

出國報告(出國類別：進修)

運用機械學習理論於憂鬱症治療

服務機關：臺大醫院精神醫學部

姓名職稱：吳其炘主治醫師

派赴國家：美國波士頓

出國期間：2018.07.16~2019.07.15

報告日期：2019.08.12

摘要

哈佛醫學院近幾年來在機械學習於醫學應用的領域，有許多相當重要的研究。Ronald C. Kessler 教授本身是流行病學學者，有相當豐富的研究經驗，近幾年來也在機械學習的領域中，有著重要的研究成果，例如自殺的預測與憂鬱症的嚴重程度。因此這次的進修，主要是在Ronald C. Kessler教授的研究團隊中學習。

我於2018年7月16日至2019年7月15日於美國波士頓哈佛醫學院，學習機械學習在流行病學的應用，並進行相關的科學研究。期間主要是針對如何運用機械學習理論，建立個人化醫療的模式。這一年當中，分析的資料，是以臺灣健保資料庫為主。原本標的疾病是以憂鬱症的治療為主，但因為思覺失調症的資料較為單純，易於分析。因此在探索階段，決定以思覺失調症的治療做為分析的主軸，日後再運用這些分析方法，再擴展於憂鬱症的治療。

Kessler教授受託於世界衛生組織，進行全世界精神疾病流行病學調查，目前已經累積接近30個國家的精神流行病學資料。在進修期間，我也參與了全世界抗憂鬱藥的使用狀況的分析。這些資料有助於未來瞭解憂鬱症的治療。

目次

壹、目的	1
貳、過程	2
一、進修單位介紹.....	2
二、機械學習理論在流行病學的應用.....	2
三、思覺失調症的個人化治療.....	3
四、抗憂鬱藥之全球流行病學研究.....	7
參、心得	10
一、思覺失調症的個人化治療.....	10
二、抗憂鬱藥的全球流行病學研究.....	11
肆、建議事項	12

壹、目的：

機械學習理論應在醫學與流行病學的研究，是目前重要的研究方向。近幾年來，健康保險資料和電子病歷等健康相關的資料快速的累積，如何善加利用這些資料，是重要的課題。

機器學習是一種數據分析方法，可以依據資料的輸入自動建構出預測模型。與統計模型相比，機器學習具有靈活處理高維度變量，變項間的非線性關係的能力，以及對於各種資料分布的假設較為寬鬆。最近，機器學習演算法已成功運用於預測自殺，暴力行為，以及重度憂鬱症的持續性和嚴重性。如何使用人工智慧來輔助臨床決策支持系統，幫助臨床醫生提高預測治療成效的準確度，以及協助處方藥物，是當前重要的議題。

在精神疾病的治療中，目前已經有各式各樣不同的精神藥物被核准使用。許多專家也根據臨床試驗的結果和專家的共識，制定了廣為人知的治療指引。即使如此，仍有相當大比例的患者對治療沒有反應。因此許多人對於臨床試驗的結果，是否可以推廣到真實世界的病患中，產生了質疑與擔憂。其主要原因在於，真實世界的患者，往往有著複雜的內外科或精神科的共病症，他們也同時服用許多不同的藥物，這跟臨床試驗中，受試者的病況往往較為單純，有明顯的不同。因此，如何達到個人化醫療，根據每個人的狀況，選擇最佳的藥物，仍是目前醫學上的挑戰。

這次進修的目標是使用大型資料庫的資料，其中包含豐富的臨床資料。我們將運用機器學習算法來構建預測模型來預測治療效果。並依據患者個別之臨床特徵來選擇最佳藥物，以達到個人化醫學(personalized medicine)之目的。

貳、過程：

一、進修單位介紹

哈佛醫學院在公共衛生與藥物流行病學術域，執世界之牛耳。Ronald C. Kessler 教授為當代精神流行病學之大師，其主持之全世界精神流行病學調查研究，為精神醫學領域，最常引用之研究之一。Kessler 教授近幾年積極發展機械學習理論於醫學上的應用；Kessler 教授團隊過去針對美國陸軍的內部資料，利用機械學習理論，大幅提高自殺的預測準確度；也利用機械學習的模型，來預測憂鬱症的預後與療效，均已發表在精神醫學界的頂尖期刊。這一年的進修期間，就是在 Kessler 教授的研究室當中，進行資料的分析；同時解決分析過程中的總總困難。

二、機械學習理論在流行病學的應用

機械學習在影像辨識的領域，雖然有顯著的進展，然而這些新穎的演算法，是否能夠勝過傳統線性迴歸模式的分析，仍有許多爭議；不同的研究的結果都不太一樣。依據最近的統合分析(meta-analysis)，線性迴歸與機械學習演算法相比較，其實並沒有明顯的差異。但如果是運用集合學習的方式(stacking ensemble methods)，結集了各種演算法(包含線性迴歸模式)的預測結果，再進行第二次的學習，這集合學習模式，我們稱之為 superlearner，或 meta-learner。運用這樣的機械學習模式，幾乎必定會勝過任何單一的演算法。因此目前機械學習模式，利用集合學習的模式，可以達到更精準的預測，是無庸置疑的。

以 SuperLearner 的套件，其產生的結果往往如下：用不同的 algorithm (例如 elastic net regression, randomforest, generalized regression model) 產生的預測

值與真實質之間的差異 (mean square error), 可以發現集合學習所預測的結果, 其誤差是最小的。

Algorithm	Mean Squared Error	Standard error	Minimal	Maximal
SuperLearner	0.15778	0.018141	0.12556	0.20425
Discrete	0.16399	0.019065	0.12556	0.22289
SL.glmnet_All	0.15835	0.018638	0.12006	0.22289
SL.rf.better_All	0.14424	0.016878	0.12556	0.16361
SL.randomForest_All	0.1431	0.016695	0.12726	0.16039

機械學習在流行病學的應用, 除了可以讓預測更加精準之外, 也能夠讓統計推論(inference)更加正確。傳統的迴歸模式, 需要假設資料的分布情形(例如常態分或二項式分布), 並以此來進行推論。但目標最大概似估計法(Targeted Maximum Likelihood Estimation)的方式, 對於資料的分布則不做任何假設, 相反的, 運用機械學習的方式, 從資料中學習該資料有的分布。進而用最貼近資料本身的分布, 來進行推論, 而達到最正確的結果。

三、思覺失調症的個人化治療

(一) 背景

思覺失調症是一種慢性且嚴重精神疾病, 對個人、家庭、或社會都帶來沉重的負擔。思覺失調症的標準治療方法是使用抗精神病藥物。藥物治療可減少精神病症狀, 並防止復發。然而, 在新診斷的思覺失調症患者當中, 只有 60%的個案, 對第一種抗精神病藥治療有反應, 而達到顯著減輕症狀

的效果；其餘的個案可能會因為療效不佳，或副作用太多，而停止或是更換藥物。這些患者中大約 33%至 72%在第一年改變或停止治療。而如果沒有規律地接受抗精神病藥的治療，可能會增加復發的風險，並導致長期功能的退化。因此，臨床醫療上，需要找出治療首次發作的思覺失調症患者的最佳藥物，換句話說，即是制定個體化的最佳治療規則 (individual treatment rule，縮寫為 ITR)。

(二) 方法

1. 資料來源

這研究的資料來源，我們使用臺灣的國家健康保險資料庫。在 2005 到 2011 年期間，我們找到 3 萬多名首次發作的思覺失調症患者，並挑選 15 種最常用的抗精神病藥物。這 15 種藥物包含 amisulpride, aripiprazole, chlorpromazine, clothiapine, flupentixol, haloperidol, olanzapine, paliperidone, quetiapine, risperidone, sulpiride, thioridazine, trifluoperazine, ziprasidone, zotepine。我們用利用集合機器學習方法來開發 ITR。將全樣本中的 70%，歸為訓練樣本，用以建立模型。剩餘的 30%的樣本，做為驗證樣本。我們使用 SAS 9.4 (SAS Institute, Inc, Cary, NC, USA) 進行數據整理。而預測模型的機器學習算法的應用和統計分析，則是使用 R 套件 (<http://www.r-project>)。

2. 最佳個體化治療規則(optimal individualized treatment rule)的建立

我們定義治療成功為在治療開始的 12 個月內，(1)沒有任何住院的情況。
(2) 沒有任何換藥或加藥的狀況。

3. 病患特徵

我們利用患者的人口統計學和臨床特徵，包括可能影響治療結果的內外科與精神科的合併症。此外，我們收集了患者目前和過去使用藥物的資料，這些資料可能反映患者的精神和身體狀況。總體而言，我們的數據分析中包括 120 個患者特徵變量。

4. 模型的建立

我們建立了預測模式，用來估算每種抗精神病藥物對於每位病患的治療成功率。如果某一個抗精神病藥物的估計的成功率愈高，則該藥物治療對於那位病患的益處越大。因為有 15 種抗精神病藥物，所以我們建立 15 種抗精神病藥預測模型，得到了每位患者對於這 15 種藥物所預測的治療效果。我們使用 R 語言中的套件，用來產生預測模型，並且建立了最佳個體化治療規則的應用。這些套件可以計算每個變數與治療組別之間的交互作用，得到一個對治療效果的系數的總合，系數愈高的，表示這個藥物的治療效果愈佳。同時，我們利用 SuperLearner 來運算。為了提高效率並避免過度擬合 (overfitting)，在超級學習者中使用 Pearson correlation 來篩選變項，如果該變項與治療成功與否的關係，其 Pearson correlation 的 p 值 < 0.1 ，才會納入預測模型當中。

5. 治療效果的評估

我們估算如果全體病患接受最佳個體化治療規則的治療成功率。我們將使用目標最大概似估計法 (TMLE) 來估計治療效果。這是一種用在觀察性研究中，同時控制藥物組別與干擾因素中的關聯，以及與治療結果 (treatment

outcome)與干擾因素之間的關聯；利用這種雙重穩健(double robust estimation)的方法來得到因果效應推估，而將產生無偏差的估計。TMLE 可以結合各種機器學習算法來估計結果和暴露機制並最小化偏差。在此分析中，我們使用包括 18 個基本演算法的超級學習者來估計。分析用於 R 中的 tmle 套件。

演算法的描述	R functions in SuperLearner
Bayesian GLM	SL.bayesglm
Generalized additive model	SL.gam
Generalized linear model	SL.speedglm
Ridge	SL.glmnet (alpha=0)
Elastic net	SL.glmnet (alpha=0.25)
	SL.glmnet (alpha=0.5)
	SL.glmnet (alpha=0.75)
LASSO	SL.glmnet (alpha=1)
Support vector machine	SL.ksvm (kernel = "rbfdot")
Neural network	SL.nnet
Polynomial spline regression	SL.polymars
Random forest	SL.ranger
Extreme gradient boosting	SL.xgboost (max_depth=1, shrinkage=0.01)
	SL.xgboost (max_depth=1, shrinkage=0.1)
	SL.xgboost (max_depth=2, shrinkage=0.01)
	SL.xgboost (max_depth=2, shrinkage=0.1)
	SL.xgboost (max_depth=4, shrinkage=0.01)
	SL.xgboost (max_depth=4, shrinkage=0.1)

(三) 結果

我們發現如果病患的治療符合最佳個體化治療規則，其治療成功率為 52%；而我們所觀察到的治療(即一般的治療)為 45%。兩者的差距為 72% (p-value <.001)。如果是隨機分配治療藥物的話，其治療成功率為 44%，略低於觀察到的治療成功率(p-value=0.03)。這個結果意味著雖然臨床醫生可能知道一點如何選擇藥物，但其效果只比隨機分再好一點點。

我們檢視最佳個體化治療規則所建議的藥物，發現 aripiprazole 和 amisulpride 是最佳個體化治療規則最常推薦的藥物（分別為 32% 和 30%）。

四、抗憂鬱藥之全球流行病學研究

(一) 背景

近幾十年來，許多國家都發現抗憂鬱藥的使用率逐年增加。在美國，抗憂鬱治療的比率從 1996 年的 6% 上升到 2005 年的 10%。同樣，加拿大和歐洲國家也出現了增長趨勢。抗憂鬱藥使用增加的趨勢，其背後的原因相當多種。可能是由於新一代抗憂鬱藥的引入，抗憂鬱藥的治療適應症的擴展，以及大眾對於精神科藥物的接受度逐年上昇。

新一代抗憂鬱藥，包括選擇性血清素再回收抑制劑（selective serotonin reuptake inhibitors）和其他具有獨特藥理作用機制的新型抗憂鬱藥，通常在治療憂鬱症時，取代了過去常用的三環和四環抗憂鬱藥（tricyclic antidepressant）和單胺氧化酶（monoamine oxidase inhibitors）。抗憂鬱藥不僅被核准用於治療情緒障礙症，還被核准用於治療各種焦慮症。然而，TCA 和其他傳統的抗憂

鬱藥，仍然經常被使用。但用途經常是於標籤外(off-label)適應症，如睡眠障礙，身體症狀，如疼痛控制，尿失禁，性心理功能障礙或功能性胃腸道症狀。

儘管抗憂鬱藥的使用增加，但在憂鬱症的治療上，仍然有很大比例的憂鬱症患者沒有得到足夠的治療。藥物遵囑性仍然很低，大多數患者在沒有足夠治療時間的情況下，自行停用抗憂鬱藥。過去對於藥物遵囑性的研究，發現好的遵囑性，下列的因素相關，包括年齡較大，使用新一代的抗憂鬱藥物，伴有焦慮症或睡眠障礙的合併症。患者的態度和主觀經驗，只有少數研究探討。美國大型的全國流行病學調查（NCS-R）的研究發現，34%的情緒障礙患者因為覺得治療後，症狀變好而停用抗憂鬱藥；因為其他理由停藥的人中，有 30.8%覺得藥物沒有幫助而停止，13.8%因副作用而停止。另一項針對臨床樣本的研究，收集了 272 名患有重度憂鬱症的患者的資料，研究者發現約半數患者在 6 個月內停用抗憂鬱治療。其中，約 50%因感覺好轉而停藥，四分之一因不良反應而停藥。

(二)研究方法與結果

我們的研究，包含了 22 個調查研究(總共 20 個國家)，48,446 名參與者；同時，我們調查了 40 種以上的抗憂鬱藥。其中包含了各種抗憂鬱藥。我們將不同國家區分為高收入與中低收入國家。

比較不同的國家，我們發現在高收入國家，其抗憂鬱藥的使用盛行率約為 4.7%；然而在中低收入的國家，其抗憂鬱藥的使用盛行率僅 1.4%。這樣的差異，並不是因為不同國家的憂鬱症盛行不同所造成。我們只看各個國家的憂鬱

症的患者，也同樣發現高收入國家，接受抗鬱藥治療的比例，遠高於中低收入國家。

抗憂鬱藥使用模式，在不同的國家也有明顯的不同。高收入國家，較常使用選擇性血清素再回收抑制劑與其他新一代的抗憂鬱藥。但中低收入國家，使用三環抗鬱藥與其他的傳統藥物的個案，仍然是大多數。

另外，我們也發現服用抗憂鬱藥的原因也有所不同。在高收入國家中，使用抗憂鬱藥主要是用來治療憂鬱和焦慮症狀。然而在中低收入國家的人，則有較多人使用抗憂鬱藥來治療睡眠障礙、身體症狀、或疼痛。高收入國家的抗鬱藥使用者，有較好的藥物遵囑性，而中低收入國用的使用者，其自行停藥的比例較高。在探索患者自己停止用藥的原因，我們發現「感覺變好了」是停止服用藥物的最常見原因。中低收入國家比高收入國家的使用者，更容易因為「感覺變好了」而停止使用抗憂鬱藥。

這些研究說明了各國抗憂鬱藥的使用狀況不同。未來在建立抗憂鬱藥個人化治療時，要考慮到不同國家之間的差異。

參、心得

一、思覺失調症的個人化治療

據我們所知，這是第一個運用機器學習算法來研究首次發作思覺失調症患者個人化治療的研究。我們根據患者的臨床特徵和用藥史，所建立的預測模型，其準確率為 81%。同，我們的研究的結果，發現運用機械學習所建立的預測模型，其能夠增加 7% 的治療成功率。

我們所建立預測模型，最常建議使用 aripiprazole 與 amisulpride。這些發現與一些統合分析的結果一致。Amisulpride 是改善症狀和減少病患自行停藥的最有效的藥物之一。Aripiprazole 則是不會造成 QTc 延長和不會造成催乳素升高等，最安全的藥物。在訓練樣本中，aripiprazole，amisulpride 和 risperidone 的治療成功率也分別高，分別為 45.7%，45.4% 和 47.2%。值得注意的是，我們的預測模型可以確定每位患者的最佳藥物，模型推薦治療的粗治療成功率為 52%，優於任何一種抗精神病藥物。

然而，我們發現最佳個人化治療規則所建議的治療，與實際觀察到的治療，有著高度的不一致。在臨床治療中，因為沒有顯著的預測因子可以用來做為選擇抗精神病藥物的依據，因此大多數臨床醫生根據他們的經驗開立抗精神病藥。例如，具有鎮靜作用的抗精神病藥可能適合於焦躁激動的患者；那些具有代謝副作用的藥物，則不適用於身體肥胖的患者。但是，這些原則，幾乎沒有實證的證據來支持這些開藥的理論。相較之下，基於利用機器學習演算法所建立的最佳個人化治療規則，則是通過豐富的醫療記錄和眾多臨床變項建立的。該模型可以提供比臨床醫生經驗更準確的預測。

調整後的最佳治療平均治療效果僅有 7%，換算為需要被治療的病人數 (Number needed to treat) 為 14。所增進的療效雖然不大，但我們認為個性化治療將在公眾的水平上，將能產生大量益處。因為思覺失調症是一種常見且嚴重的精神疾病。此外，該預測模型具有幾個優點。首先，該模型是根據患者的人口特徵，內外科和精神科共病、以及同時使用藥物所建構的。這些資料都可以從目前的健保卡中擷取，不需要進一步萃取或與連結其他的資料庫。因此這個預測模型可以很容易地在目前的電腦系統中使用。其次，預測模型可以生成一系列抗精神病藥物治療的選項，這些選項可以按治療成功的機率來排序。醫生將能夠選擇自由選擇成功率較高的藥物。在臨床上的使用性相當方便。

二、抗憂鬱藥的全球流行病學研究

這個研究說明了各國抗憂鬱藥的使用狀況不同。臺灣的研究雖然未納入這次全球流行病學調查當中。但我們根據過在臺灣研究的結果，顯示臺灣的使用抗憂鬱藥的模式，比較接近高收入國家。我們過去的研究發現，臺灣抗憂鬱藥的使用盛行率從 2000 年的 3.2%，逐沖增加到 2009 年的 4.6%。其中抗抑鬱藥使用的盛行率在老年人（≥65 歲）中最高。同時我們過去也發現三環抗憂鬱藥的使用率逐年下降。而選擇性血清素再回收抑制劑與其他新一代抗憂鬱藥逐年上升。就抗抑鬱藥處方的適應症而言，我們發現抗抑鬱藥用於睡眠障礙和調節障礙的趨勢越來越明顯；然而，用於情緒障礙、焦慮症和所有非精神病症適應症的抗抑鬱藥使用率下降。

臺灣的使用抗憂鬱藥物的模式，愈來愈接近高收入國家。未來在建立抗憂鬱藥個人化治療時，要考慮到不同國家之間的差異。

肆、建議事項

一、研究本身的限制

在思覺失調症個人化治療的研究中，我們的研究結果應該謹慎地解讀。首先，因為健保資料庫所包含的資訊有限，我們沒有辦法評估患者的症狀嚴重程度、日常生活的功能、以及生活型態。我們認為治療藥物的變化和住院與否，多少反映了治療效果的好壞。然而，這些指標可能對於症狀減輕和功能恢復與否，等輕微的改變並不甚敏感。其次，我們用該預測模型的效果，在觀察研究中進行了測試。然而，有幾種可能的干擾因素可能影響了最佳個人化規則治療的療效的估計。我們雖然使用 TMLE 來消除干擾因素的影響；然而，無法測量的干擾因可能會影響了估計的結果。健保資料庫不包括一些重要資訊，例如社會經濟因素，患者的偏好、和症狀嚴重程度。因此，要驗證預測模型對首發思覺失調症患者的個性化治療的有效性，仍然需要臨床試驗來證實。第三，大多數預測模型並不完美，我們的預測模型的準確率為 81%。如果可以獲得更多重要的資訊，我們準確度應該還會再提高。然而，預測模型將因此而改變，所建議的抗精神病藥物，也可能會有所改變。最後，本研究僅包括首發思覺失調症患者。該模型不能推廣到慢性思覺失調症的患者或難治型思覺失調症的患者。是否可以在這些人群中最佳個體化治療規則需要進一步調查。

二、未來的方向

機械學習演算法仍然持續有新的進展，未來在從事流行病學的研究時，仍然要隨時注意機械學習理論的演進。另外，個人化醫療的研究，雖然有初步的成果。未來有兩個方向需要努力。一個是如果讓預測能力更加精準。目前的資

料少了一些重要的資料，將來可以利用文字探勘的技巧，從電子病歷中取得重要的資料；另外也可能從抽血得到生物指標，以及未來基因檢驗愈來愈普及，而找到更好的預測因子，來讓預測模型更加精準。另外，機械學習所建立的預測模型，雖然在觀察性研究中得到證實。但未來仍需要在臨床試驗中驗證其效果。