學習類型中監督式學習的方法，主要架構由核心函式(Kernel Function)與超平面(Hyperplane)兩項組成，這也是影響 SVM 模型優劣的好壞。SVM方法將資料轉換到特徵空間，找出不同類別的超平面(hyperplane)，使超平面距離最大，以區分出不同類別，此類學習方式過程稱為結構化風險最小誤差法(Structural Risk Minimization, SRM)，期望讓分類器在期望誤差中找到最小值。

根據氣象局測站南部地區(台南、高雄及屏東縣市)日雨量資料中(表1)，排除日雨量0mm之兩日最大日雨量為688公厘，最小為0.5公厘，平均為30.14公厘，全資料中位數為11公厘，第三分位為34.5公厘，故將門檻值訂為10公厘與40公厘。另外大雨標準為日雨量80公厘，故增加80公厘為門檻值。

本研究使用NCEP全球再分析資料之氣象因子做為資料輸入元，利用SVM訓練與建立模型，進行判識南部地區(台南、高雄及屏東縣市)日雨量達強降雨之門檻值(圖二)。門檻值首先以10mm/day進行篩選，大於10mm/day的個案中，再分為大於40mm/day及小於40、大於10mm/day的個案(後簡稱為10-40mm/day門檻值)。將判斷是否小於80、大於10mm/day或是大於80mm/day的門檻值，稱為10-80mm/day。而是否小於80、大於40mm/day或是大於80mm/day的門檻值，稱為40-80mm/day門檻值。



圖二、流程示意圖

表1、1989年至2018年五、六月日雨量統計表

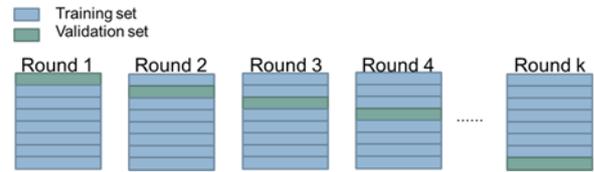
DATA:1989-2018 5-6月 日雨量>0.0					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.5	2.5	11	30.14	34.5	688

三、結果分析

將1989至2017年5、6月份資料為訓練資料，其資料組合10mm門檻值，無/有個案比例為45%與55%，10-40mm為48%與52%，40-80mm比例為56%與44%(如表2)。驗證方式使用交叉驗證方法(k-fold cross validation)，其方法為避免因依賴某一特定的訓練和測試樣本資料而產生偏差，方法如圖三所示，本研究使用10次交叉驗證(10 k-fold cross validation)進行校驗與分析(Su et al., 2018)。

表2、1989-2017年南部地區門檻值個案數與比例

	10 mm		10-40mm		40-80mm	
	N	Y	N	Y	N	Y
筆數	859	1032	492	540	301	239
比例	45%	55%	48%	52%	56%	44%



圖三、交叉驗證示意圖

本研究測試7個判識模組。針對南部地區目標日強降雨之判識，分別使用當日當地時間8時分析場(00Z,後稱00HR)、前一日當地時間晚上8時(12Z,後稱12HR)、前一日當地時間8時分析場(00Z,後稱24HR)、前二日當地時間晚上8時(12Z,後稱36HR)、前二日當地時間8時分析場(00Z,後稱48HR)、前三日當地時間晚上8時(12Z,後稱60HR)、前三日當地時間8時分析場(00Z,後稱72HR)，各自進行模組訓練，共訓練7個模組，針對不同預報時段之模組進行校驗，分析其預報可信度。

針對不同預報時段進行判識，以交叉驗證方法進行校驗，分別對於正確率(Accuracy, Acc)、陽性預測值(Positive predictive value, PPV)、陰性預測值(Negative predictive value, NPV)、可偵測機率(Probability of Detection, POD)以及預兆得分(Threat Score, TS)進行分析，將三個門檻值之平均校驗結果如圖四所示。



圖四、機器學習判識南部地區強降雨平均校驗結果

分析不同預報時段的校驗結果可發現，在正確率與陽性偵測值隨判識目標日與輸入之分析場的時間越久，校驗分數隨之遞減。00HR至三天前之72HR的七個模組判識之校驗能力中，正確率約為六成至七成，陽性預測值的結果也約在六成至七成。而在

可偵測機率中大致約為六成，其中在兩天前的36HR的分析場所判識之結果，在可偵測機率中較其他分析場來說略為不足，約為56%。這樣的結果，導致綜合分數中在前兩天12Z分數最低。綜合正確率、陽性預測值、陰性預測值、可偵測機率以及預兆得分五個校驗分數中，表現最好為當日00Z之結果其次分別為12HR、前兩日48HR。考慮實際應用的時效性，將測試12HR與48HR的分析場所建立之模組進行平行測試。

利用2018年5、6月資料進行平行測試，將訓練模組中12HR與48HR的模型進行測試。測試結果如表3所示。12HR模組中，三個門檻值正確率在七成至八成，平均正確率為七成六，陽性偵測值在八成至九成，平均陽性偵測值為八成一。在48HR模組中，正確率在六成至七成，平均為六成一，陽性偵測值在三成五至八成，平均陽性偵測值為五成，其結果與交叉驗證結果近似。在48HR模組中，在10-40mm和40-80mm的結果略為落後。

表3、平行測試之校驗分數

	Accuracy PPV			Accuracy PPV	
10mm_12hr	0.73	0.80	10mm_48hr	0.70	0.79
40mm_12hr	0.68	0.52	40mm_48hr	0.60	0.42
80mm_12hr	0.86	0.67	80mm_48hr	0.75	0.35
平均_12hr	0.76	0.66	平均_48hr	0.68	0.52

四、結論

本研究使用全球分析場氣象因子作為輸入因子，利用機器學習之SVM演算法，針對南部地區發生降雨的三個情境狀況(10mm、10-40mm以及40-80mm)進行資料訓練與判識模組建立。此外，本研究亦使用不同時間00HR至72HR之分析場，建立相同目標日之七個判識模組，了解由不同分析場所建立的模組判識能力。由比較結果可知道，隨著輸入分析場的時間距離目標日愈久，其判識能力隨亦之降低，平均正確率約在六成至七成，陽性偵測值亦落在六成至七成之間。

考慮訓練模組校驗預報能力與實際運作有效時段，本研究進行前12HR與前48HR分析場判識模組的平行測試。使用2018年5月至6月的分析場進行平行測試，測試結果與訓練之交叉驗證結果近似，表示其模組具有一致性，平均前12HR分析場判識模組之表現正確率為七成六，陽性偵測值為八成一；前48HR分析場判識模組平均正確率則為六成一，陽性偵測值為五成二。其結果顯示在距離目標日較近的12HR分析場判識模組表現較優，而雖然利用前48HR的分析場所建立的模組，在80mm降雨的判識上，具有一定的掌握能力，但是仍有進一步優化的空間。

參考文獻

- 陳泰然，2007：最近之梅雨研究回顧。 *大氣科學*， **35:4**， 261-286。
- 陳景森、陳宇能、林秉承、林沛練、劉哲伶、蘇炯瑞、彭文君，2007：梅雨季臺灣西南部豪雨之探討。 *大氣科學*， **35:4**， 287-304。
- 蘇元風、傅鏗漩、林又青、王俞婷、張志新、黃柏誠、于宜強，2017：20170601 豪雨災害報告。 *災害防救電子報*， **146**。
- 陳淡容、廖信豪、吳宜昭、于宜強，2020：2020年5月19-24日梅雨鋒面豪雨事件氣象分析。 *災害防救電子報*， **180**。
- Ma J., K. A. Bowley, and F. Zhang, 2019: Evaluating the Forecast Performance of the Meiyu Front Rainbelt Position: A Case Study of the 30 June to 4 July 2016 Extreme Rainfall Event. *Atmosphere*, **10(11)**, 648; <https://doi.org/10.3390/atmos10110648>
- McGovern, A., K. Elmore, D. Gagne, S. Haupt, C. Karstens, R. Lagerquist, T. Smith, and J. Williams, 2017: Using artificial intelligence to improve real-time decision making for high-impact weather. *Bull. Amer. Meteor. Soc.*, **98**, 2073–2090, <https://doi.org/10.1175/BAMS-D-16-0123.1>.
- Saha, S., Moorthi, S., Pan, H.L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer, D., Liu, H., Stokes, D., Grumbine, R., Gayno, G., Wang, J., Hou, Y.T., Chuang, H.Y., Juang, H.M.H., Sela, J., Iredell, M., Treadon, R., Kleist, D., van Delst, P., Keyser, D., Derber, J., Ek, M., Meng, J., Wei, H., Yang, R., Lord, S., van den Dool, H., Kumar, A., Wang, W., Long, C., Chelliah, M., Xue, Y., Huang, B., Schemm, J.K., Ebisuzaki, W., Lin, R., Xie, P., Chen, M., Zhou, S., Higgins, W., Zou, C.Z., Liu, Q., Chen, Y., Han, Y., Cucurull, L., Reynolds, R.W., Rutledge, G. and Goldberg, M., 2010: The NCEP climate forecast system reanalysis. *Bulletin of the American Meteorological Society*, **91**, 1015–1057.
- Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.T., Chuang, H.Y., Iredell, M., Ek, M., Meng, J., Yang, R., Mendez, M.P., van den Dool, H., Zhang, Q., Wang, W., Chen, M. and Becker, E., 2014: The NCEP climate forecast system version 2. *Journal of Climate*, **27**, 2185–2208.
- Su S.-H., J.-L. Chu, T.-S. Yo, and L.-Y. Lin, 2018: Identification of synoptic weather types over Taiwan area with multiple classifiers. *Atmospheric Science Letters*, **19**:e861. DOI: 10.1002/asl.861.

Identification of Southern Heavy Rainfall in Mei-Yu Season Using the Machine Learning Scheme

Chou-Chun Chiang Jung-Lien Chu Chieh-Ju Wang Li-Huan Hsu An-Hsiang Wang Yi-Chiang Yu
National Science and Technology Center for Disaster Reduction

Abstract

This study aims to conduct the machine-learning-based modules to identify heavy rainfall during Mei-Yu season for the region of southern Taiwan. The support vector machine (SVM) scheme is used for conduction. During the training period of modules conduction, predictors are taken from the Climate Forecast System Re-Analysis(CFSR) data, which is derived by National Centers for Environmental Prediction(NCEP). Meanwhile, the rainfall records observed by Central Weather Bureau over the region of southern Taiwan are used as the predictand for identification. The results reveal that the heavy rainfall (> 40 mm/day) happened frequently in the mountain area of southern Taiwan. Based on the result of modules evaluation for different lead time, modules with shorter lead time would give higher skill scores. Generally, the modules based on 12Z(12 hours lead) and 00Z(48 hours lead) perform better. To evaluate the skill of identification for the heavy rainfall of Mei-Yu season in 2018, the accuracy could reach to 0.76 and 0.68 for the modules of 12Z and 00z respectively. The positive predictive value and the probability of detection are also higher than 0.5 for both of the two modules. According to the preliminary result, the SVM-based module has the potential in the identification of heavy rainfall over the region of southern Taiwan during Mei-Yu season.

Keyword : SVM, Machine Learning, Mei-Yu