

機器學習 讓你瞭解這世界上最晦暗的戰亂和貧困

A Life in Extreme Poverty

亞萌

整個人類似乎發展地很好：網路經濟、奧林匹克運動會、火箭發射、太空漫遊.....然而這樣繁榮的景象，並不是人類的全部，世界上還依然充滿了貧困和戰亂。就像這個在敘利亞戰亂剛剛遭遇空襲的5歲小男孩Omran Daqneesh，出人意料地安靜、麻木和茫然，不哭不鬧地擦掉自己臉上的血，而誰又知道他經歷了怎樣的殘酷！

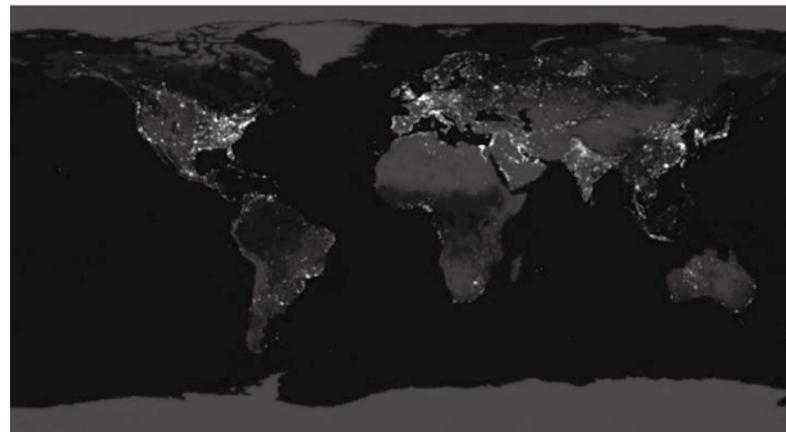
儘管有那麼多國際機構聲稱。要去要去幫助、去改變，然而由於無法瞭解基本資訊，很難對症下藥。

他們的夜晚連燈光都沒有

長久以來，研究人員會有一個評估某一地區的發展程度的指標：夜晚燈光的亮度和密度。但是這個方法有個局限，那就是它只能看到亮著的地區，而對於那些夜晚的黑暗之地，我們只能判斷它們大致是貧困的，而究竟貧困到什麼程度，就不得而知。



人們樂於分享幸福與富裕，而戰亂和貧困地區人的日常和生活水準，我們很難得知。



「非洲最窮的地方，是我們最關心的。但在晚上，那裡幾乎全部都是黑暗之地。」史丹佛大學的電子工程與電腦科學博士Neal Jean說道。

兩個村莊，一個靠近湖邊，一個靠近森林，在燈光分布圖裡，都可能是全暗的，但是它們臨近不同的自然資源，富裕程度不一樣。

而且，一般來說，瞭解一個地區的經濟水準，需要進行調察研究，而這一方法在戰亂和衝突地區是無法展開的，而恰巧這些地區就是世上最窮之地。

為什麼要知道哪裡是最窮的？

國際組織和當地政府決策者，最在乎這些數據。比如，一個公益性國際組織要對不發達的地區進行資金援助，資金是有限的，他們就需要確切知道，哪些地區的需求最懇切，所謂「把錢花在刀口上」。

對於政策制訂者而言，瞭解本國個地區的經濟發展狀況，是一項必須進行的工作，因為這樣才能制訂出針對性的發展方案。但事實是什麼呢？世界銀行數據顯示，在2000-2010年的十年時間裡，59個非洲國家中，有39個國家只進行了不到2次的貧困水準調查。

新的方法：更聰明、更便宜

史丹佛大學的這批以Jean為領導的研究人員，他們採取的是現在很流行的卷積神經網路的機器學習算法，所用的數據分白天的和晚上的，白天的數據來自衛星拍攝的地球高清圖像，夜晚的數據就是燈光分布圖。整個的算法分兩步。第一步叫做「轉移學習」（Transfer Learning），研究人員把五個國家的白天和夜晚的數據「餵給」系統，包括尼

日、坦桑尼亞、烏干達、馬拉維、盧安達這五個非洲國家。電腦需要學習衛星照片上的物體，與夜晚燈光亮度的關係。

比如系統可以學習到：一片湖區和一片森林，他們在晚上應當是黑暗的，這與該地區是否貧困並無關聯；但是，一個房屋聚集的村莊，理應夜晚亮度比較高，如果這片區域是黑暗的，那麼極有可能就是貧困地區。

透過這樣的學習，系統將貧困地區的數據都提取出來。接下來，就進行第二步，叫做「嶺回歸模式」（Ridge Regression Model）。研究人員將經過真實調查的數據指標，再次「餵給」系統，比如世界銀行生活水準測量研究等。

這次，系統會把未經過調查統計的地區，大致的貧困程度預測出來。比如，系統已經知道一個10戶房屋的村莊A的夜晚大致燈光亮度，而調查報告裡有一個與A村地理位置相近，而且燈光亮度相似的10戶村莊B的資訊，報告顯示B村的家庭收入水準是1.9美元/天。那麼系統透過交叉對比，就可以判斷出A村與B村有著相似的經濟水準。

這個分兩步走的算法系統，比單一的依靠燈光預測的方法更為準確，在具體的實驗裡，其準確率達到81%-99%。

另外，由於該系統使用的數據，都是來自公共機構，比如衛星圖像和燈光分布圖，就是來自NASA，而調查報告來自於美國政府和國際組織，所以整個系統的花費並不昂貴，可供有需要的公眾和機構免費使用。

Jean表示，系統目前數據只是來自非洲的五個國家，下一步的工作就是用其他國家的數據來訓練，以繪製出全球貧困地圖，幫助政府和機構更好地減輕世界貧困端零件。