

科展作品說明書

科 別：普通科

組 別：電腦與資訊學科

作品名稱：探討遺忘機制對於神經網路學習的幫助

關 鍵 詞：droupout、遺忘機制、CNN(卷積神經網路)

編 號：電-01

作者: 21308李卓岳 21302許智禕 21328戴偉璿	指導老師: 資訊科 鄭景元 老師 生物科 陳馨 老師
---	----------------------------------

目次

壹、研究動機	- 1 -
貳、研究目的	- 2 -
參、研究設備及器材	- 2 -
肆、研究過程或方法	- 3 -
伍、研究結果	- 13 -
陸、討論	- 16 -
柒、結論	- 19 -
捌、參考資料及其他	- 20 -

摘要

資訊科技是在發展什麼?人類資訊科技一開始是在研究"如何記憶", IBM公司在1956年製造了史上第一個硬碟, 不過體積有兩台冰箱大小, 容量只有5MB。隨著時代的演進, 硬碟容量越來越大, 體積越來越小, 儼然已經發展到了一個地步; 然而, 就在這硬碟已經逐漸成熟的同時, 人類卻開始研究, **如何忘記**相信絕大多數的人不喜歡忘記的感覺。在出門前尋找半天, 就因為忘記外套放哪, 到了學校才捶胸頓足, 就因為又忘了要帶作業(是真的忘記了嗎?)。然而, "忘記"其實也有些好處, 遺忘某些事情, 其實可以幫助大腦運作得更順暢。

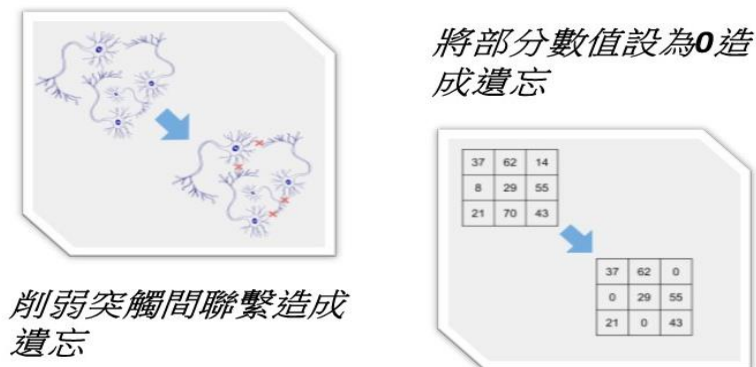
本篇研究主要目的就是研究"遺忘"對人工智慧之影響。修改一個AI模型的dropout值, 觀察其精準度是否有變化, 以及探討人類大腦與遺忘相關之機制。

壹、研究動機

我們三個人都是金庸武俠小說的愛好者, 尤其對"倚天屠龍記"感到印象深刻。在"倚天屠龍記"中"有一個經典的橋段:趙敏郡主率領大內高手前往武當山挑戰, 張無忌在學習太極劍法的時後, 將一切的招式全忘記了, 只留下劍法的核心, 於是我們想研究遺忘是否有能改善AI的預測結果。

從生物學的角度(AI的神經網路是模仿生物大腦做出來的)去看, 遺忘其實是一種大腦的自動清理機制, 將一個無意義的事情忘記, 有助於我們將注意力集中在更重要的事情上。同時, 遺忘也可以幫助大腦避免過擬合的狀況發生, 簡單講, 就如同張無忌學劍一樣。遺忘掉表面上特定的案例, 只記得案例背後的原理, 避免因為資訊太多而造成過擬合。

在查詢各式資料後, 我們發現人工智慧中也正巧有類似於遺忘的機制, 人工智慧在訓練的過程中, 會使用到含有極大量資訊的資料庫進行運算, 然而, 當AI訓練到一定程度後會發現, AI的訓練準確率和驗證準確率差距越來越大, 這時就能夠透過將部分資料設為0的方式(等同遺忘), 避免此情形的發生, 於是我們就想透過實驗觀察此方式究竟能對AI的預測帶來甚麼影響, 並以人類大腦遺忘的角度試著解析這個方法。



貳、研究目的

本研究主要目的為研究**遺忘機制能否改善AI的預測結果**，並且嘗試使用Excel呈現；其二為研究**人腦遺忘與機器學習遺忘之關聯性**。詳如下：

1. 針對不同資料庫建構AI模型，並修改AI模型之dropout值，觀察其對AI判斷結果的影響，並解釋原因
2. 比較人類與AI之記憶和遺忘機制的異同，並且思考如何改善

參、研究設備及器材

一、硬體部分：

ASUS B9450F 筆記型電腦、ASUS X415JA筆記型電腦、微星科技 GT76 Titan筆記型電腦

二、軟體部分：

Python(3.8.3 64bit)、CUDA(11.2)、微軟office Excel 2016、Spyder3、Tensorflow 2.1.0、keras 套件

肆、研究過程或方法



一、文獻探討:

根據國家教育研究院雙語詞彙、學術名詞暨辭書研究網，我們會遺忘一件事情其實是因為新的記憶和舊的記憶競爭，導致某一部分的記憶被抑制；從生物學的觀點，我們建立一新的記憶其實就是讓數個神經元之間增加突觸，簡單講，就是增加神經元之間的連結數量，讓人體神經元內的電訊號藉由稱為**神經傳遞物質**之化學物質傳遞至另一神經元，一旦由訊號所引發的動作電位強度超過閾值，該神經元就會被激發，從而產生新的電訊號及化學訊號，前往其他的神經元，形成傳遞網絡。而記憶的提取，也就是回憶，需要強化神經元間的連結，因此若連結阻塞就無法提取記憶，而產生遺忘的現象。用於儲存或回憶的神經基質被稱為「印跡」，當人經歷某種體驗而活化神經元，產生一系列物理/化學變化，就會形成一個印跡。隨後若獲得該體驗的相關線索時，會激活印跡，而引起記憶恢復。

二、尋找人類大腦與記憶及遺忘相關之機制

1.干擾理論

由於(長期)記憶有相同的提示，因此在提取記憶時，會互相干擾，形成遺忘現象。干擾分為兩種，一種為順攝抑制，為先前學習到的資訊對後來學習資料的記憶產生干擾；另一種為倒攝抑制，指後來的學習對先前學習資料的記憶產生干擾。

2.遺忘

人在記憶時，大腦神經元會建立新的突觸，用以傳遞電訊號或化學信號。遺忘的可能性有兩種，一為突觸的聯繫消失，二為突觸作用減弱。突觸聯繫消失的生理機制為突觸後細胞表面AMPA受體或NMDA受體的消失，導致無法接收訊號。突觸具有可塑性，亦即神經間的連結可以增強、減弱或重塑。突觸連結減弱為突觸具可塑性的表現之一。當訊息傳遞過程中產生的動作電位傳到軸突末端，會激發鈣離子通道，使鈣離子進入突觸後細胞，促使囊泡中的神經傳導物質被釋放。因此鈣離子通透性減少、囊泡釋放機率減少皆有可能使突觸連結減弱。

3.神經可塑性

神經可塑性可分為兩種，一種為結構性神經可塑性，另一種為功能性神經可塑性。結構性神經可塑性為樹突棘的消失或發生，而樹突棘是興奮性突觸的突觸後部位，其功能與型態對記憶重大影響。功能性神經可塑性涉及神經傳導物釋放、突觸後受體狀態等機制的改變，引發長期抑制作用或長期增強作用。

1.長期抑制作用

神經元突觸在強烈或長期的刺激後，會持續數小時至數天受到抑制。高濃度的 Ca^{2+} 為誘發長期抑制作用的必要條件之一，而引起 Ca^{2+} 信號來源為平行纖維與攀爬纖維。一般情況下平行纖維會先於攀爬纖維活化，兩者同時活化而釋放 Ca^{2+} 才會引起長期抑制作用。另一抑制原因為突觸後磷酸酶活化導致突觸AMPA受體(影響興奮性突觸傳遞)被內吞至突觸後Purkinje細胞，而降低對Schaffer側枝，亦即影響神經可塑性以保留記憶之軸突，所釋放之神經傳遞物-L-谷氨酸的敏感性，影響神經訊號之傳遞。

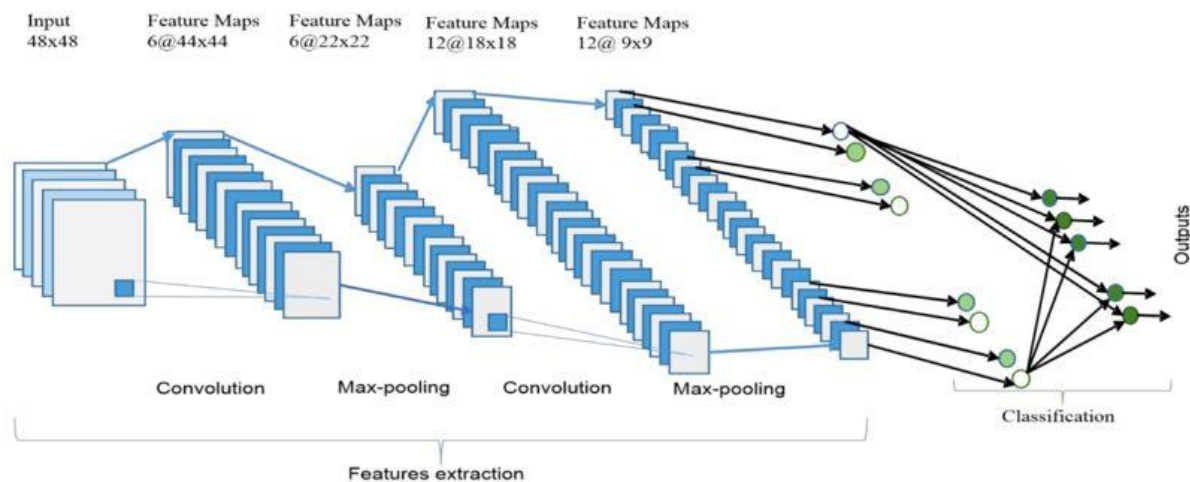
2.長期增強作用

此現象為同步刺激兩個神經元而使兩個神經元間信號傳輸持久的增強。蛋白激酶的持續活動能引發AMPA受體磷酸化而增強活性，使後續興奮性刺激產生更大的突觸後反應，此為長期增強作用的前期。長期增強作用的後期則是由基因表達及蛋白質合成引發的變化所誘導，此過程由長期增強作用前期激活的蛋白激酶之持續激發所引起。

三、名詞釋義

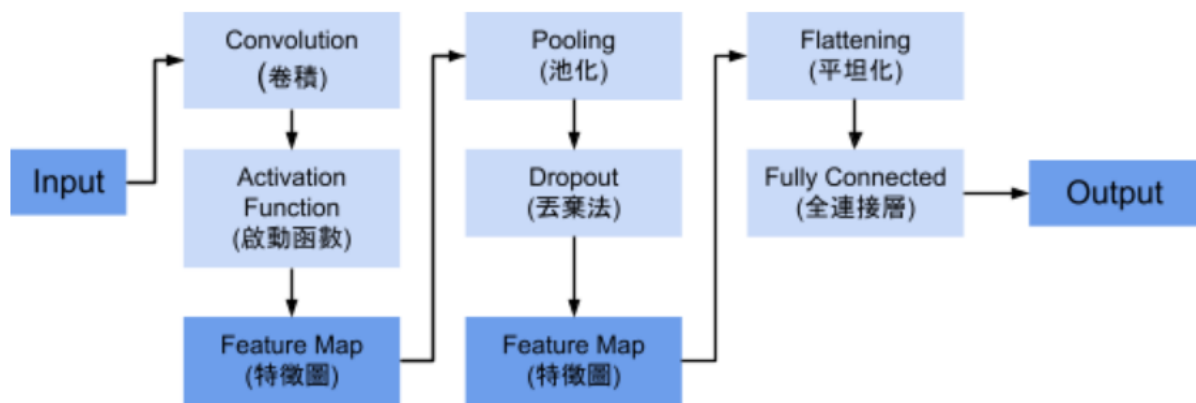
1.卷積神經網路(CNN)

卷積神經網路(Convolutional Neural Network,簡稱CNN)，為目前深度學習主要的研究領域之一，其於圖片判斷的領域上已超越了人類，卷積神經網路的基本架構為卷積層(Convolution Layers)和池化層(Pooling Layers)，大致架構如下圖：



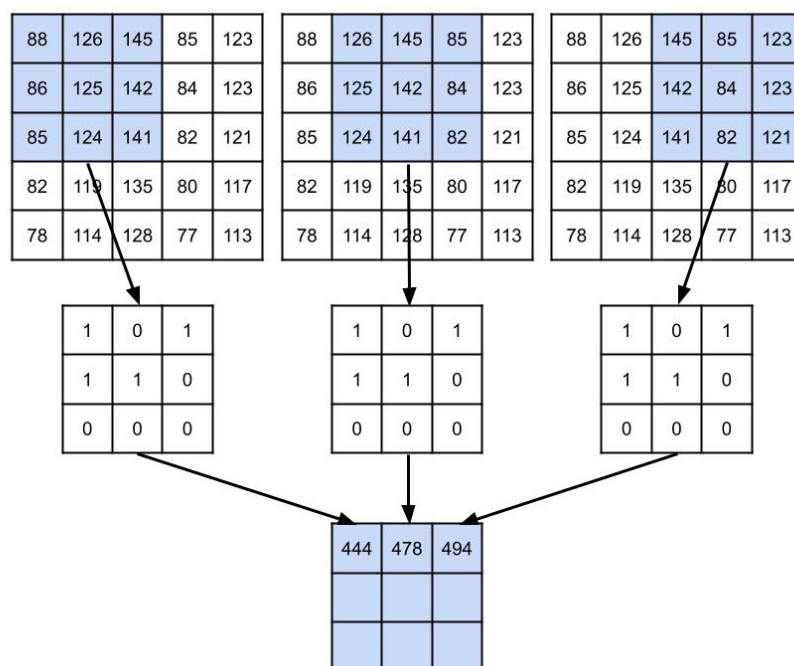
圖一、CNN架構

我們以 5×5 的輸入圖片和兩個 3×3 的卷積核來展示CNN的運算處理過程，下列是執行一次卷積，一次池化，一次丟棄和平坦化的卷積神經網路正向傳播流程圖：



圖二、CNN正向傳播流程圖

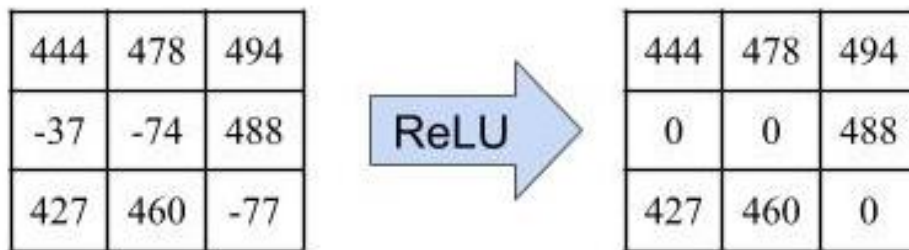
- **卷積層**：使用多個過濾器或稱卷積核，一步步掃描圖片並執行卷積運算來萃取出特定的特徵形成特徵圖，可以理解為過濾器就是此卷積層的權重，大致流程如下圖。



圖三、卷積層

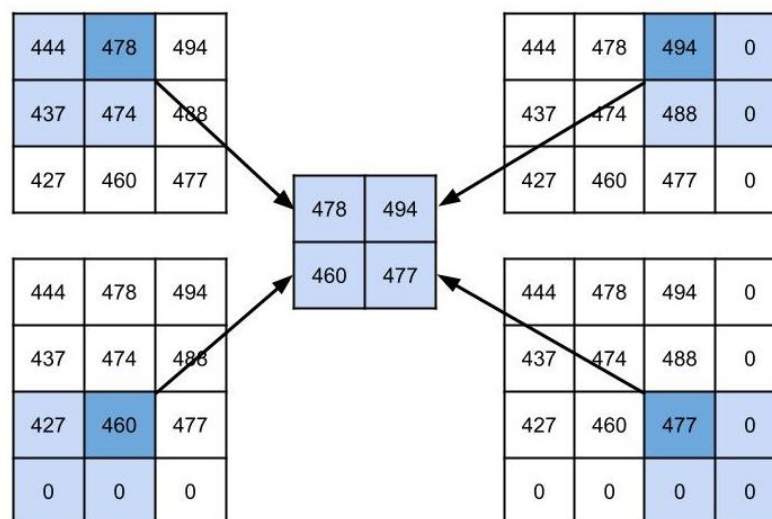
- **啟動函數**：我們模型使用ReLU啟動函數來輸出特徵圖如果數值小於0就設為0，否則維持原本的線性關係，ReLU定義如下所示：

$$ReLU(x) = \begin{cases} 0, & \text{if } (x < 0) \\ x, & \text{if } (x \geq 0) \end{cases}$$



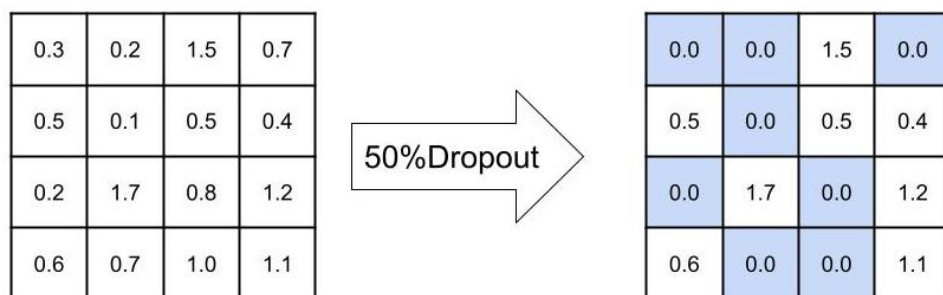
圖四、ReLU

- **池化層**：在池化層主要的目的為壓縮資訊，來獲得所需的重要資訊，其目的是讓卷積層專注於圖片中是否存在此特徵，而拋棄特徵所處的位置資訊，以下是執行Max Pooling的過程，即為取出四格中的最大值，大致流程如下圖。



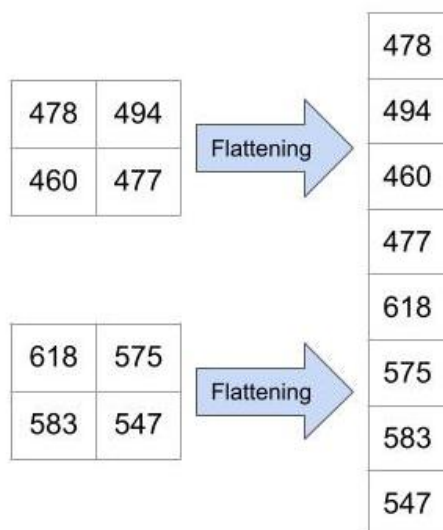
圖五、池化層

- **丟棄層**：假設我們新增一個50%的Dropout層，神經網路在訓練時，會隨機將池化層輸出的特徵圖中50%的資料設為0，如下圖所示：



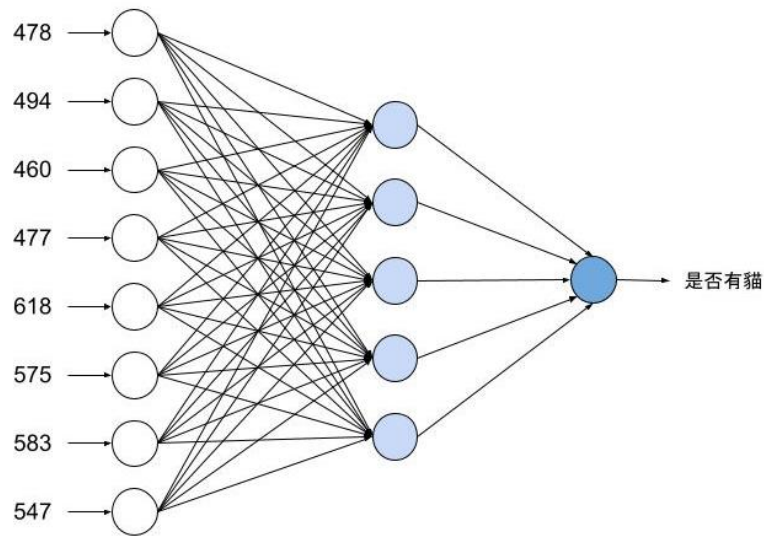
圖六、丟棄層

- **平坦化**：在計算出Dropout的特徵圖後，會將結果平坦化，讓原本的矩陣轉換成向量，以便送入全連接層的神經網路處理分類。



圖七、平坦化

- **全連接層**：將平坦化後的數據，送入多層感知器(Multilayer Perceptron,MLP)透過密集的神經元運算來獲得結果，並使用反向傳播法更新卷積層的權重，也就是卷積核(filters)，如下圖所示：



圖八、全連接層

(二)丟棄法(Dropout)

丟棄法(Dropout)是最有效和最常使用的神經網路常規化技術之一。神經網路的丟棄法主要是在訓練期間隨機丟棄(dropping out，把feature值設為0)，減少資料中所包含的特徵資訊量，丟棄法的核心想法是在layer的輸出值加入雜訊。

(三)熵(Entropy)

簡單講，**熵就是"有多亂"的概念**。熵的概念最早於1862年由魯道夫·克勞修斯以**融化的冰塊**提出，而資訊熵又與熱力學中的熵有些許不同，資訊熵(Information Entropy)有以下性質：

- 為接收每條訊息所包含資訊的平均量，與能傳輸的資訊量呈正相關
- 熵的單位取決於定義用到對數的底，單位有:bit、Sh、nat、Hart
- 熵可視為隨機變量的期望值

接下來解釋以上內容：

1.熵的概念：

熵為資訊的單位的概念。諸如質量的單位有公斤(kg)，長度的單位有公尺(m)，訊息的單位的概念為熵。不過由於訊息的不同，所使用的單位亦會不同(第二點會提到)。為何熵會被視為**包含資訊的平均量**呢？由於熵所代表的意義為**有多亂**，倘若一道訊息可以被精準預測，則此訊息存在的意義不大。在機器學習領域中，倘若一個用以預測貓咪圖像的模型都使用貓咪A的圖片來訓練，則容易產生過擬合的情型，訓練成熟後，這個AI就只能判斷**這個物體是否為貓咪A**。此時他可以準確預測貓咪A，此時的熵已經趨近於0，代表其已經擷取了貓咪A的所有特徵，鮮少出現判斷錯誤。因此再給他貓咪A的照片對於訊息的增加量不大。

但若換成貓咪B的照片給他呢？這個貓咪AI沒有看過貓咪B啊！此時的熵極大，代表他無法判斷貓咪B為貓咪，因此他需要修改神經網路中的權重。這就代表貓咪B圖像給他一大堆的資訊。

由以上的例子，我們可以得知：熵越大所傳遞的資訊量越多。

2.熵的計算

為了要將資訊熵量化，我們先給予資訊熵一個定義：**在一個事件中平均需要幾個問題才可以成功判斷出某一個條件**。即所需決策樹層數的期望值

有了定義後，我們可以歸結出以下幾個公式

$$\text{資訊熵 } H(x_i) = P(x_i) \cdot -\log_2 P(x_i)$$

其中， $P(x_i)$ 為一事件中某條件發生的機率，因此我們可以再歸結出以下公式：(此公式即夏農公式)

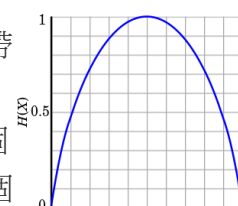
$$H(X) = \sum_i P(x_i) \cdot -\log_2 P(x_i)$$

舉一個實際的例子。假使要計算**拋擲一枚公正的硬幣(正反不同)**的熵，可以先列舉出所有可能的狀況以及機率：(1)正面 0.5 (2)反面 0.5，將其帶入計

$$H(X) = 0.5 \times -\log_2 0.5 + 0.5 \times -\log_2 0.5 = 1$$

算可得因此，此事件的熵為1bit。附圖為機率不同時的熵值，可發現當某一個機率為0或者是1的時候，熵值都為0。這是因為我們100%可以肯定某個事件是否發生，因此此事件的不確定性為0。(如圖九)

本研究中使用之損失函數(loss)為交叉熵(cross-entropy)



圖九、熵值

3.Cross-entropy(交叉熵)

交叉熵是使用資訊熵來評估兩組機率向量之間的差異程度，我們知道熵的大小與事件的混亂程度成反相關，因此當交叉熵越小，代表兩組機率向量越接近，以下為公式：

$$H(X,Y) = - \sum_{i=1}^n P(x_i) \log_2 P(y_i)$$

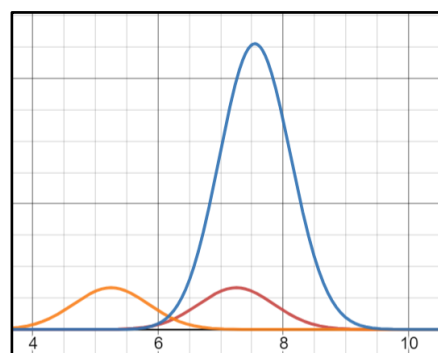
其中， $P(x_i)$ 為預測數值的機率分布圖， $P(y_i)$ 為實際數值的機率分布圖

交叉熵為計算實際機率分布與預測機率分布之間的差異，如圖十：

圖片來源：<https://www.desmos.com/calculator/zytm2sf56e?lang=zh-TW>

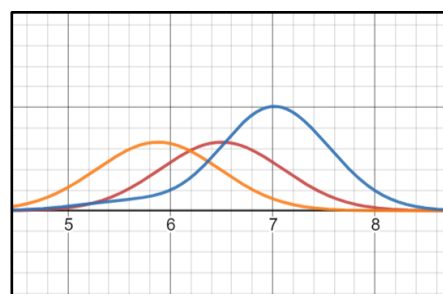
圖中的橘色線條為預測的機率分布，紅色線條為實際的機率分布，而藍色線條則是在某一個x下的交叉熵計算結果。真實的交叉熵則為藍色線與x軸夾出的面積。

由圖可知，當兩者的機率分布相差極大時，交叉熵亦會變大，交叉熵的數值即為損失函數的回傳值。再利用反向傳播修改權重，以增加準確率。



圖十、實際機率分布與預測機率分布之差異-1

再由圖十一可發現，當兩者機率分布相差不大時，交叉熵的數值亦變小，此即代表已經可以大概預測，因此交叉熵的資訊量降低，回傳後對於權重的修改較少。



圖十一、實際機率分布與預測機率分布之差異-2

有了我們的夏農公式後，我們可以進行極值的推導。已知，z軸座標與x,y座標的關係為：

$$H(X,Y) = - \sum_{i=1}^n P(x_i) \log_2 P(y_i)$$

此時，我們只需要推導出藍色線與x軸所夾的面積之極值，也就是交叉熵的極值，就可以知道在怎樣情況下的交叉熵會回傳最小的損失函數，瞭解怎樣的遺忘數值可以達到最佳的結果。

四、研究過程

1.研究AI遺忘機制

(1)建構CNN模型

我們以CNN作為這次主要的實驗模型，以keras套件中內建的mnist,cifar10,fashion_mnist資料庫做為輸入資料。

(2)修改CNN模型

在初步完成CNN模型後套過修改卷積、池化層數、激活函式和神經元數目，建構出最適合此資料庫之模型後，進行下一步的實驗。

Algorithm: Interpreting Mnist through CNN to classify handwritten digits

Input: Mnist Datasets

Output: Numbers predicted by CNN

Let: X_train **be** Mnist Datasets' s images ,

Y_train **be** Mnist Datasets' s labels

Update: X_train **by** Orthogonalization

Set: X_train to 2D Tensor (samples, features)

Building: model **by**

Layer (type)	Output Shape	Param #
conv2d	(None, 26, 26, 32)	320
max_pooling2d	(None, 13, 13, 32)	0
dropout	(None, 13, 13, 32)	0
conv2d_1	(None, 11, 11, 64)	18496
max_pooling2d_1	(None, 5, 5, 64)	0
dropout_1	(None, 5, 5, 64)	0
flatten	(None, 1600)	0
dense	(None, 512)	819712
dropout_2	(None, 512)	0
dense_1	(None, 10)	5130

Initialize training

For epochs = 9 **do**

fitting X_train and Y_train

End for

Get: Numbers predicted by CNN

Return Numbers

(3)調整dropout值

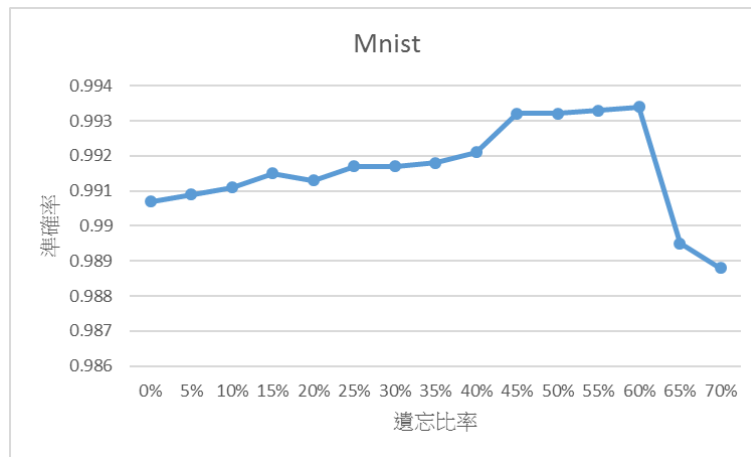
在建構出模型後，我們透過調整dropout層的遺忘比率來進行實驗，每個資料庫的模型都由**遺忘5%到遺忘95%**，紀錄每個不同的遺忘比率模型訓練完後的驗證損失值，驗證準確率，訓練損失值、訓練準確率，來進行分析。最終找出最合適（即準確率最高）的比例。

伍、研究結果

(一)AI遺忘機制之影響

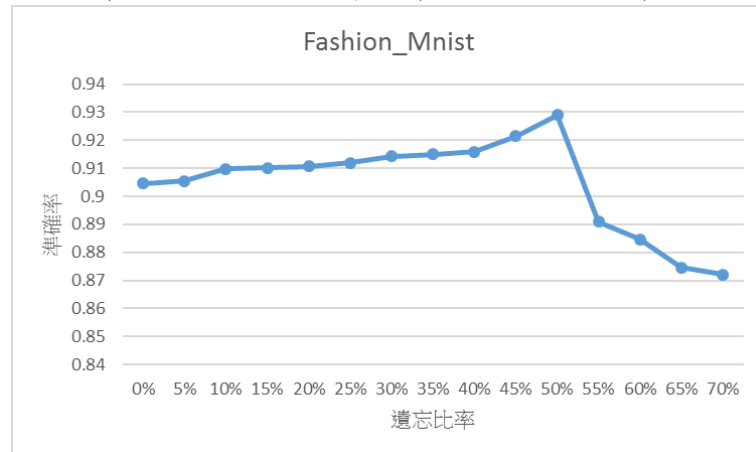
我們將CNN模型不同遺忘比率的驗證準確率畫成圖表後可以發現，在遺忘比率不大於50%時，AI的預測準確率會隨著遺忘比率的增加而提升，而在遺忘比率大於50%之後，會因為遺忘掉的資訊太多造成訓練AI時無法順利地收斂找出通式而造成準確率大幅下降，以下分別是Mnist，Fashion_Mnist，cifar10遺忘比率與準確率關係之折線圖：

下圖為Mnist資料庫的遺忘比率與準確率關係之折線圖，Mnist資料庫由來自 250 個不同人手寫的數字構成，含有60,000 個訓練樣本和10,000 個測試樣本，是AI深度學習的經典資料庫，觀察下圖可知，在未遺忘掉任何資訊時，準確率已經到達0.99，而遺忘比率到達50%時準確率會提升到0.993，可以清楚的觀察到在遺忘比率不大於50%時，準確率有隨著遺忘比率的增加而提升之現象。



圖十二

下圖為Fashion_Mnist資料庫的遺忘比率與準確率關係之折線圖，Fashion_Mnist是一個涵蓋10個種類的服飾正面灰階圖片，含有60,000 個訓練樣本和10,000 個測試樣本，很適合用來嘗試各式各樣的model，觀察下圖可知，在未遺忘掉任何資訊時，準確率也到達0.90的高準確率，雖然如此但是當遺忘比率到達50%時準確率會提升到接近0.93，可以更清楚的觀察到在遺忘比率不大於50%時，準確率有隨著遺忘比率的增加而提升之現象。



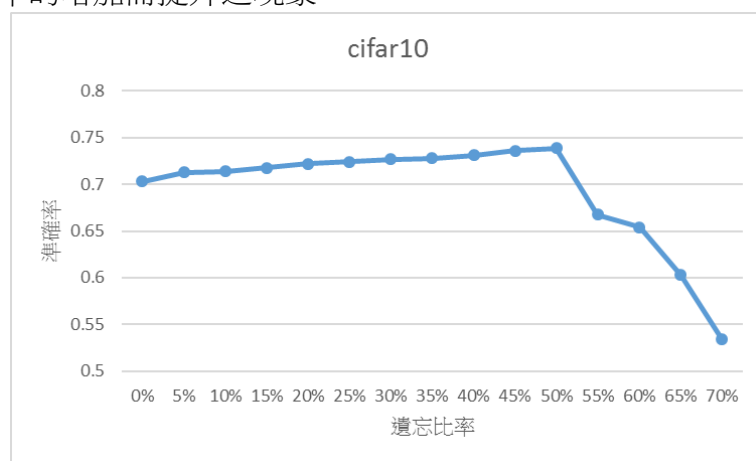
圖十三

圖十四為Cifar10資料庫的遺忘比率與準確率關係之折線圖，Cifar資料集一共包含60 000張圖片，其中50000張圖片作為訓練集，10000張圖片作為測試集。

Cifar-10 的所有圖片被分為 10 個類別 (以 0~9 數字作為 Label 之編碼)：

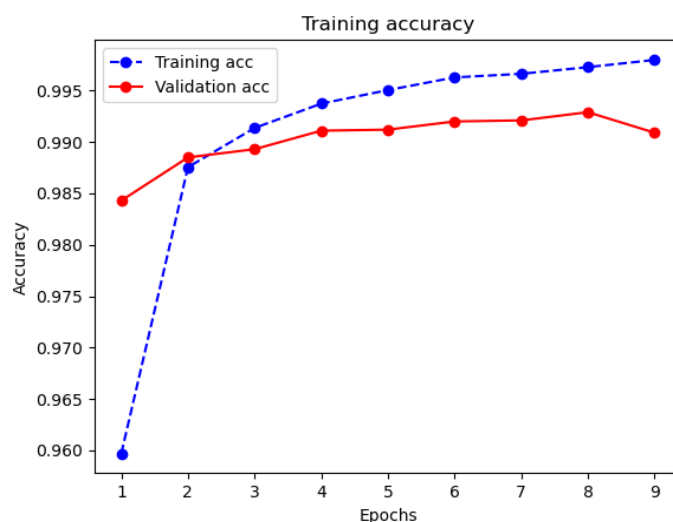
0 : airplain (飛機)	3 : cat (貓)	6 : frog (青蛙)	9 : truck (卡車)
1 : automobile (汽車)	4 : deer (鹿)	7 : horse (馬)	
2 : bird (鳥)	5 : dog (狗)	8 : ship (船)	

觀察下圖可知，在未遺忘掉任何資訊時，準確率到達了0.7左右，而當遺忘比率到達50%時準確率則會提升到接近0.74，可以清楚的觀察到在遺忘比率不大於50%時，準確率有隨著遺忘比率的增加而提升之現象。



圖十四

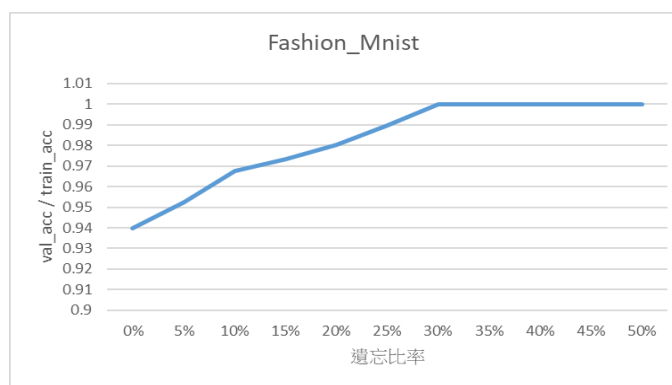
除了發現Dropout能夠提升AI準確率之外，我們也發現Dropout能夠避免過擬合的狀況發生。以Mnist資料庫Dropout值為0%之模型作為範例:觀察圖十五可發現，訓練準確率和驗證準確率在訓練次數(Epochs)達兩次之後的趨勢差距越來越大，這時就是AI模型發生了過擬合的狀況。而會發生過擬合則是因為驗證資料集的資料和訓練集資料不同，在AI記住太多訓練集資料中不必要的資訊後會導致AI在分析驗證資料集中沒看過的圖片時，錯誤率增加。



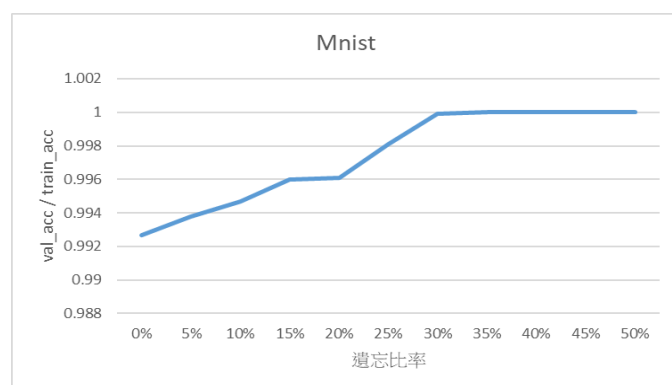
圖十五

而在我們多次實驗後發現，當dropout數值到達30%之後就能將訓練準確率和驗證準確率的趨勢差距縮到最小，但遺忘超過50%的資訊將會導致AI在訓練過程中所需要的特徵圖資訊量不足而無法透過訓練找出正確的判斷通式，導致AI的訓練準確率降低，更連帶影響到AI在驗證的預測預測不準確，導致驗證準確率也大幅降低。

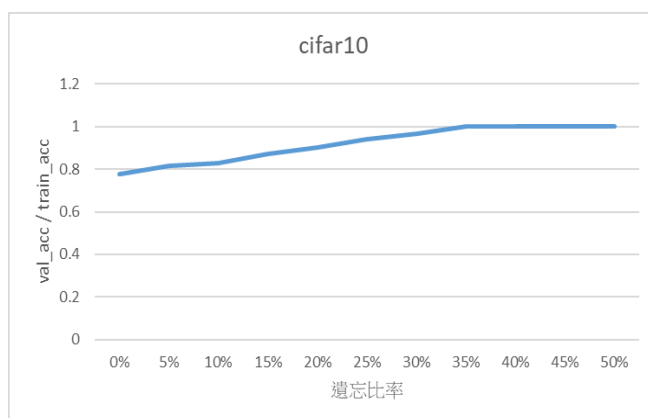
如下圖(圖十六製圖十八)所示，以遺忘比率當作橫軸，將驗證準確率除以訓練準確率之值當作縱軸來分析。可發現在Mnist，Fashion_Mnist，Cifar10資料庫中的實驗數據都能觀察到這現象。



圖十六



圖十七



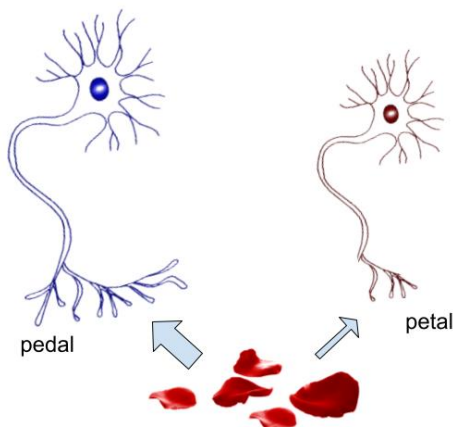
圖十八

陸、討論

一、對於dropout能提升準確率之解釋

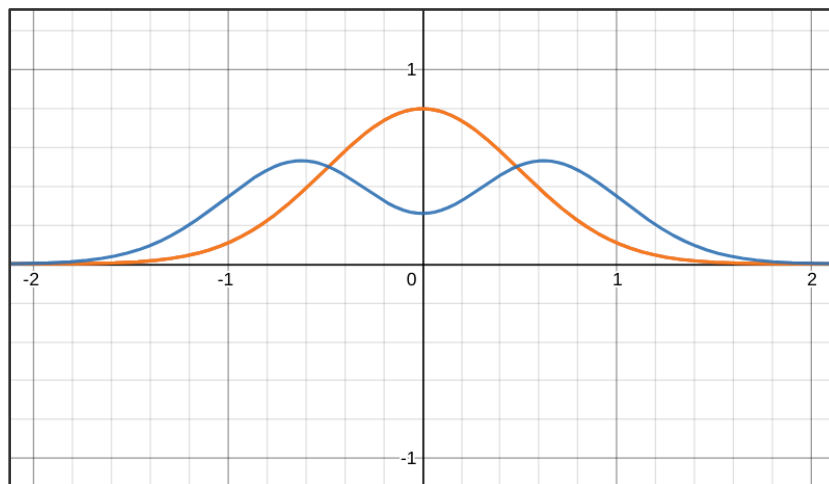
(一)在人類腦神經科學中，存在長期抑制作用及干擾理論。長期抑制作用，也就是突觸受到抑制的現象，源自於突觸受到強烈或長期的刺激。對於AI來說，訓練所需之大量數據即為強烈的刺激，過多的特徵資料抑制了AI準確判斷之能力。因此若減少此刺激，也就是透過dropout減少數據中的特徵資料，便可減少對於AI判斷能力的抑制。而干擾理論其中一種作用簡單來說為後來學習的內容會影響先前學習的內容，因此我們認為dropout作用等同於減少在多次學習之後所記得的瑣碎內容之干擾了，而不會影響先前學習的內容，此作用表現於AI判斷的準確率提升之現象。由於AI為高度仿生之科技，因此我們認為dropout可提升準確率之現象，能以人類大腦的這兩種機制來解釋。

(二)以大腦建立連結的機制而言，當有新刺激產生時，大腦更傾向於連結已有較多神經分支的神經元。以下圖為例，pedal為先前學習之單詞，意為踏板。petal意為花瓣，為後來學習之單詞，所建立之連結強度較pedal弱。當此人看到花瓣時，由於pedal的連結較強，大腦更容易將之聯想為pedal，而非正確之單詞-petal。因此，若能減弱pedal之連結，大腦應較容易聯想至正確的單詞petal，而AI使用dropout之效果即類似於此，AI的判斷應該要傾向於找出通式，而非過度的連結到單一特徵上導致只有擁有類似特徵的事物都會連結到錯誤的判斷，故dropout能提升判斷之準確率。



圖十九

(三)若是由本神經網路之損失函數:交叉熵(Cross-Entropy)之觀點來看，觀察交叉熵的公式，可以發現這是一個雙變數函數，其一為 x ，其二為 i 。因為 $P(y_i)$ 代表的是真值的機率分布，因此不會改變。在了解到他是一個雙變數函數後，我們可以知道他依定會有一個相對極大值。換個角度看，訓練AI的目的就是為了要將損失函數降到最低，而這個相對極大值就是損失函數的最小值，接下來再進行訓練只會將精準度降低而已。而前面提到熵的概念為"不確定性的期望值"，當準確率越高，熵的值也會隨之降低，這也可以解釋為何找到擁有最小值的交叉熵時可以有準確率最高的模型出現。以高斯機率分佈為例，在高斯機率分佈中，即使即使兩組的機率分佈完全吻合，交叉熵的面積仍不為0。(如圖十九)



圖十九

到這邊，我們來統整一下關於熵以及交叉熵跟我們報告的關係

- 熵是熱力學的名詞，跟我們的科展沒有太大關聯性
- 資訊熵是一種資訊量的概念，交叉熵需要使用到資訊熵進行計算
- 資訊熵可視為不確定性的期望值，亦或者是判斷一事件所需決策樹深度的期望值
- 交叉熵利用資訊熵進行計算，為一種損失函數，計算交叉熵需要使用到積分技巧
- 找到交叉熵的極值，即可計算出遺忘多少可以有最佳預測結果

二、對於改善過擬合狀況之解釋

首先，我們要先了解為何訓練準確率和驗證準確率的趨勢在多次訓練後差距越來越大，在CNN的模型運作的過程中，其實就是透過卷積和池化層來不斷的從大量資訊中取得越來越多的特徵圖，並以此來做為判斷依據，而正是因為驗證資料集的資料和訓練集的資料不同，導致AI在訓練集資料中記住太多不必要的資訊而導致AI在看驗證資料集中沒看過的圖片時，準確率下降，而我們透過Dropout來遺忘掉部分資料中的特徵，減少後來學習的內容影響先前學習的內容的程度，讓AI能夠更有效率地從資料中找出通式，而不是過度的連結於其中幾個特徵，導致只要出現類似的特徵，就會導向到錯誤的答案，故Dropout能夠改善過擬合的情形。

三、比較人類與AI之記憶和遺忘機制之差異

- 人類之遺忘過程無法由人類自主控制，而AI之dropout過程則可由人類控制
- AI遺忘過程為"隨機"刪除數值，人類則是傾向於遺忘聯繫較少之事件
- AI進行dropout的好處在於能提升判斷的準確率，而人類適度遺忘的好處則在於能增加空間以記憶大腦認為更加重要的內容

柒、結論

一、總結

(一)人類和AI遺忘機制的關聯性

人類記憶的干擾機制以及長期增強作用可做為資訊量過多而影響準確率之解釋。兩者之遺忘機制皆對其有幫助，分別為增加大腦容量及增加準確率。

(二)遺忘對於學習有幫助

電腦最擅長的就是記憶，能夠記憶起比人類還要多得多的資訊，而經過我們上列的各項實驗結果可知，適度的遺忘確實能對AI學習帶來幫助，也許記得最多的事不一定是最強大的，透過適度的設置Dropout層來遺忘某些資訊，不但可以提升AI預測的準確率，還能大幅度的改善訓練過程中會產生的過擬合之狀況，我們希望能透過研究更多人類遺忘機制，並在AI中使用這些方式，也許就能讓AI在未來的發展中變得更加強大，更類似於人類。

二、未來展望

(一)提升準確率之好處與應用

經過我們上列的各項實驗結果可知，適度的遺忘確實能對AI學習帶來幫助在調整Dropout值後發現其能夠改善CNN的預測結果，也許在研究更多人類遺忘機制後，能夠將其運用在不同的深度學習模型像是運用在RNN中讓其增加對長期或短期的預測準確率，亦或是運用在GAN中提升AI生成圖片和辨識假圖片的效果。

(二)與人類相異之處及發展性

人類遺忘之內容與事件之相關性與重要性有關，而目前dropout的過程為隨機刪除數值，因此未來可連結資訊熵的領域，透過觀察資訊熵的大小進行研判，看看是否能夠訓練AI自行判斷各資料之重要性差異，並學會決定刪除哪些數值以最大程度的改善判斷結果，或是使訓練準確率和驗證準確率的趨勢差距縮到最小。

捌、參考資料及其他

一、中文部分

【實體書】

1. François Chollet (2019). 深度學習必讀-Keras大神帶你用Python實作。臺北市：旗標。
2. 涌井貞美(2021)。圖解AI與深度學習的運作機制。臺北市：臺灣東販。
3. 陳允傑(2019)。TensorFlow與Keras - Python深度學習應用實務。臺北市：旗標。

二、網路資源

(一)、中文部分

1. 謝秉男(民105年8月3日)。雜談突觸可塑性(上)。民111年2月21日，取自：
<https://highscope.ch.ntu.edu.tw/wordpress/?p=73324>
2. 謝秉男(民105年8月3日)。雜談突觸可塑性(下)。民111年2月21日，取自：
<https://highscope.ch.ntu.edu.tw/wordpress/?p=73325>
3. 陳燕(民96年)。神經元的突觸可塑性與學習和記憶。民111年2月21日。取自：
http://www.ibp.cas.cn/sourcedb_ibp_cas/tsg/thesis/202105/P020210615593488125400.pdf
4. 許恆修(民108年1月14日)。何謂Cross-Entropy(交叉熵)。民111年2月25日。取自：
<https://r23456999.medium.com/%E4%BD%95%E8%AC%82-cross-entropy-%E4%BA%A4%E5%8F%89%E7%86%B5-b6d4cef9189d>

(二)、英文部分

1. Werner Stangl(2022). Was geschieht im Gehirn, wenn wir vergessen?
Retrieved February 21, 2022, from <https://arbeitsblaetter.stangl-taller.at/GEDAECHTNIS/Vergessen.shtml>
2. Werner Stangl(2022). Das Vergessen - Einige Forschungsergebnisse zum Erinnern und zum "False-Memory-Syndrome"
Retrieved February 21, 2022, from
<https://arbeitsblaetter.stangl-taller.at/GEDAECHTNIS/VergessenForschung.sht>
3. University of Bristol. AMPA Receptor
Retrieved February 21, 2022, from
<http://www.bristol.ac.uk/synaptic/receptors/ampar/>
4. Long-term depression.
Retrieved February 21, 2022, from
https://en.wikipedia.org/wiki/Long-term_depression#Characterisation