



打造韌性城市之 停電預測挑戰賽

成果分享暨專家座談

指導單位 經濟部

主辦單位



台灣電力公司

執行單位

dsp 智庫驅動



第三名

Frank & Louis

kao.boruei@gmail.com

louisjyli@gmail.com

github: <https://github.com/frankbrkao/PFPC>

摸索過程

- 問題: 預測颱風時, 各村里的停電戶數
- 攻防問題 (攻擊強度 vs 抗災能力)
 - 停電: 輸電網路的中斷 (電線脫落、電桿/電塔倒塌)
 - 攻防主體: 行政區 (縣市、鄉鎮、村里) 的輸電網路
 - 攻擊手段: 強風、豪雨、落雷
 - 抗災能力: 電網組成 (電桿類型)、耐久度 (設置多久、是否經常維護)、行政區地形地勢
- 找出攻防因素之間的關聯, 建立損害模型

期望的資料 (理想)

- 資料類型
 - 風勢、雨量、落雷數
 - 電網組成 (電桿類型與數量)
 - 電網損壞、維護、與更新紀錄
 - 區域特色 (e.g. 地形地勢)
 - 停電戶數
- 資料特性
 - 精細度: 村里
 - 須包含歷史、歷程資訊

整合資料 (實際)

- 風勢 (縣市/鄉鎮)

風速、最大陣風

- 雨量 (縣市/鄉鎮/村里)

最大日降雨量、最大 24h / 12h / 6h / 3h / 1h 雨量

- 電桿數 (村里)

3T桿, H桿, 木併桿, 木桿, 水泥併桿, 水泥桿, 用戶自備桿, 鋼併桿, 鋼桿, 電塔

- 戶籍數 (村里)、用電戶數 (村里)

解題策略 (1)

- 基本構想
 - 將損害模式近似的資料分群，個別建構模型
 - 使用 "現況條件近似" 的模型進行預測
- 建立模型
 - Random forest (thanks for Li Cheng En's sharing)
 - 分颱風 & 縣市 / 鄉鎮市區 建立模型
 - 計算預測目標之天候與歷史紀錄之相似性
(R, matrix distance, proxy::dist)
 - 選擇相似程度高之對應模型來預測

解題策略 (2)

- 資料清理 (缺失值)
 - 自動測站無 最大陣風
 - 部分自動測站無 最大平均風速
 - 部分縣市無局屬氣象測站
- 模式調整 (e.g. random forest, set ntree=5000)
- 排除未曾停電之區域
 - 降低資料變異性, $\sim 24\% = 1890/7851$
- 設定上界 (依據歷史資料)
- 蒐集解答 (比賽專用)

成功模型？

- 最高分 (有爭議, 無法實用)
 - RandomForest
 - 單一模型 (所有颱風/村里)
 - 混用前次最高分結果
- 建議模型 (尚未成功)
 - RandomForest
 - 分群建模

還可以怎麼做

- 分析輸電網路拓樸 (行政區 => 網路節點)
 - 目前作法只考慮行政區內的影響因素
 - 風災對電網節點的穩固性 (robustness) 影響
 - 電網節點損壞對停電數的影響
- 分損害程度 (零損害/ 輕度 / 重度) 建構模型
- 僅預測 "鄉鎮市區" 層級
- 其他模型: Xgboost, SVM
- 混和多模型預測結果

對比賽的建議

- 開放資料
 - 續辦比賽 (e.g. ILSVRC)
 - 以目前的成果為基礎 (資料集、分析方法)
 - 降低資料蒐集的比重
 - 評估評分指標的適用性
- 評分指標不一定代表準確率 (誤差)

評分指標不一定代表準確率

- 相同的誤差，高估的分數會 (遠) 高於低估
- 相同的誤差，預測值的數量級越低，分數越低
- 線性放大預測值與實際值，誤差線性放大，但分數不變
- 預測值與實際值的數量級越高，分數的鑑別性越低
- 無法處理 0
- 總差距較小，分數卻較低

評估指標測試

CM:

https://dspim.gitbooks.io/data-challenge-guide/content/morisita-horn_similarity.html

<https://dspim.gitbooks.io/data-challenge-guide/content/ting-dian-yu-ce-tiao-zhan-sai-ping-liang-shuo-ming.html#語言程式碼>

Case 1: Diff = 10, over-estimation > under-estimation

CM(20, 30) # 0.9230769 - overestimation

CM(20, 10) # 0.8 - underestimation

Case 2: Diff = 10, over-estimation >> under-estimation

CM(11, 21) # 0.8220641 - overestimation

CM(11, 1) # 0.1803279 - underestimation (small order of magnitude / close to zero)

Case 3: Diff = 1000, scale-up a, b, c, compare to case 1, scores are the same

CM(2000, 3000) # 0.9230769 - overestimation

CM(2000, 1000) # 0.8 - underestimation

Case 4: Diff = 1000, low discrimination with a large order of magnitude

CM(5000, 6000) # 0.9836066 - overestimation

CM(5000, 4000) # 0.9756098 - under-estimation

Case 5: Diff = 10, cannot handle zero

CM(10, 20) # 0.8 - overestimation

CM(10, 0) # 0.0 - underestimation

Case 6: Low diff but low score

CM(5000, 6000) # 0.9836066 - diff = 1000

CM(10, 20) # 0.8 - diff = 10

CM(10, 0) # 0 - diff = 10

資料集 (1)

- 颱風資料庫
<http://rdc28.cwb.gov.tw/>
- 氣象局 - 觀測資料查詢系統
<http://e-service.cwb.gov.tw/HistoryDataQuery/index.jsp>
- 用電戶數 - 台電 - 縣市住商用電資訊 - 各縣市村里售電資訊
<http://www.taipower.com.tw/content/announcement/ann01.aspx?BType=37>
- 政府資料開放平台 - 人口戶數資料 - 政府資料開放平臺 - 村里戶數、單一年齡人口
<https://data.gov.tw/dataset/32973#r0>
- 政府資料開放平台 - 有人/無人氣象測站基本資料
<https://data.gov.tw/dataset/45128>
<https://data.gov.tw/dataset/34517>
- 政府資料開放平台 - 台灣電力公司_電桿坐標及桿號
<https://data.gov.tw/dataset/33305>

資料集 (2)

- 政府資料開放平台 - 一年觀測資料-本局屬地面測站一年觀測資料
<https://data.gov.tw/dataset/33029>
- 政府資料開放平台 - 每月氣象-過去9年局屬地面測站每月氣象資料
<https://data.gov.tw/dataset/23827>
- 政府資料開放平台 - 落雷監測歷史資料
<https://data.gov.tw/dataset/9712>