

# 資料探勘HW1

AI三 B1228005 胡樂麒

(a) 對 confidence 分數前十名的的每一個推薦規則，計算並印出出 confidence 與 support, 並加上六種 correlation 的分數的計算, lift and  $\chi^2$ , all confidence, max confidence, Kulczynski, and cosine.

== 前 10 名推薦規則的各種指標分析 ==

Rule	Conf	Supp	Lift	Chi <sup>2</sup>	AllConf	MaxConf	Kulcz	Cosine
If [Silence of the Lambs, The (1991), Return of the Jedi (1983)] → [Star Wars (1977)]   1.000   0.171   1.990   38.23   0.340   1.000   0.670   0.583								
If [Empire Strikes Back, The (1980), Fugitive, The (1993)] → [Raiders of the Lost Ark (1981)]   1.000   0.171   2.689   66.06   0.459   1.000   0.730   0.678								
If [Contact (1997), Empire Strikes Back, The (1980)] → [Raiders of the Lost Ark (1981)]   1.000   0.141   2.689   51.96   0.378   1.000   0.689   0.615								
If [Toy Story (1995), Return of the Jedi (1983), Twelve Monkeys (1995)] → [Star Wars (1977)]   1.000   0.186   1.990   21.07   0.210   1.000   0.605   0.458								
If [Toy Story (1995), Empire Strikes Back, The (1980), Twelve Monkeys (1995)] → [Raiders of the Lost Ark (1981)]   1.000   0.101   2.689   34.63   0.270   1.000   0.635   0.520								
If [Pulp Fiction (1994), Toy Story (1995), Star Wars (1977)] → [Raiders of the Lost Ark (1981)]   1.000   0.131   2.689   47.47   0.351   1.000   0.676   0.593								
If [Pulp Fiction (1994), Toy Story (1995), Return of the Jedi (1983)] → [Star Wars (1977)]   1.000   0.090   1.990   17.47   0.180   1.000   0.590   0.424								
If [Toy Story (1995), Silence of the Lambs, The (1991), Return of the Jedi (1983)] → [Star Wars (1977)]   1.000   0.121   1.990   24.80   0.240   1.000   0.620   0.490								
If [Toy Story (1995), Empire Strikes Back, The (1980), Return of the Jedi (1983)] → [Star Wars (1977)]   1.000   0.126   1.990   26.08   0.250   1.000   0.625   0.500								
If [Pulp Fiction (1994), Toy Story (1995), Shawshank Redemption, The (1994)] → [Silence of the Lambs, The (1991)]   1.000   0.101   2.843   37.88   0.286   1.000   0.643   0.535								

(b) test data 的 confidence 分數

Rule #1 Rule: If a person recommends Silence of the Lambs, The (1991), Return of the Jedi (1983) they will also recommend Star Wars (1977) - Train Confidence:1.000 - Test Confidence:0.936
Rule #2 Rule: If a person recommends Empire Strikes Back, The (1980), Fugitive, The (1993) they will also recommend Raiders of the Lost Ark (1981) - Train Confidence:1.000 - Test Confidence:0.876
Rule #3 Rule: If a person recommends Contact (1997), Empire Strikes Back, The (1980) they will also recommend Raiders of the Lost Ark (1981) - Train Confidence:1.000 - Test Confidence:0.841
Rule #4 Rule: If a person recommends Toy Story (1995), Return of the Jedi (1983), Twelve Monkeys (1995) they will also recommend Star Wars (1977) - Train Confidence:1.000 - Test Confidence:0.932
Rule #5 Rule: If a person recommends Toy Story (1995), Empire Strikes Back, The (1980), Twelve Monkeys (1995) they will also recommend Raiders of the Lost Ark (1981) - Train Confidence:1.000 - Test Confidence:0.903

Rule #6 Rule: If a person recommends Pulp Fiction (1994), Toy Story (1995), Star Wars (1977) they will also recommend Raiders of the Lost Ark (1981) - Train Confidence:1.000 - Test Confidence:0.816
Rule #7 Rule: If a person recommends Pulp Fiction (1994), Toy Story (1995), Return of the Jedi (1983) they will also recommend Star Wars (1977) - Train Confidence:1.000 - Test Confidence:0.970
Rule #8 Rule: If a person recommends Toy Story (1995), Silence of the Lambs, The (1991), Return of the Jedi (1983) they will also recommend Star Wars (1977) - Train Confidence:1.000 - Test Confidence:0.933
Rule #9 Rule: If a person recommends Toy Story (1995), Empire Strikes Back, The (1980), Return of the Jedi (1983) they will also recommend Star Wars (1977) - Train Confidence:1.000 - Test Confidence:0.971
Rule #10 Rule: If a person recommends Pulp Fiction (1994), Toy Story (1995), Shawshank Redemption, The (1994) they will also recommend Silence of the Lambs, The (1991) - Train Confidence:1.000 - Test Confidence:0.794

### (c) 結果觀察與現象分析

#### 1. min\_support

- a. 透過設定 `min_support`，篩選出在使用者之間具有一定受歡迎程度的電影，避免冷門或極端情況下的電影干擾關聯規則的準確性。

#### 2. Confidence

- a. 在訓練資料中，只要符合前提條件，用戶就「一定」會喜歡結論的電影  $\text{confidence} = 1$ ，顯示這些是非常強的推薦規則。
- b. 在測試資料中， $\text{confidence}$  也維持在  $0.79\sim0.97$  之間，表示這些規則對新用戶仍具有很高的預測能力。

#### 3. Lift

- a. 表示這些規則不只是因為結論電影本身受歡迎，而是真的因為「前提電影組合」提高了用戶喜歡結論電影的機率。
- b. 例如：「*Toy Story + Return of the Jedi + Twelve Monkeys → Star Wars*」的 *Lift* 為  $1.99$ ，顯示關聯性強。

#### 4. $\chi^2$

- a.  $\chi^2$  都很大，代表前提與結論間的關聯顯著高於隨機分佈，不是偶然出現。

#### 5. Cosine

- a. Cosine 值在 0.4 ~ 0.67 之間表示前提與結論在用戶群中的重疊度高，顯示這些電影往往吸引同一類型的觀眾。
- b. 多數規則都牽涉到同類型或系列電影（Star Wars、Toy Story、Pulp Fiction）。

#### 6. All confidence 與 Kulczynski

- a. All confidence 與 Kulczynski 都大約在 0.2~0.6 之間表示雖然 confidence 高，但規則只適用於「部分」用戶（不是所有用戶都會有前提行為）。
- b. Kulczynski 值穩定，代表規則正向與反向的強度接近。

#### 7. 規則內容多為經典熱門電影組合

- a. Star Wars、Return of the Jedi、Empire Strikes Back、Raiders of the Lost Ark、Toy Story、Pulp Fiction 這些都是知名大片，符合大部分人都喜歡的直覺。

## 總結：

從分析結果可以看到，Apriori 演算法能有效找出用戶偏好  
的「電影組合」。這些規則不僅在訓練集上 confidence 高達  
1.000，在測試集上也維持 0.79~0.97，顯示規則具有良好的泛化  
能力。而 Lift、 $\chi^2$ 、Cosine 等指標進一步證明了這些電影間的  
關聯性，多數規則都是由經典或熱門電影組成，與人們觀影偏好相  
符，說明了關聯分析在電影推薦上的實用性。