

# 基金推薦方法

---

陳一宏/徐祥恩

分析應用科/DBM

2017-6-30

# 推薦算法

- 用戶交易數據

- 協同過濾法 (Collaborative Filtering)

- 以用戶為基礎 (User based)

基於統計性質

- 以物品為基礎 (Item based)

- 隱語意模型

- 隱藏因子模型 (Latent Factor Model)

基於模型基礎

- 矩陣分解法 (Matrix Factorization)

- 以內容為基礎 (Content based)

- 物品特徵

- 用戶特徵

# 用戶交易數據 — 以用戶為基礎(協同過濾法)

User\Item	甲	乙	丙	丁	戊
A	1	1		1	
B	1		1		
C		1			1
D			1	1	1



用戶向量

$$Sim_{AB} = \frac{|N(A) \cap N(B)|}{\sqrt{|N(A) \cup N(B)|}} = \frac{1}{\sqrt{3}\sqrt{2}} = \frac{1}{\sqrt{6}}$$

$$Sim_{AD} = \frac{|N(A) \cap N(D)|}{\sqrt{|N(A) \cup N(D)|}} = \frac{1}{\sqrt{3}\sqrt{3}} = \frac{1}{3}$$

依用戶相似性推估  
用戶A選擇物品丙  
的興趣分數為**0.74**

$$\begin{aligned}
 P(A, \text{丙}) &= Sim_{AB} \times 1(B\text{買}) + Sim_{AD} \times 1(D\text{買}) \\
 &= +\frac{1}{\sqrt{6}} \times 1 + \frac{1}{3} \times 1 = \mathbf{0.74}
 \end{aligned}$$

# 用戶交易數據 — 以物品為基礎(協同過濾法)

User\Item	甲	乙	丙	丁	戊
A	1	1	1	1	
B	1	1	1	1	
C		1			1
D			1	1	1

物品向量

$$Sim_{甲丙} = \frac{|N(甲) \cap N(丙)|}{\sqrt{|N(甲) \cup N(丙)|}} = \frac{1}{\sqrt{2}\sqrt{2}} = \frac{1}{2}$$

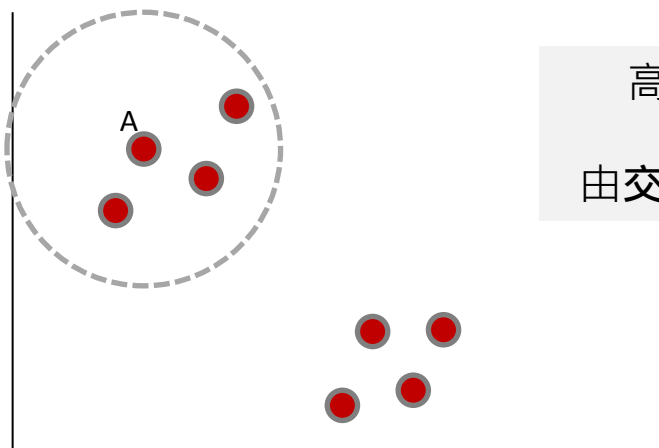
$$Sim_{乙丙} = \frac{|N(乙) \cap N(丙)|}{\sqrt{|N(乙) \cup N(丙)|}} = \frac{1}{\sqrt{3}\sqrt{2}} = \frac{1}{\sqrt{6}}$$

依物品相似性推估  
用戶A選擇物品c的  
興趣分數為1.72

$$\begin{aligned}
 P(A, 丙) &= Sim_{甲丙} \times 1 + Sim_{乙丙} \times 1 + Sim_{丙丁} \times 1 \\
 &= \frac{1}{2} \times 1 + \frac{1}{\sqrt{6}} \times 1 + \frac{2}{\sqrt{6}} \times 1 = 1.72
 \end{aligned}$$

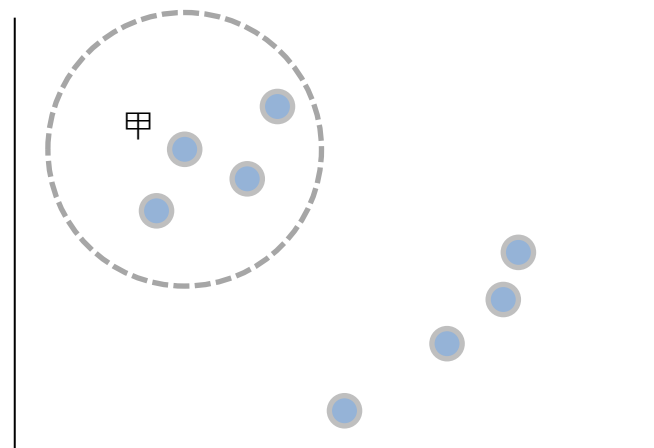
# 用戶交易數據 — 協同過濾法示意

用戶分布概況



高維空間 —  
由交易資料構成

物品分布概況

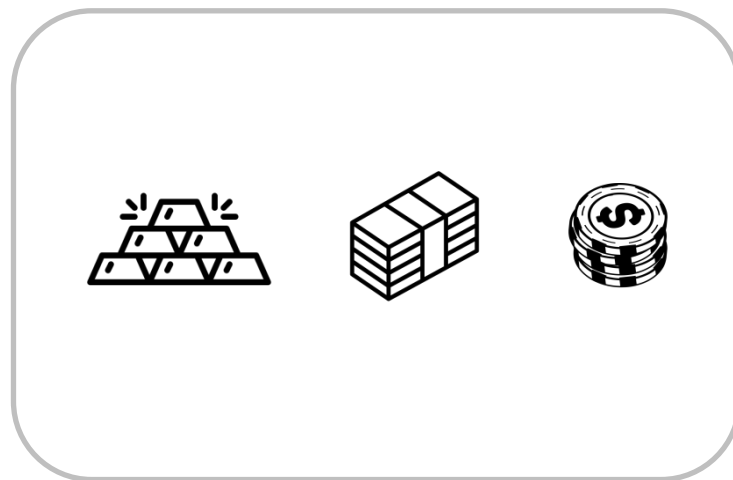
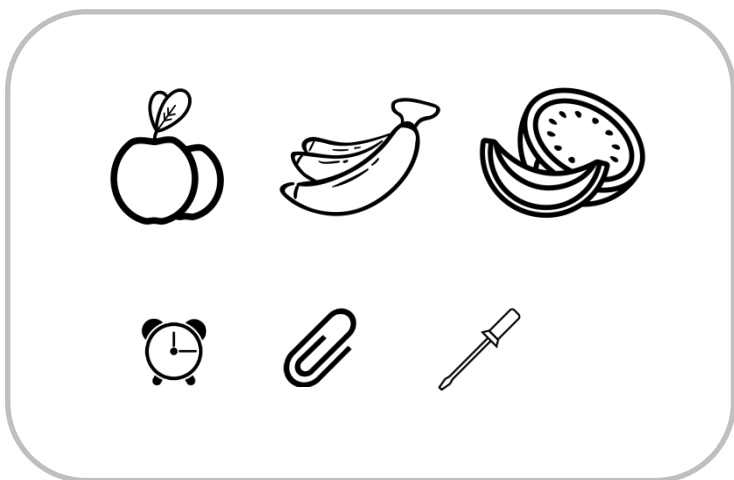


針對**用戶A**來說，選擇最近鄰的N個**相似用戶**，發掘相似用戶購買的**物品(基金)**

針對**用戶A**有買的**物品甲**來說，選擇最近鄰的N個**相似物品**，發掘相似**物品(基金)**

# 以內容為基礎

1. 參考用戶購買過的物品，推薦最**相關(相似)**的物品
2. 參考用戶的屬性，推薦最相似用戶購買的物品



相關程度由，系統對物品/用戶的背景知識定義

# 評估方法

訓練集

測試集

U-I matrix	甲	乙	丙	丁
A	1	1		
B	1		1	
C		1		1
D	1		1	1
E			1	1

U-I matrix	甲	乙	丙	丁
F	1		<del>1</del>	
G		1		<del>1</del>

將資料分為訓練/測試集合，透過訓練集產生相似關係，在測試集中，透過遮蔽部分答案驗證準確度/召回度。

# 本行基金交易資料

---

1. 協同過濾法
2. 物品混合協同過濾法
3. 特徵抽取協同過濾法



# 資料涵蓋

---

- 時間區間：2015.01-2017.05.25
- 考慮申購基金檔數  $\geq 2$
- 用戶數計：26,621
- 基金數計：2,177

# 1. 協同過濾法

透過交易資料評估用戶最可能購買...

交易紀錄

UI-Matrix	甲	乙	丙	丁	戊
A	1	1		1	
B	1		1		
C		1			1
D			1	1	1

用戶向量 (blue arrow pointing right)

物品向量 (red arrow pointing down)

基於物品相似度- IBCF

物品相似度	甲	乙	丙	丁	戊
甲	0	xx	xx	xx	xx
乙		0	xx	xx	xx
丙			0	xx	xx
丁				0	xx
戊					0

UI-Matrix	甲	乙	丙	丁	戊
A	1	1	0.9	1	0.7
B	1	1.3	1	0.4	0.5
C	1.4	1	1.3	0.2	1
D	1.1	1.5	1	1	1

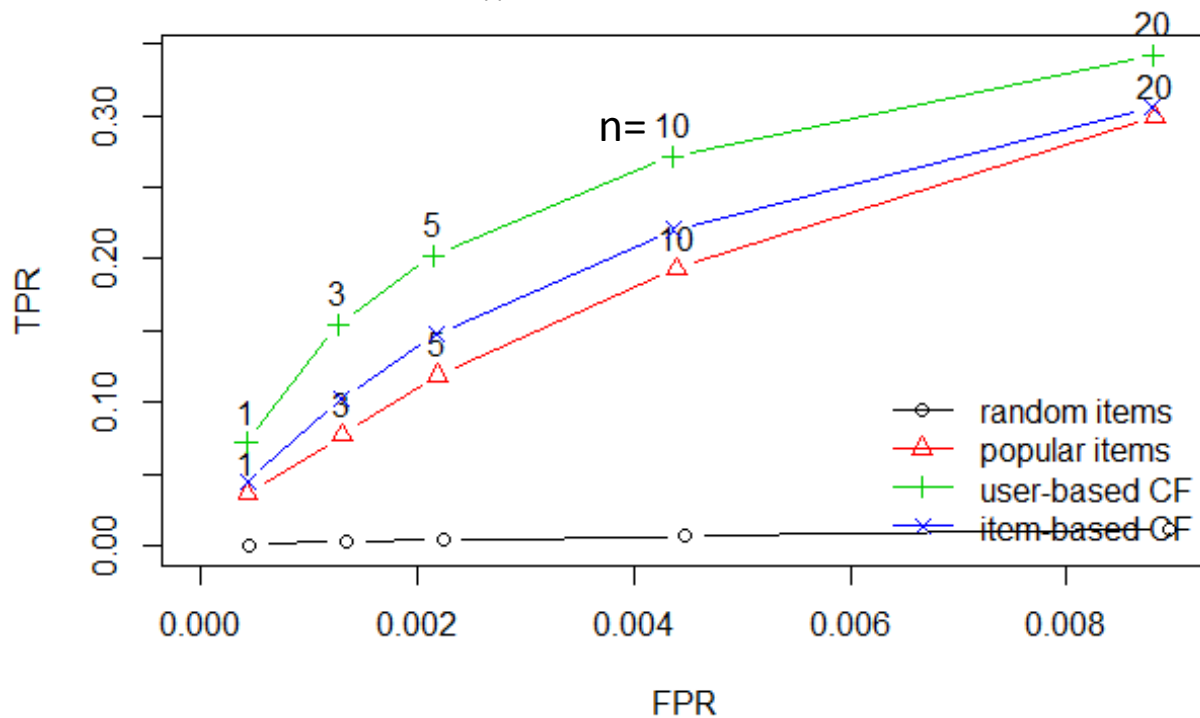
基於用戶相似度- UBCF

用戶相似度	A	B	C	D
A	0	xx	xx	xx
B		0	xx	xx
C			0	xx
D				0

UI-Matrix	甲	乙	丙	丁	戊
A	1	1	1.2	1	0.8
B	1	1.3	1	1.4	0.7
C	1.1	1	1.5	0.3	1
D	1.0	1.2	1	1	1

# 1.評估算法－協同過濾法

ROC－受試者工作曲線



- 推薦 1,3,5,10,20 個物品評估結果
- 根據ROC圖判斷，基於用戶相似度 (UBCF)算法，得到較佳結果

## 2.物品混合型協同過濾法 — 物品屬性

除交易資料以外，另考慮非交易型資料 — 基金屬性

分析面向	分類	欄位範例
基本資訊	境內外	國內外基金註記
	規模大小	基金目前規模區間
	成立時間	基金成立幾年
	基金公司	基金公司代碼
投資標的	投資型態別	股票型/債券型/平衡型/指數型
	商品風險屬性	RR1-RR5
操作績效	淨值	
	夏普指數	
	貝他指數	
	累積報酬率	一個月/三個月/半年/一年/三年 /五年累積報酬率
	熱門基金	熱門銷售註記

國內股票型基金

國外債券型基金

國外股票型基金

## 2.物品混合型協同過濾法 – 交易+非交易型資料

物品相似度 – 基於交易資料 (S1)

物品-物品 相似矩陣	甲	乙	丙	丁
甲	0	xx	xx	xx
乙		0	xx	xx
丙			0	xx
丁				0

物品相似度 – 基於非交易資料(S2)

物品-物品 相似矩陣	甲	乙	丙	丁
甲	0	xx	xx	xx
乙		0	xx	xx
丙			0	xx
丁				0

$$S = (1 - W) \times S_1 + W \times S_2$$

內容權重

混合物品相似度: 交易資料 + 物品非交易資料

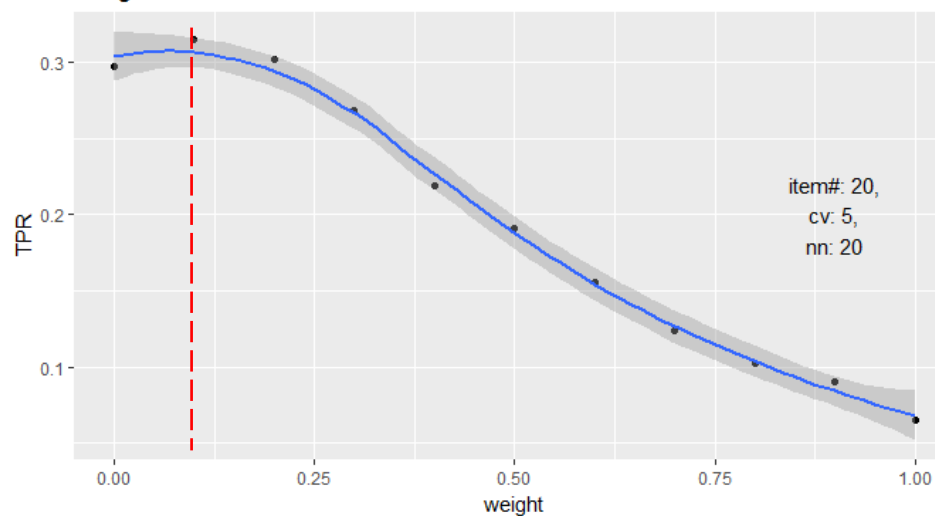
混合用戶相似度: 交易資料 + 用戶非交易資料

## 2.評估算法 – 物品混合型協同過濾法

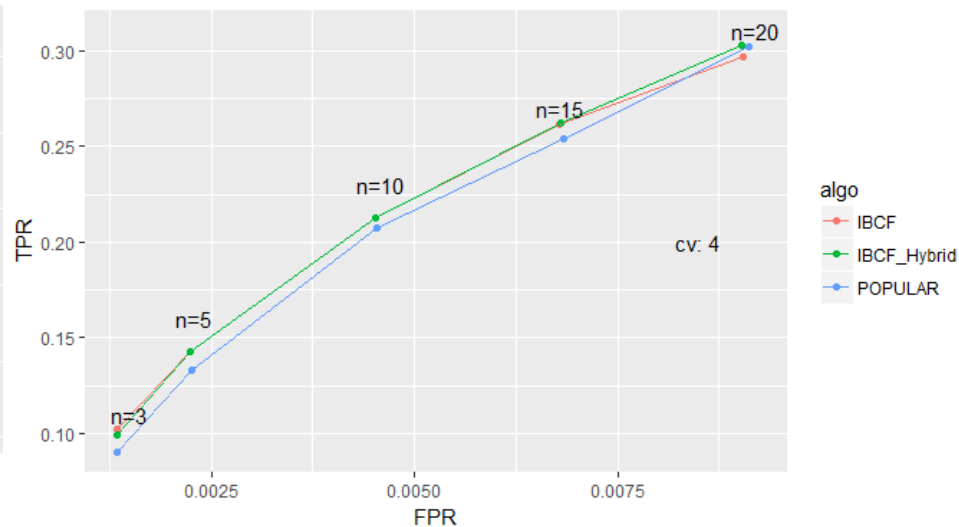
物品屬性影響結果

$$S = (1 - W) \times S_{trans} + W \times S_{Content}$$

Weight of item-content



ROC



加入少量物品內容元素提升“些許”的準確度

### 3.特徵抽取協同過濾法-原理1

國內股票型基金(ex: 甲)

U-I matrix	甲	乙	丙	丁
A	1	1		
B	1			
C		1		1
D	1		1	1
E			1	1

U-I matrix	甲	乙	丙	丁
A	1	1	0.9	0.8
B	1	0.7	0.6	0.5
C		1		1
D	1	0.3	1	1
E			1	1

國外債券型基金(ex: 乙)

U-I matrix	甲	乙	丙	丁
A	1	1		
B	1			
C		1		1
D	1		1	1
E			1	1

U-I matrix	甲	乙	丙	丁
A	1	1	0.8	0.6
B	1			
C	0.5	1	0.33	1
D	1		1	1
E			1	1

國外股票型基金(ex: 丙丁)

U-I matrix	甲	乙	丙	丁
A	1	1		
B	1			
C		1		1
D	1		1	1
E			1	1

U-I matrix	甲	乙	丙	丁
A	1	1		
B	1			
C	0.1	1	0.33	1
D	1	0.2	1	1
E	0.7	0.5	1	1

U-I matrix	甲	乙	丙	丁
A	1	1		
B	1			
C		1		1
D	1		1	1
E			1	1

### 3.特徵抽取協同過濾法-原理2

U-I matrix	甲	乙	丙	丁	Features Table	國內股票型	國外股票型	國外債券型	AUM 100萬內	AUM 100~300萬
A	1	1			A	1	0	1	1	0
B	1				B	1	0	0	0	1
C		1		1	C	0	1	1	0	1
D	1		1	1	D	1	1	0	1	0
E			1	1	E	0	1	0	1	0

對C評分－

1.基於國內股票型

× 0

2.基於國外債券型

U-I matrix	甲	乙	丙	丁
C	0.33	0	0	0

3.基於國外股票型

U-I matrix	甲	乙	丙	丁
C	0.33	0	0.667	0

4.基於AUM100~300萬

U-I matrix	甲	乙	丙	丁
C	0	0	0.1	0

甲(0.66)  
丙(0.76)

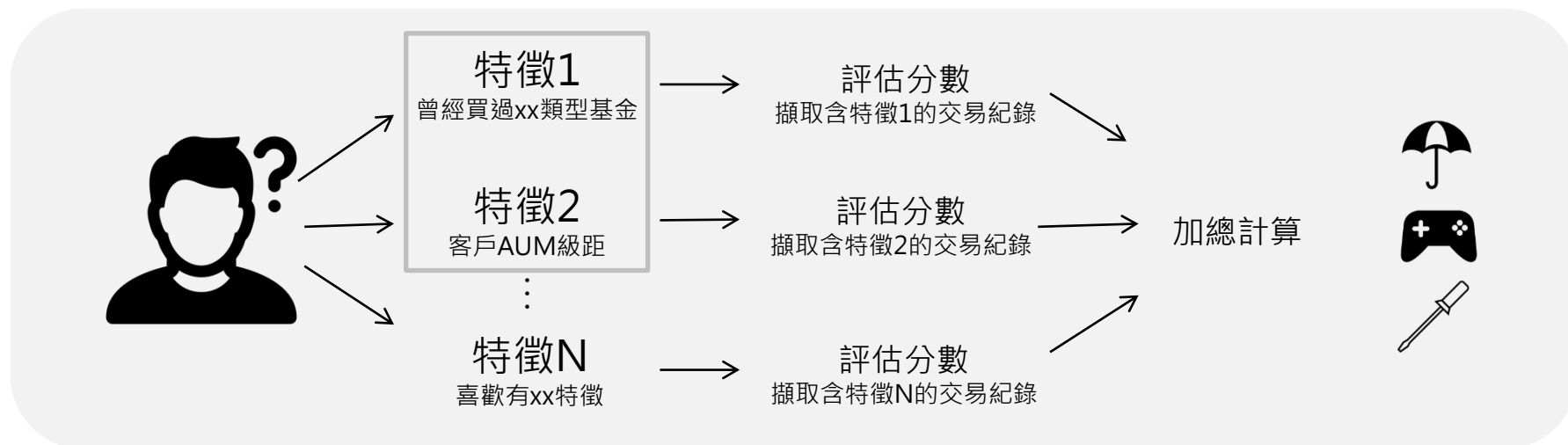


### 3.特徵抽取-選擇面向

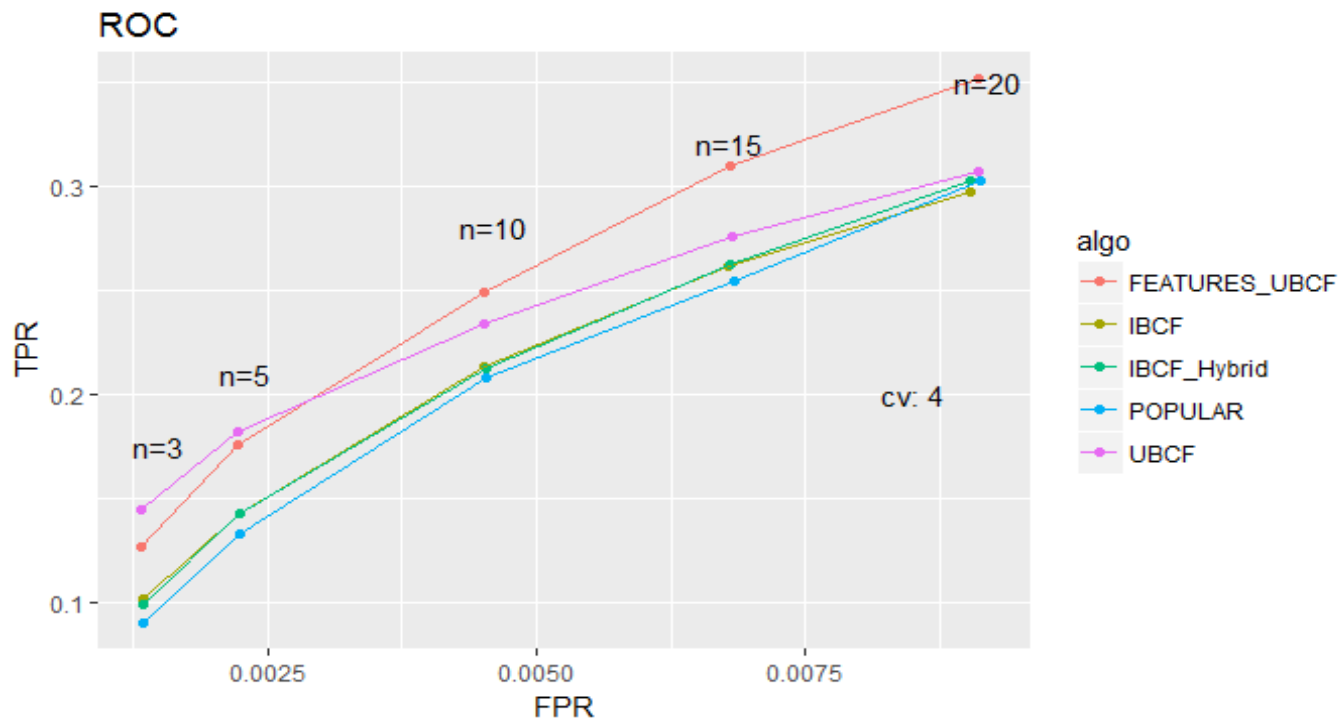
面向	類別	說明
基金特徵	國內股票型基金	基於分群結果
	國外股票型基金	
	國外債券型基金	
用戶特徵	AUM級距	0
		0-100萬
		100-300萬
		300萬以上

Features Table	國內股票型	國外股票型	國外債券型	AUM=0	AUM<100 萬	100~300萬	AUM>300萬
A123456789	1	0	1	0	1	0	0
A567890123	1	0	0	0	0	1	0
B134567891	0	1	1	0	0	0	1
B134889123	1	1	0	0	0	1	0
A111023456	0	1	0	0	1	0	0

### 3.特徵抽取-示意



### 3.評估算法 — 特徵抽取協同過濾法



加入特徵抽取方法(AUM級距/基金種類)，有效提升推薦準確度

推薦20個基金，正確率(TPR)約35%

### 3.推薦清單- 特徵抽取協同過濾法

Id	已知購買	國內股票	國外債券	國外股票	AUM	預測	結果	正確/錯誤
47457400	63Z_第一金全球多重資產入息平衡基金A(累積)	1	1	1	300萬以上	J99_摩根多重收益基金-JPM(美元對沖)A股(每月派息)	J99_摩根多重收益基金-JPM(美元對沖)A股(每月派息)	O
	GK0_復華全球債券組合基金					T34_安聯收益成長基金-AM穩定月收類股(美元)		
	H17_貝萊德中國基金 A2 美元					J84_摩根俄羅斯基金-JPM(美元)A股(分派)		
	GK0_復華全球債券組合基金					GJ8_第一金中國世紀基金		
	JOA_摩根多重收益基金-JPM					L91_聯博全球高收益債券基金AA(穩定月配)級別美元)		

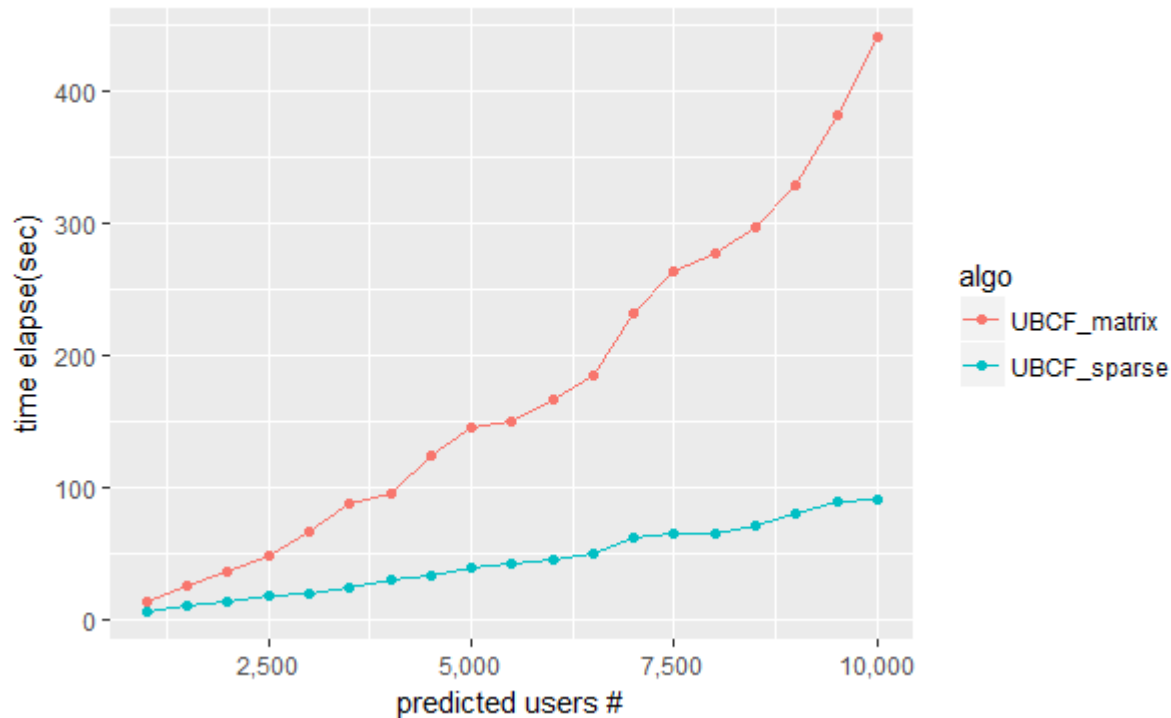
Id	已知購買	國內股票	國外債券	國外股票	AUM	預測	結果	正確/錯誤
195660120	297_瀚亞亞太豐收平衡基金(不配息)-人民幣	1			100~300萬	T34_安聯收益成長基金-AM穩定月收類股(美元)	312_富邦中國新平衡入息基金(人民幣累積)	X
	327_永豐人民幣貨幣市場基金(人民幣)	1				C20_霸菱東歐基金 (美元)		
						MW6_貝萊德拉丁美洲基金 (美元)		
						Y15_富蘭克林坦伯頓全球拉丁美洲基金(年配權)		
						317_聯博全球高收益債券基金TA月配(人民幣)		

### 3.推薦次數 — 特徵抽取協同過濾法

推薦次數 排名	基金名稱	熱銷排名	
1	T34_安聯收益成長-AM穩定月收類股(美元)	1	
2	J99_摩根多重收益-JPM(美元對沖)A股(每月派息)	5	
3	UN8_永豐滬深300紅利指數	2	
4	63Z_貝萊德中國 A 2 ( 美元 )	3	
5	L91_聯博全球高收益債券AA(穩定月配)級別美元	6	
6	V01_瑞銀 ( 盧森堡 ) 生化股票(美元)	4	
7	AE7_柏瑞印度股票基金A	10	
8	MU7_貝萊德世界礦業 ( 美元 )	8	
9	GJ8_第一金中國世紀	12	國內股票型基金
10	78F_施羅德環球股息AX配息(美元)	17	國外股票型基金
			國外債券型基金

# 挑戰 – 效能提升

Time Elapse for different algos



相似度計算耗用  $O(N^2)$

問題

- ① 記憶體耗用
- ② 時間消耗

隨用戶增加倍增

改用

- ① 稀疏矩陣  $O(N \log N)$
- ② 向量化計算

結果

26,621 用戶

套件 – 2,708 sec  
修改 – 348 sec

# 結論

---

- 新客戶(New-New)
  - 問卷 → 基金套餐
- 未曾買過(或僅有一檔)基金客戶
  - 客戶相似度分析 → 個人化推薦
- 已經購買兩檔以上用戶
  - 特徵抽取協同過濾 → 個人化推薦

# To Do

---

- 新客戶/僅購買1檔基金
- AB test 方法

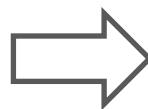


# 附件

---

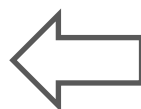
# 資料涵蓋

	申購
時間區間	2015.01-2017.05.25
用戶數	45,350
憑證數	207,408
涵蓋基金數	2,334



MMA架上涵蓋有 2,235檔基金  
用戶數 45,288 人

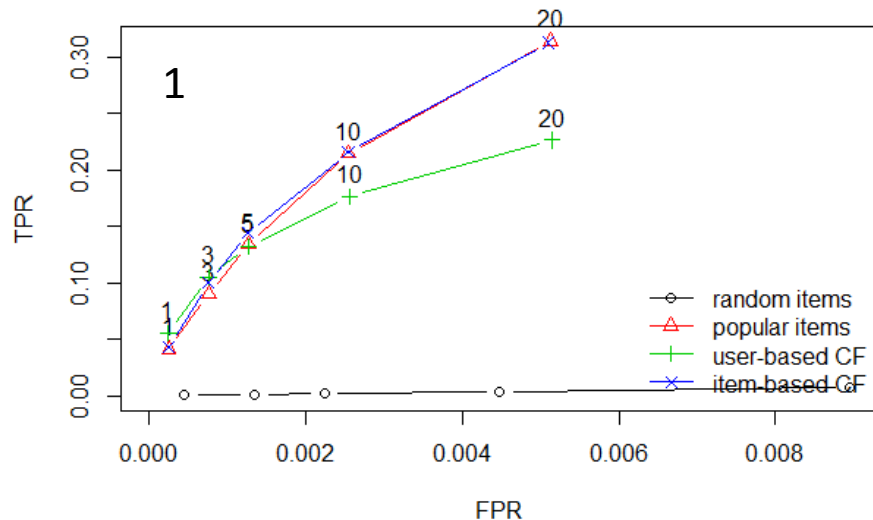
僅考慮申購基金數  $\geq 2$  ,  
用戶數計：26,621 (扣除1關戶)  
基金數計：2,177



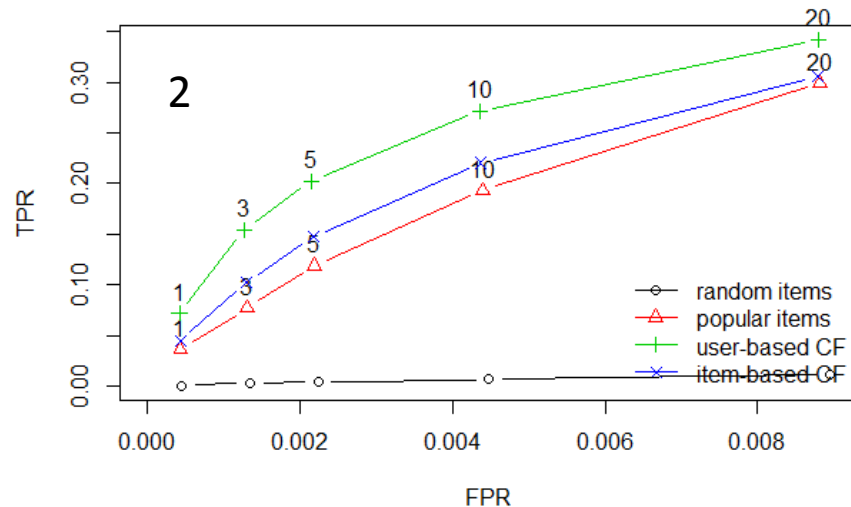
申購基金數	用戶數
1	18,290
2	9,243
$\geq 3$	17,399

# 模型評估 – ROC

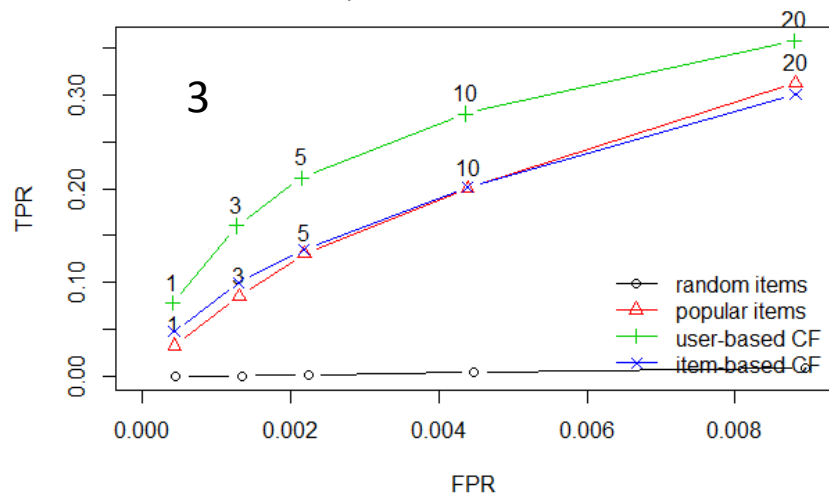
至少申購一檔基金



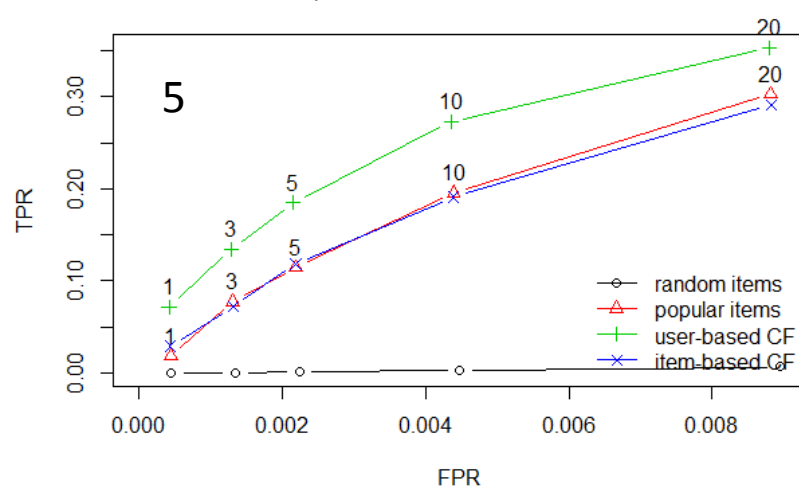
至少申購兩檔基金



至少申購三檔基金



至少申購五檔基金



# 基金分群-1

基金分類	國內外基金註記	基金目前規模區間	區域別	商品投資屬性	淨值
Length:546 cluster:1 國內股票型	國內0 ---- 542	d.10~100億元台幣 :208	TW :452	RR1 :49	Min. : 0.31
	國外1 ----- 4	b.1~5億元台幣 :143	CN :67	RR2 :11	1st Qu.: 10.11
		c.5~10億元台幣 :111	IN : 7	RR3 :31	Median : 12.44
		a.<1億元台幣 :58	BR : 3	RR4 :279	Mean :16.66
		e.100~500億元台幣:23	JP : 3	RR5 :175	3rd Qu.: 18.17
		f.500~5000億元台幣: 3	LU : 3	NA's: 1	Max. :176.96
		(Other) : 0	(Other): 11	NA	NA
基金代碼	國內外基金註記	基金目前規模區間	區域別	商品投資屬性	淨值
Length:1031 cluster:2 國內外全球債券(高收益)/平衡型	國內0 --- 190	h.1~5兆元台幣 :286	LU :594	RR1 :26	Min. : 0.32
	國外1 --- 841	i.>10兆元台幣 :225	TW :151	RR2 :233	1st Qu.: 10.12
		i.5~10兆元台幣 :131	US :133	RR3 :742	Median : 13.79
		f.500~5000億元台幣:106	IE :94	RR4 :24	Mean :106.35
		g.5000億~1兆元台幣: 75	CN :28	RR5 : 4	3rd Qu.: 39.11
		(Other) :200	GB :12	NA's: 2	Max. :4956.25
		NA's : 8	(Other): 19	NA	NA's :4
基金代碼	國內外基金註記	基金目前規模區間	區域別	商品投資屬性	淨值
Length:1200 cluster:3 國外股票型	國內0 ---- 0	h.1~5兆元台幣 :485	LU :658	RR1 : 0	Min. : 0.38
	國外1 ---- 1195	f.500~5000億元台幣:243	US :145	RR2 : 0	1st Qu.: 14.39
		g.5000億~1兆元台幣:232	IE :89	RR3 :58	Median : 26.91
		i.5~10兆元台幣 :119	JP :87	RR4 :732	Mean : 325.15
		i.>10兆元台幣 :67	CN :41	RR5 :399	3rd Qu.: 82.79
		(Other) :49	(Other):175	NA's: 11	Max. :59618.00
		NA's : 5	NA's : 5	NA	NA's :1

# 基金分群-2

基金分類	淨值	Sharpe	Beta	六個月累積報酬率(%)	五年累積報酬率(%)	熱賣基金註記	投資型態別
Length:546 cluster:1 國內股票型	Min. : 0.31 1st Qu.: 10.11 <b>Median : 12.44</b> Mean : 16.66 3rd Qu.: 18.17 Max. : 176.96 NA	Min. : -47.5600 1st Qu.: 0.1000 <b>Median : 0.3000</b> Mean : -0.5903 3rd Qu.: 0.4500 Max. : 4.3300 NA's : 24	Min. : -0.1800 1st Qu.: 0.3600 <b>Median : 0.7100</b> Mean : 0.6754 3rd Qu.: 0.9400 Max. : 1.7700 NA's : 24	Min. : -13.590 1st Qu.: 2.007 <b>Median : 6.220</b> Mean : 6.507 3rd Qu.: 10.330 Max. : 27.090 NA's : 18	Min. : -32.86 1st Qu.: 16.93 <b>Median : 39.31</b> Mean : 42.81 3rd Qu.: 61.08 Max. : 142.69 NA's : 166	否0: 544 是1: 2	a.股票型:325 d.平衡型: 65 b.債券型: 52 c.貨幣型: 49 e.組合型: 37 g.指數型: 12 (Other): 6
基金代碼	淨值	Sharpe	Beta	六個月累積報酬率(%)	五年累積報酬率(%)	熱賣基金註記	投資型態別
Length:1031 cluster:2 國內外全球債券(高收益)/平衡型	Min. : 0.32 1st Qu.: 10.12 <b>Median : 13.79</b> Mean : 106.35 3rd Qu.: 39.11 Max. : 4956.25 NA's : 4	Min. : -12.6200 1st Qu.: 0.1600 <b>Median : 0.3500</b> Mean : 0.3312 3rd Qu.: 0.5700 Max. : 5.9600 NA's : 22	Min. : -2.5700 1st Qu.: 0.5400 <b>Median : 0.7600</b> Mean : 0.7271 3rd Qu.: 0.9400 Max. : 3.7000 NA's : 22	Min. : -6.630 1st Qu.: 2.047 <b>Median : 5.080</b> Mean : 4.929 3rd Qu.: 7.190 Max. : 16.130 NA's : 19	Min. : -48.34 1st Qu.: 7.15 <b>Median : 20.78</b> Mean : 22.31 3rd Qu.: 32.63 Max. : 142.33 NA's : 387	否0: 1024 是1: 7	b.債券型:800 d.平衡型:109 a.股票型: 59 e.組合型: 33 c.貨幣型: 26 h.其他型: 3 (Other): 1
基金代碼	淨值	Sharpe	Beta	六個月累積報酬率(%)	五年累積報酬率(%)	熱賣基金註記	投資型態別
Length:1200 cluster:3 國外股票型	Min. : 0.38 1st Qu.: 14.39 <b>Median : 26.91</b> Mean : 325.15 3rd Qu.: 82.79 Max. : 59618.00 NA's : 1	Min. : -0.130 1st Qu.: 0.320 <b>Median : 0.440</b> Mean : 0.439 3rd Qu.: 0.550 Max. : 2.230 NA's : 18	Min. : -0.0800 1st Qu.: 0.6600 <b>Median : 0.8600</b> Mean : 0.8269 3rd Qu.: 1.0100 Max. : 1.7000 NA's : 18	Min. : -12.99 1st Qu.: 8.95 <b>Median : 13.21</b> Mean : 12.87 3rd Qu.: 17.10 Max. : 32.04 NA's : 7	Min. : -43.85 1st Qu.: 36.70 <b>Median : 63.02</b> Mean : 60.99 3rd Qu.: 85.31 Max. : 249.75 NA's : 244	否0: 1190 是1: 10	a.股票型:1168 d.平衡型: 26 ND : 5 c.貨幣型: 1 b.債券型: 0 e.組合型: 0 (Other): 0