1. 各位口試委員、老師早安。我是吳承翰，今天要跟大家報告我的論文”以DAE向量雜訊移除為基礎之新進使用者冷啟動推薦”，這篇論文主要是想運用深度學習模型DAE，試著解決推薦系統領域中常見的冷啟動問題。

我會在之後的內容逐一地介紹什麼是推薦系統中的冷啟動問題以及DAE模型是什麼。在這之前，首先我們必須要了解推薦系統存在的重要性。

2. 想像一個情境，當我們在周五晚上忙完了一整周的事後，想要好好的窩在家裡的電腦前享受Netflix上的好片時，是否曾經遇過不知道該看什麼片的窘境。

這種情況凸顯了推薦系統存在的必要性，對於使用者來說如果沒有推薦系統或推薦系統的推薦效果不好時，很有可能使用者就會放棄使用該平台。

3. 換一個角度來看，對於平台方來說，好的推薦系統可以透過降低用戶的流失，帶來以下兩點的好處:

1. 可以增進現存用戶的終生價值

2. 可以降低尋找新客戶的成本 (廣告費)

在”The Netflix Recommender System: Algorithms, Business Value, and Innovation”這篇論文中就有提到光是透過改善推薦系統得到以上兩點的好處，每年可以幫助Netflix節省一百萬美元的花費。

5. 我們的研究主題是推薦系統中的冷啟動問題，會先介紹什麼是冷啟動問題，以及冷啟動問題的難處。接著會介紹我們該使用什麼深度學習模型以及，如何應用深度學習的模型來解決冷啟動問題

6. 在一般非冷啟動的情境下，我們可以利用用戶彼此之間的共同評分來推薦使用者可能會喜歡的商品。常見的例子有:Amazon網路書店中，購買了此商品的用戶也購買了哪些商品的推薦。

而在冷啟動的情境下，由於使用者大多剛進入平台沒多久，因此與其他使用者在商品的評分上並沒有重疊，而難以推薦商品給使用者

7. 介紹完冷啟動問題後，接著就是為什麼我們想使用深度學習來解決推薦系統冷啟動問題，我們個別從業界與學術界找到一些例子。

既然在業界與學界中有愈來愈多的人投入相關的研究，就能得知以深度學習為基礎的推薦顯然是一個無法避免的趨勢。

那麼問題又來了，千百種深度學習模型中，我們又該使用哪一種模型呢?

經過各方的survey後，我們選用了Denoising Autoencoder，原因在於Denoising Autoencoder有兩個很棒的特性:

8. 但是直觀上來說，推薦系統的冷啟動問題跟Noise好像沒什麼關係，那為什麼我們會想使用DAE模型呢?

我們可以將一個cold start user視為rich user不斷受到noise覆蓋掉評分的產物，如此一來就可以使用DAE來將cold start user還原回原本的rich user。

13. 在DropoutNet這篇論文中，提出了一個想法來解決cold start問題。當某個使用者或某個商品處於cold start狀態時，由於我們只有少數關於使用者或商品的評分資訊，因此我們會將對應評分的輸入設定為0，

以利模型用其他完整的content information去學習如何估計使用者對於商品的評分。而這種方法被該論文的作者稱為Dropout。

14. 這篇論文則提出representative-based matrix factorization這個方法來找出具代表性的商品。

舉個實例來說:~

\* 將user用k個representative items來表示，並且將items也用k個representative items來表示

15. 而這篇論文主要是想用來改進前一篇論文RBMF中，所有使用者共用一組具代表性商品這樣的缺點，而使用類似tree-based的方法將所有使用者分成不同組，分完組後再去做矩陣的拆解。

16. 在論文方法中，首先我們會詳細介紹DAE模型，接著介紹我們提出的返老還童方法，最後則是我們的方法如何進行訓練與推薦。

17. 在介紹Denoising Autoencoder之前，我們要先了解Autoencoder是什麼

20. 因此我們提出了一個適用於推薦系統而且可解釋性高的雜訊產生方法，稱其為”使用者返老還童”。

使用者返老還童方法主要分成兩個部分，其中選擇具代表性商品又分成好幾個子流程~

24. 將使用者分群後，接著我們會使用RBMF為每一組的使用者找出具代表性的商品