

姓名：周嘉禾

學號：D1166506

Netflix 或 YouTube 會需要進行影片推薦

請針對此大數據推薦系統考慮要如何進行

## 蒐集資料

用戶端互動資料：影片點擊、用戶評分

註冊資料：年齡、性別、所得、渠道、註冊時間、手機機型

影片資料：名稱、類型、創作者、上傳時間、描述、評論

外部資料：使用者的不同類型喜好

## 儲存、流動、轉換格式

影片格式(MP4、AVI、MKV、MOV、FLV)，並根據畫質、效率等的需求選擇最適合的格式，像是為了快速播放影片而選擇 FLV、為了高畫質而選擇 MOV 或 AVI 以及為了相容性而選擇 MP4 等。

## 大數據相關系統

關聯式資料庫、雲端儲存、伺服器、超級計算機

## 非大數據相關系統

用戶介面、播放器、支付系統、認證系統、內容管理系統

## 分析方式

推薦演算法主要有以下幾步：

- 召回：當用戶以及內容量比較大的時候，往往先通過召回策略，將百萬量級的內容先縮小到百萬級。
- 過濾：對於內容不可重複消費的領域，例如實時性比較強的新聞等，在使用者已經曝光和點選後不會再推送到使用者面前。
- 精排：對於召回並過濾後的內容進行排序，將百萬級的內容並按照順序推送。
- 混排：為避免內容越推越窄，將精排後的推薦結果進行一定修改，例如控制某一型別的頻次。
- 強規則：根據業務規則進行修改，例如在活動時將某些文章置頂。

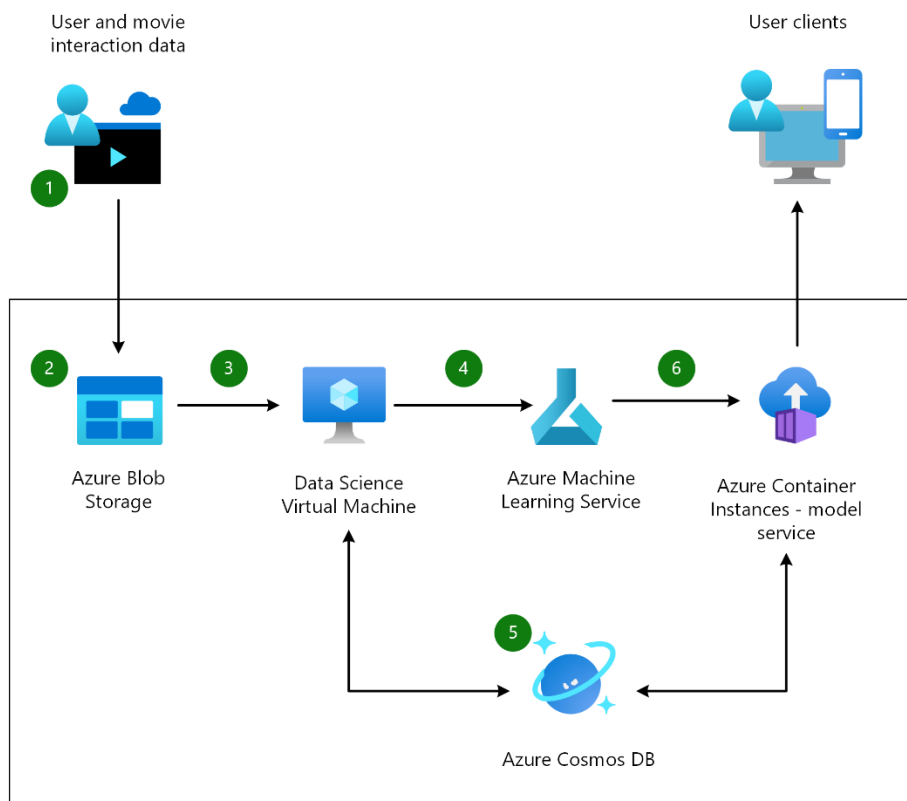
並透過機器學習、演算法、標籤製作等方式計算推薦的影片，如：

- 個人化影片評比 ( Personalized Video Ranker, PVR )  
這個演算法顧名思義是以巨量數據分析後的個人觀影喜好為依據，計算出使用者喜歡哪種類型的影片，以及這類型影片的出現次序
- 最佳推薦評比 ( Top-N Video Ranker )  
這個評比演算法著重計算影片庫中所有的影片，並且透過使用者的喜好數據，找出使用者最喜歡的哪幾部片，而不像個人影片評比演算法一樣，會先計算出使用者喜歡的影片類型，再從中排出次序。
- 最新趨勢評比 ( Trending Now )  
掌握使用者可能會想要看哪些影片，這種演算法主要的依據像是每年都會有的節慶，例如情人節，或者現在剛好發生了某些災禍，現在是鎂光燈焦點的特定事件等，再將這些資料與使用者的喜好數據結合，就能夠推薦使用者目前的趨勢影片。
- 繼續收看評比 ( Continue Watching )  
這個演算法是特別針對已經看過的影片，在已經看過但卻沒有看完的影片中，嘗試找出哪些使用者還會想要繼續欣賞的影片。在這個演算法中，有一些使用者的數據會被分析，包括這部影片的觀看時間，在哪個時間點離開觀賞（影片中間、一開始或結尾前），中斷觀賞後是不是還看了其它影片，以及用哪些裝置觀賞。
- 影片相似度評比 ( Video-Video Similarity )  
這是針對使用者曾經看過的影片中，依照每一部影片都挑選出相似的影片來推薦，在評比影片的相似程度時，演算法完全不會加入任何使用者喜好，完全是影片與影片之間來做評比。但評比完後，要將哪幾列使用者曾經看過影片的相似影片，推薦給使用者，就會加上使用者喜好。

- 頁面產生評比 ( Page Generation: Row Selection and Ranking )

這個演算法是用來決定，到底有哪些影片列可以出現在使用者的首頁上。在透過上述提及的演算法得出，究竟有哪些候選影片列使用者可能會想觀賞後，Netflix 會再透過「頁面產生評比」，排出一個與使用者相關且兼具多樣性的影片推薦首頁。

## 系統串接方式



圖片來源：Microsoft Azure

## 參考資料

選擇適當的影片格式

<https://www.adobe.com/tw/creativecloud/video/discover/best-video-format.html>

大數據轉化為多種演算法 成就 Netflix 精準影片推薦系統

<https://www.cool3c.com/article/131945>

推薦系統架構與演算法流程詳解

<https://www.gushiciku.cn/pl/g4JM/zh-tw>

使用機器學習建立電影建議系統

<https://learn.microsoft.com/zh-tw/azure/architecture/example-scenario/ai/movie-recommendations-with-machine-learning>