Big Data Analysis x 91 APP Final Project —

會員行為模型、分群與商品推薦

第7組

資管三 杜沛慈 | 經濟三 胡南圳 | 圖資四 凌麗

圖資四 林奕萱 | 會研所 陳詩婷 | 會研所 江泓葦

說明影片: https://youtu.be/zfDZwNq-mYM

目錄

- 01 動機、問題與解方
- 02 資料處理
- 03 模型分析
- 04 總結

01

- 01 動機、問題與解方
- 02 資料處理
- 03 模型分析
- 04 總結



動機與商業問題

分析客戶資訊,了解客戶行為與商品購買之關聯,發掘目標客群將行銷資源花在刀口上,建立精準行銷

實作方式

會員行為模型

透過行為資料加權 建立模型,預測目標會 員是否會購買特定商品



會員分群

進行會員分群, 搜尋與目標會員有 類似偏好的的會員



商品推薦

將目標會員可能會購買 的商品推薦給其他在 同一群集的會員

02

- 01 動機、問題與解方
- 02 資料處理
- 03 模型分析
- 04 總結

建立商品標籤



使用CKIP對SalePageTitle進行斷詞

2. 篩選含有中文且出現次數 高於5次的token作為標籤

避免選到太冷門的token作為標籤, 選擇出現5次以上的token作為標籤

3. 手動篩選長度為一個字的標籤

一個字的token多不具意義,

需手動刪除無意義的標籤

4. 刪除長度為四個字以上的標籤

四個字的token大多可切分為二字或三字, 且已含在二字和三字的token中,故删去不用

建立17742商品 x 2054個詞的矩陣

SalePageId	SalePageTitle	token	モモ	不修邊	手機	光蝴蝶結	抗皺	結麻	•••	緊緻	縮小	化苞釦帶	細絨紅	訂製	電鍍	真牛	金屬釦	筆記本	人氣
u1n1twlQk7jvA==	123	['123']	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0
sHYTfVeuHuw==	Ann'S品牌止滑墊	["Ann'S", '品 牌', '止滑墊']	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0
ZQiTcisLdcJw==	Care bear海灘提袋	['Care', ' bear', '海灘', '提袋']	0	0	0	0	0	0		0	0	0	0	0	0	0	0	0	0





行為資料處理

1. 準備行為資料

僅保留瀏覽商品、加入購物車、結帳 3種行為,並取出所需欄位

MemberIDBehaviorSalePageIDwhuoifjbwViewProductwduvihjdkuaqnjviuhnjAddtoCartcjioduhrujiuwzjiogqjisCheckOuteoihbjquiyej

2.建立3種行為的商品標籤矩陣

將每一筆行為與其商品套入上述商品標籤 矩陣中, 並依照行為分割成三個表

View	SalePageID	紅色	簡約	高跟
會員 A	wduvihjdku	0	1	0
會員 A	cjioduhrujiu	1	0	1
會員A	eoihbjquiye	1	1	0

3.計算每位會員的標籤分數

將資料Group By會員ID, 並將商品標籤 次數相加。得到每位會員在三種行為中的 商品分數

View	紅色	簡約	高跟
會員 A	10	7	6

Add	紅色	簡約	高跟
會員 A	3	2	2

Check	紅色	簡約	高跟
會員 A	1	0	2



標籤分數加權

1. 計算全店行為比例, 定義分數權重

計算瀏覽商品、加入購物車、結帳 3種行為數量,轉成倒數比例

透過觀察全店行為比例,可以得知購物各階段的轉換機率, e.g. Add / View= 1 / 18.25, 代表顧客平均瀏覽 18.25個商品後會加入一個商品至購物車,由此作為行為權重

Behavior	行為數量	加權比例
ViewProduct	30293877	1
AddToCart	1660374	18.25
CheckOut	399861	75.66

2. 計算出加權後的標籤分數

將三種行為標籤分數乘上權重, 總和成每位會員在每個標籤的分數 後續分析將使用此加權過後的分數進行

Sum	紅色	簡約	高跟
會員 A	120	62	48
會員 B	38	280	75
會員 C	0	135	20

03

- 01 動機、問題與解方
- 02 資料處理
- 03 模型分析
- 04 總結



選擇目標會員

行為資料、購買資料 兩者筆數皆多者

目標會員相關商品

篩選出目標會員瀏覽、 加入購物車、結帳之商品

目標會員商品向量空間

從所有商品特徵詞向量空間中,取出目標會員相關商品部分

會員A

紅色高跟鞋 簡約高跟鞋

	紅色	簡約	高跟
紅色高跟鞋	1	0	1
簡約高跟鞋	0	1	1



目標會員行為向量空間

取出目標會員經行為加權(瀏覽、加入購物車、結帳)後的商品特徵詞向量空間

相關商品×行為加權

將目標會員的相關商品、行為加權商品特徵詞向量空間數值相乘,得出該會員經行為加權的相關商品

	紅色	簡約	高跟
會員 A	3	2	7

會員A	紅色	簡約	高跟
紅色高跟鞋	3	0	7
簡約高跟鞋	0	2	7



目標會員購買商品

從行為資料中, 篩選出目標會員購買之商品

新增購買標籤

新增目標會員相關商品購買行為標籤 (1 - 購買、0 - 未購買) 作為模型預測使用

紅色高跟鞋

會員A	紅色	簡約	高跟	購買
紅色高跟鞋	3	0	7	1
簡約高跟鞋	0	2	7	0

_____ X ______ Y



■ 切分訓練集與測試集

01動機、問題與解方

- ── 訓練 80%、測試 20%
- 將資料放入模型
 - **T** XGBoost
- 模型結果
 - ____**預測準確率:84.3%**

	precision	recall	f1-score	support
0.0	0.90	0.93	0.91	80
1.0	0.14	0.11	0.12	9
accuracy			0.84	89
macro avg	0.52	0.52	0.52	89
weighted avg	0.83	0.84	0.83	89

XGBoost	預測購買	預測不購買
真實購買	1	8
真實不購買	6	74



商品購買預測

■ 輸入新商品 ■ 斷詞 ■ 建立向量空間

模型預測結果:

	SalePageTitle	巴黎	氣質	端莊	底涼	粉	金	綿羊 皮	韓式	療癒	 美女	短靴	平底 鞋	方頭鞋	聯名	閃耀	牛皮	拉鍊	鉚釘	model_output
0	商品A	0	1	1	0	0	0	0	0	0	 1	0	1	0	1	0	0	0	0	0
1	商品B	0	0	0	0	1	0	1	1	1	 0	0	1	0	0	0	0	0	0	0
2	商品C	0	0	0	1	1	1	0	1	1	 1	1	0	0	0	0	0	1	0	1
3	商品D	1	1	0	0	1	0	1	0	0	 0	1	0	0	1	1	0	0	1	0
4	商品E	0	0	0	0	1	1	0	0	0	 0	1	0	0	1	1	1	1	1	0
5	商品F	0	0	0	1	0	0	1	0	0	 0	1	0	1	1	1	0	1	1	0



建立每個會員的商品關鍵字

1. 整理每位客戶的商品標籤

將每位客戶的商品標籤按加權分數大小排序,排名越前面代表其為越能代表客戶的關鍵字

2. 建立使用function

建立一個function, 使用者只要輸入會員編號, 就可以產生該會員的前10 關鍵字及前50關鍵字

02 資料處理

3. 關鍵字呈現

前10關鍵字將按排名以表格方式呈現(含標籤及分數) 前50關鍵字將以文字雲的方式呈現



輸入會員編號 : zzzHm2cdbc30xk0pPbND0hTJC9KvNVW3vvOpTfzWsvo=



	Memberid	KeyWords	Count
0	zzzHm2cdbc30xk0pPbND0hTJC9KvNVW3vvOpTfzWsvo=	平底	246.66
1	zzzHm2cdbc30xk0pPbND0hTJC9KvNVW3vvOpTfzWsvo=	拖鞋	245.66
2	zzzHm2cdbc30xk0pPbND0hTJC9KvNVW3vvOpTfzWsvo=	涼拖	244.66
3	zzzHm2cdbc30xk0pPbND0hTJC9KvNVW3vvOpTfzWsvo=	涼拖鞋	244.66
4	zzzHm2cdbc30xk0pPbND0hTJC9KvNVW3vvOpTfzWsvo=	平底涼	238.66
5	zzzHm2cdbc30xk0pPbND0hTJC9KvNVW3vvOpTfzWsvo=	底涼	238.66
6	zzzHm2cdbc30xk0pPbND0hTJC9KvNVW3vvOpTfzWsvo=	最高	231.00
7	zzzHm2cdbc30xk0pPbND0hTJC9KvNVW3vvOpTfzWsvo=	現折	231.00
8	zzzHm2cdbc30xk0pPbND0hTJC9KvNVW3vvOpTfzWsvo=	高現	230.00
9	zzzHm2cdbc30xk0pPbND0hTJC9KvNVW3vvOpTfzWsvo=	高現折	230.00



前50關鍵字文字雲

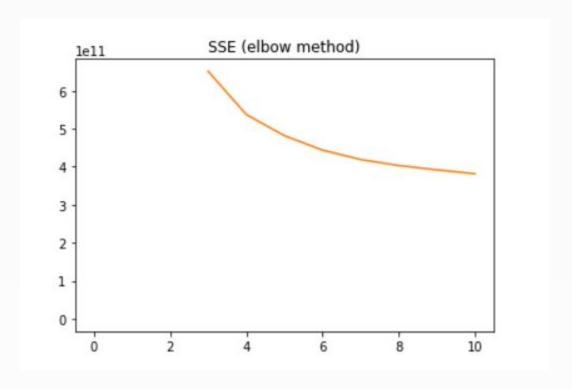
前10關鍵字



客戶分群

K-means

- 因會員數量過於龐大,我們以前十萬筆會員 為例進行分群
- 使用會員行為 (瀏覽、加入購物車、結帳) 的商品特徵詞表, 進行 k-means 分群
- 以手肘法找到最適合的群集數量大約為7

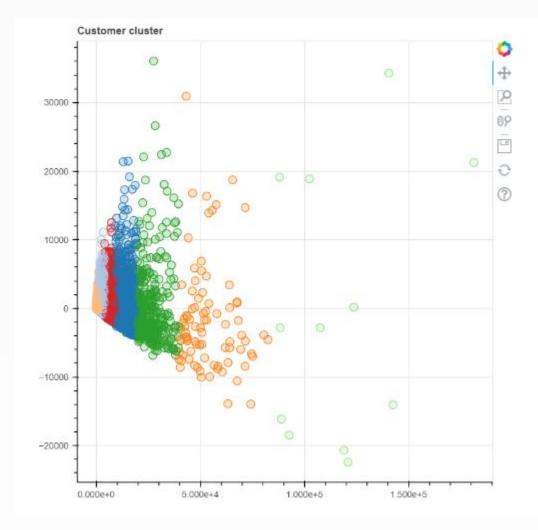




客戶分群

K-means

- 根據分群的結果可找出目標客群
- 將商品以「其他人也看過」等方式,推 薦給位於同一群集的會員



04

- 01 動機、問題與解方
- 02 資料處理
- 03 模型分析
- 04 總結



總結



產出為透過分析會員的行為資料,找出單一會員的喜好標籤,並使用機器學習方法產生個人化推薦商品列表



將商品喜愛類別相似的會員分群 ,製作其他人(同一集群)也看過 的推薦系統,發掘同一集群內的 潛在購買



藉由這次的專案較精準的推薦會員他可能有興趣的商品,以提升瀏覽對購買的轉換率與商店營收

未來展望

- 動態更新進行會員商品貼標,如:移動回測
- 因電腦硬體能力有限,目前藉由分析代表性會員向相關群體進行商品推薦 期望未來能針對每個消費者進行更精準的個人化商品推薦

Thank You