

國立中央大學

資訊工程學系
碩士論文

透過矩陣分解之多目標預測方法預測使
用者於特殊節日前之瀏覽行為變化

Predicting User's Browsing Tendency
During Holidays by Matrix Factorization
based Multi-objective Method

研究生：白國臻

指導教授：陳弘軒 博士

中華民國 一百零七年 六月

透過矩陣分解之多目標預測方法預測使用者於特殊節日前之瀏覽行為變化

Predicting User's Browsing Tendency During Holidays by Matrix Factorization based Multi-objective Method

中文摘要

近幾年電子商務公司在特殊節日的銷售額大於平日，然而從使用者瀏覽網頁的歷史紀錄中，我們發現僅有並非所有使用者在節日間皆比平時更常造訪電商網站，此現象表示銷售額的提升可能只是來自少數使用者所產生的影響，因此電子商務公司若能透過系統化的方式判別使用者行為，並給予不同行為表現使用者相對應的行銷手段，便能夠使市場行銷策略發揮更大的效能。

我們提出 Matrix Factorization based Multi-objective Method，當同時有多個待預測的目標時，本模型能有效利用多個預測目標之間可能存在互相影響的隱性因子。相較於為每個目標分別建立獨立的模型，本方法能有效減少模型需學習之參數，因此在訓練樣本數受限的情況，依然能達到有效的訓練結果。我們使用此方法同時預測使用者於特殊節日期間在不同類型網頁的行為變化，結果顯示：本方法在大多數時候能勝過單目標之訓練模型。

關鍵字：監督式學習、Matrix Factorization

透過矩陣分解之多目標預測方法預測使用者於特殊節日前之瀏覽行為變化

Predicting User's Browsing Tendency During Holidays by Matrix Factorization based Multi-objective Method

ABSTRACT

It is reported that sales by e-commerce companies were greater than usual during shopping holidays and festivals. However, based on users' browsing logs, we found that not all users visit e-commerce websites more often than they normally do during holidays. Therefore, the increase in sales may come from the purchase behaviors of a small number of users. If the e-commerce companies can systematically assess and analyze user behaviors, they might be able to apply customized marketing method to maximize the effectiveness of their sales strategies.

This study proposes a matrix factorization based multi-objective method, which effectively uses the latent variables that represent possible interactions among multiple targets. Compared with establishing separate models for each target, this method can effectively reduce the parameters that the model needs to learn, and can, therefore, achieve an effective training outcome even when training samples are limited. We use this method to simultaneously predict users' behaviors on different types of web pages during shopping holidays and festivals. The results show that this method can outperform the single target training model most of the time.

Keyword: Supervised learning, Matrix Factorization

目 錄

中文摘要.....	i
ABSTRACT.....	ii
目錄.....	iii
圖目錄.....	v
表目錄.....	vi
一、緒論.....	1
1.1 研究動機	1
1.2 研究目標	1
1.3 研究貢獻	2
1.4 論文架構	3
二、相關研究.....	4
2.1 使用者在 Pinterest 使用行為分析	4
2.2 使用者圈影響力之推薦策略	5
2.3 Matrix Factorization 介紹	5
三、MF based Multi-objective Method	8
3.1 MF based Multi-objective Method 概述	8
3.2 訓練模型	9
3.3 測試模型	10
3.4 Multi-objective method 與 Single-objective method 的比較	10
四、實驗介紹.....	12
4.1 問題描述	12
4.2 資料集介紹	14

4.2.1	真實資料集	14
4.2.2	建立擬真資料集	17
4.3	真實資料集前處理	19
4.4	特徵說明	20
4.5	預測分類器介紹	22
五、	實驗分析.....	23
5.1	實驗介紹	23
5.2	模型預測效果之評估指標	23
5.3	真實資料集之結果分析	24
5.4	資料量對 MF based Multi-objective Method 之影響分析	26
六、	結論與未來展望.....	29
6.1	結論	29
6.2	未來展望	30
	參考文獻.....	32

圖目錄

圖 2.1	購買日前 28 天，每日購買者某種行為比例除以非購買者特定行為比例（灰色部分為標準差）	4
圖 2.2	以矩陣表示使用者對事件的評分（問號表示為未知的評分）	6
圖 3.1	Matrix Factorization based Multi-objective Method	9
圖 4.1	使用者在 12 月每日點擊購物網頁次數	13
圖 4.2	使用者瀏覽網頁數量之經驗分布函數	15
圖 4.3	使用者相關資訊分布，（a）性別（b）年齡（c）感情狀態	15
圖 4.4	三種資料集之使用者人數集合關係	16
圖 4.5	所有使用者每天瀏覽預測目標之餐廳、購物、旅遊網頁佔總瀏覽量比例	16
圖 4.6	所有使用者每天瀏覽預測目標之教育、娛樂、遊戲網頁佔總瀏覽量比例	17
圖 4.7	使用者人口資訊分布統計，（a）性別（b）年齡（c）感情狀態	17
圖 4.8	使用者人口結構分布概念圖	18
圖 4.9	網頁依受歡迎程度排序之使用者點擊網頁量累計	19
圖 4.10	五種最受歡迎的網頁類型佔所有網頁的瀏覽比例	20
圖 5.1	判斷測試表	24
圖 5.2	MF based Multi-objective Method	27
圖 5.3	general linear model	27
圖 5.4	訓練資料量對預測結果之影響	28

表 目 錄

表 4.1	使用者瀏覽網頁數量之綜合統計	14
表 4.2	使用者人數與瀏覽個類型網頁之行爲變化統計	16
表 5.1	真實資料集在四種模型預測不同網頁類型之 F_1 score	26

一、緒論

1.1 研究動機

當特殊節日來臨前，例如：情人節、中秋節、聖誕節等，廠商為因應消費者有過節送禮的需求，會釋出應景商品的優惠活動，藉以吸引更多消費者，並增加產品銷售。

而現在大多數人的生活行為已不可缺少瀏覽網頁，對於生活中突發狀況或行程規劃，都可能透過瀏覽網頁來解決，因此使用者的生活事件會表現在瀏覽網頁的歷史紀錄中，消費者會開始透過網頁瀏覽對節日進行安排，節日來臨前網頁瀏覽行為的變化將會表現在網頁歷史紀錄之中。

由於實驗室擁有一份較為完整由 600 位使用者所提供瀏覽網頁的歷史紀錄以及個人資訊的資料集，在進行資料前處理的過程中，我們發現並不是所有使用者都會增加瀏覽電子商務網站，因此，如果能透過使用者的個人資訊及瀏覽網頁的習慣，找出哪些使用者在特殊節日來臨前會增加瀏覽電商網站比例，讓電子商務公司找出要積極對哪些用戶進行銷售策略，就能夠增加商品的推薦效果。

1.2 研究目標

本論文主要研究目標是透過使用者的個人資訊及以瀏覽網頁的歷史紀錄當作瀏覽習慣，來預測使用者在特殊節日來臨前，是否會改變使用者習慣的瀏覽行為，而使用者對節日的活動安排不會只有一種，因此從生活所需(食、衣、住、行、育、樂)之網頁類型進行研究。

考量到使用者在瀏覽不同網頁的行為之間可能存在著某種相互影響的關聯性，若每次僅針對一種網頁類型進行研究，可能忽略此重要影響，因此提出以 Matrix Factorization 為基礎的 Matrix Factorization based Multi-objective Method，希望能同時針對多種目標進行預測，找出潛在影響，同時提升模型的學習效果。

此模型擁有 Matrix Factorization 有效降低訓練參數之特性，使訓練資料有限的情況，透過訓練較少參數達到較佳的預測效果，為了討論資料量對此模型預測效果之影響，我們除了使用真實資料進行實驗之外，同時也根據使用者在節日前可能的瀏覽行為變化建立擬真資料集，再對模型的預測效果進行分析。

1.3 研究貢獻

對於預測多種目標之任務，在不同目標之間可能存在著不容易直接觀察的潛在關係，因此我們提出 Matrix Factorization based Multi-objective Method，透過同時對多種目標進行學習，找出目標之間相互的影響，並利用預測任務的特徵進行預測。並且當資料集人數受限的情況，利用減少學習參數，使模型進行有效的預測。

我們從使用者瀏覽各類網頁的行為發現，並不是所有使用者在節日來臨前，都會比平常更常訪問電商網站，此結果表示電子商務公司的銷售額增加 [1, 2]，可能是來自少數使用者所造成的影響。

透過本篇論文提出的研究成果，能夠幫助電子商務公司辨別不同使用行為的使用者，給予他們不同的行銷策略，例如：對於在節日來臨前可能增加瀏覽購物網頁的使用者，廠商能夠增加廣告或釋出更多優惠商品訊息，來吸引消費者購買。

1.4 論文架構

本篇論文共分爲六個章節，其架構如下：

第一章、介紹本篇論文的研究動機、目標以及貢獻，並說明論文架構。

第二章、介紹推薦系統的相關研究，並詳細介紹 Matrix Factorization。

第三章、介紹 Matrix Factorization based Multi-objective Method 的參數設置及訓練和測試模型。

第四章、詳細說明研究問題及實驗的資料集與資料前處理的過程。

第五章、展現實驗在不同資料集及不同模型的效能評估、討論。

第六章、本篇論文之結論與未來展望。

二、 相關研究

本章節將介紹推薦策略的相關研究，另外針對我們提出的 Matrix Factorization based Multi-objective Method 之基礎 Matrix Factorization 做介紹。

2.1 使用者在 Pinterest 使用行為分析

隨著網路購物行為增加，了解消費者線上購物行為的變化更加重要，其中的關鍵在於使用者的行為隨著時間改變，所表現的購買意願變化。原因在於消費者的購買意願不是立即決定購買，而是隨著時間慢慢累積，因此此應用對比購買者與非購買者在 Pinterest 上的行為（搜尋、關注、點擊到外部網頁及保存），分析購買者的線上行為隨著時間的變化 [3]。

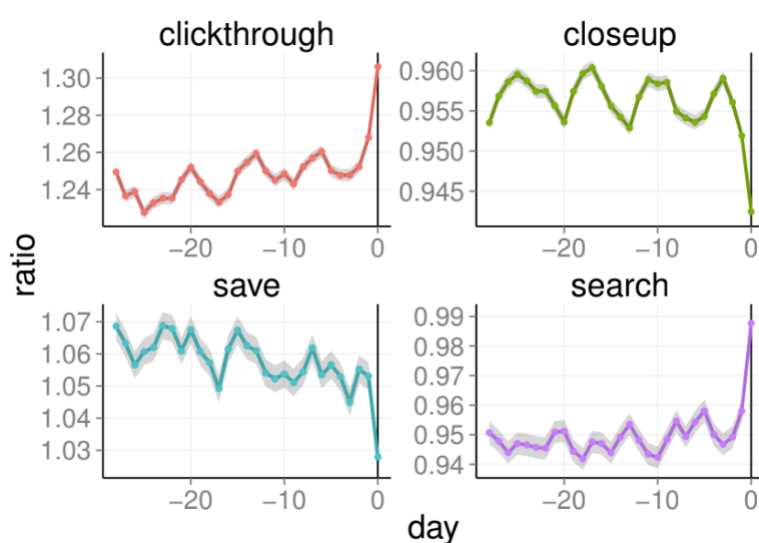


圖 2.1: 購買日前 28 天，每日購買者某種行為比例除以非購買者特定行為比例（灰色部分為標準差）

透過圖 2.1表現購買者與非購買者在購買行為約前五天的使用行為呈現出明顯差異，因此利用使用者在購買前 N 天在 Pinterest 上的使用行為，預測此使用者在未來是否會成為購買者。

此應用是利用使用者在 Pinterest 上的使用行為進行預測，透過使用者訪問特定電商，並留下瀏覽行為進行推薦，而本篇論文則是利用使用者跨站的瀏覽歷史紀錄，預測使用者是否會增加瀏覽特定網頁的比例。

2.2 使用者圈影響力之推薦策略

廠商的行銷手法除了釋出優惠通知及廣告訊息之外，同時能透過明星效應影響使用者購買意願，使用者可能受到自己喜愛的偶像或是朋友圈中較受歡迎的對象而對某些商品產生興趣，同樣的效應在社群網路上也存在相似的情況。

社群網路是透過使用者與其他人互動所組成，因此能夠利用社群網路中人與人互動之行為進行分析使用者間彼此相互影響的程度，找出社群網路中具有較大影響力的成員 [4, 5]。藉由提供免費產品的試用要求分享自己的使用經驗，讓產品在朋友圈中傳播，而吸引更多的使用者對此產品產生興趣，便能夠達到新產品宣傳的效果。

此應用的推薦策略是透過社群網路找出較有影響力的使用者，使產品在使用者圈發揮感染力，超越廣告策略的影響，而本篇論文則是以不同角度進行推薦，透過使用者的瀏覽紀錄，找出可能增加瀏覽比例的使用者，並增加廣告強度，藉此提升產品的銷量。

2.3 Matrix Factorization 介紹

作為推薦系統的方法有很多，考慮到預測任務之間可能存在某種不容易直接觀察卻相互影響的關係，而利用尋找不同任務之間的隱性因子（latent factor）之 Matrix Factorization 為基礎設計方法，以下將介紹 Matrix Factorization [6]。

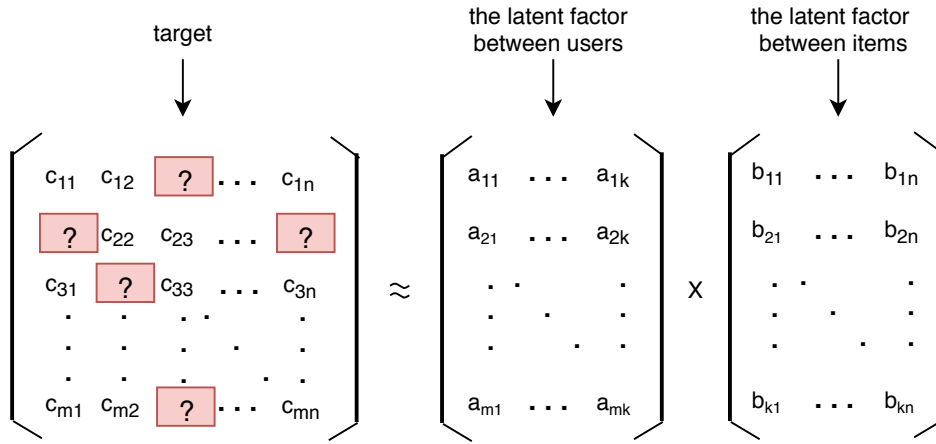


圖 2.2: 以矩陣表示使用者對事件的評分（問號表示為未知的評分）

假設擁有使用者對事件的評分，則將每位使用者對事件的評分存成矩陣形式，以矩陣 C 做表示，每列（row）表示為一位使用者，每行（column）表示為事件的評分，但其中可能有幾項使用者的評分未知，如（圖 2.2），因此，透過學習兩個較小維度矩陣 A 及矩陣 B 進行矩陣相乘的結果盡可能逼近矩陣 C 中已知的值，來預測矩陣 C 中未知的值。當矩陣中的參數經過反覆學習，則兩個小矩陣能夠共同表示事件被評分的背後所存在的影響原因，其中矩陣 A 表示影響使用者對事件評分的隱性因子，而矩陣 B 表示被評分事件之間存在的隱性因子。

矩陣 A 及矩陣 B 的學習方法是預先給定隨機參數值為初始化參數，以 gradient descent 的方式，將損失函數（loss function）分別對矩陣 A 及矩陣 B 偏微分，透過反覆迭代學習其中的參數。而兩矩陣相乘為預測結果，希望結果與目標矩陣 C 內已知值的誤差越小越好，因此以（式 2.1）表示最小化損失函數。

$$Loss = \min_{a,b} \sum_{\forall m,n} (c_{mn} - \sum_{\forall k} a_{mk} \cdot b_{kn})^2 \quad (2.1)$$

其中損失函數的決定方法有很多，如果目標矩陣 C 擁有的已知值數量太少，容易導致過擬合（overfitting）的情況發生，因此加入正規化（式 2.2），使參數

學習的大小受到限制與懲罰，並利用 λ 控制正規化的影響程度。

$$Loss = \min_{a,b} \sum_{\forall m,n} (c_{mn} - \sum_{\forall k} a_{mk} \cdot b_{kn})^2 + \lambda (\|a\|^2 + \|b\|^2) \quad (2.2)$$

以 Matrix Factorization 為設計基礎的方法及應用很廣泛，除了商品推薦，也可應用於處理文件分析 [7]，原因在於 Matrix Factorization 擁有找出不同任務之間可能存在的關聯性之特性，有效透過文章之詞彙歸納文章主題。

三、 MF based Multi-objective Method

以 Matrix Factorization 為基礎，設計 Matrix Factorization based Multi-objective Method，此方法能夠同時預測多種目標，並藉由 Matrix Factorization 以兩種較小維度矩陣學習不同任務之間隱藏關聯性的想法，幫助減少需學習之參數，達到提升預測準確度之效果。

3.1 MF based Multi-objective Method 概述

藉由 Matrix Factorization 將預測目標拆解為表示潛藏的影響因子之概念，我們將表示潛藏在預測目標背後之關係矩陣（矩陣 R ）拆解為兩個較小維度矩陣（矩陣 P 及矩陣 Q ）進行矩陣相乘，矩陣 P 及矩陣 Q 之初始值存放隨機產生之亂數，當模型進行多次訓練，使預測結果逼近真實結果，反覆更新參數，則能夠學習到由矩陣 P 及矩陣 Q 所共同決定的隱性因子。透過使用者特徵矩陣 F 與矩陣 P 及矩陣 Q 的學習結果，達到在預測時同時預測多項目標之效果，預測目標矩陣 T 之每一行（column）表示為其中一項預測結果，如圖 3.1。

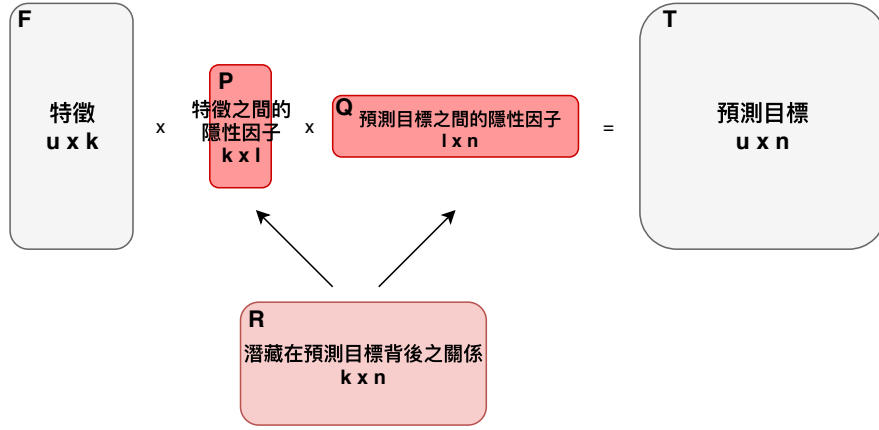


圖 3.1: Matrix Factorization based Multi-objective Method

3.2 訓練模型

訓練模型時，必需透過反覆學習，每次微調矩陣 P 及矩陣 Q 內的所有參數，達到減少目標與預測結果之間的差異，我們透過損失函數（式 3.1）來表示計算兩者間差異的方式，其中 \hat{t}_{un} 表示為 u 對目標 n 的預測結果。

$$Loss = \sum_{\forall un} (\hat{t}_{un} - t_{un})^2 \quad (3.1)$$

每次對矩陣 P 微調的方法以式 3.2 表示， α 為學習率，微調的大小為損失函數對矩陣 P 的參數偏微分（式 3.3）。

$$p_{kl} = p_{kl} - \alpha \frac{\partial loss}{\partial p_{kl}} \quad (3.2)$$

$$\frac{\partial loss}{\partial p_{kl}} = 2 \sum_{\forall un} \{ [\sum_{\forall k} f_{uk} \cdot (\sum_{\forall l} p_{kl} \cdot q_{ln})] - t_{un} \} \cdot (f_{uk} \cdot q_{ln}) \quad (3.3)$$

每次對矩陣 Q 微調的方法以式 3.4 表示， α 為學習率，微調的大小為損失函數對矩陣 Q 的參數偏微分（式 3.5）。

$$q_{ln} = q_{ln} - \alpha \frac{\partial loss}{\partial q_{ln}} \quad (3.4)$$

$$\frac{\partial loss}{\partial q_{ln}} = 2 \sum_{\forall un} \{ [\sum_{\forall k} f_{uk} \cdot (\sum_{\forall l} p_{kl} \cdot q_{ln})] - t_{un} \} \cdot (f_{uk} \cdot p_{kl}) \quad (3.5)$$

3.3 測試模型

測試模型時，將所有測試使用者之特徵存入矩陣 F ，並與訓練得矩陣 P 及矩陣 Q 進行三個矩陣之矩陣相乘，則得到表示預測目標的矩陣 T 。

預測目標為使用者瀏覽行為可能上升或下降，屬於二元分類問題，但將矩陣 F 、矩陣 P 及矩陣 Q 進行矩陣相乘，所產生的結果並非 1（上升）或 0（下降），因此，在訓練時我們統計所有使用者行為下降數為 i 位，對應訓練結果值由小排序到大，第 i 位使用者之矩陣乘積結果作為行為變化臨界值，若模型在訓練及測試之預測結果值大於行為變化臨界值，表示 1，反之則表示為 0。

3.4 Multi-objective method 與 Single-objective method 的比較

一般線性學習每次僅預測單一目標，並學習所有特徵對預測目標之權重，每次學習 k 個未知參數，若預測目標增加為 N 種，必須反覆學習 N 次，則總共學習 $k \cdot n$ 個未知參數。我們提出的方法能夠同時預測多種目標，將矩陣 R 拆解為矩陣 P 及矩陣 Q ，使學習未知參數有效減少為 $(k \cdot l) + (l \cdot n)$ 個，其中要求 l 應該遠小於 k 及 n ，當所擁有的學習資料量受限，減少需學習之參數數量，可相對於一般線性學習，達到更好的預測效果。因此此方法能夠有效降低模型發生 overfitting 的機率，並且減少模型進行訓練的時間。

Matrix Factorization based Multi-objective Method 演算法如下：

Algorithm 1 Matrix Factorization based Multi-objective Method

Input:

t_u : training users;
 p : latent factor p ;
 q : latent factor q ;
 $test_user$: test users

Output:

$training_predict$
 $test_predict$

```
1:  $p, q \leftarrow$  initial with random values
2: while termination condition is not met do
3:   for  $(u, n)$  in the target matrix do
4:     for all  $k \in p$  do
5:       for all  $l \in q$  do
6:          $pred\_t \leftarrow t_u \cdot p \cdot q$ 
7:          $p_{kl} \leftarrow p_{kl} - \alpha \frac{\partial loss}{\partial p_{kl}}$  (according to equation 3.3)
8:          $q_{ln} \leftarrow q_{ln} - \alpha \frac{\partial loss}{\partial q_{ln}}$  (according to equation 3.5)
9:       end for
10:    end for
11:  end for
12: end while
13:  $training\_target \leftarrow t_u \cdot p \cdot q$ 
14:  $test\_target \leftarrow test\_user \cdot p \cdot q$ 
```

四、實驗介紹

本章節主要討論五個部分，首先第一節完整說明整篇論文主要問題，包含建構時的想法與過程。第二節介紹在實驗中主要使用資料集的詳細組織架構，以及將資料統計後，發現使用者所表現的特殊行為。第三節介紹真實資料的前處理過程與處理想法。第四節說明在真實資料集中，選用為預測特徵的原因。第五節將與我們提出的方法進行比較的監督式學習法做基本介紹。

4.1 問題描述

考量到使用者在節日時可能會安排特殊活動進而改變平常的瀏覽習慣，因此我們收集使用者瀏覽網頁歷史紀錄做為主要使用的資料集，其中要求這些提供瀏覽網頁歷史紀錄的使用者都必須擁有線上購物的經驗，因此起初將“瀏覽購物網頁”行為定義為主要研究目標。但事實上使用者可能有多項活動安排，因此希望透過同時考慮多種不同網頁類型的瀏覽行為，找出不同網頁類型的瀏覽習慣之間可能存在著某些互相影響的隱性因子，所以將研究行為延伸到瀏覽食：餐廳類、衣：購物類、住行：旅遊類、育：教育類、樂：娛樂類及遊戲類之網頁行為，希望能預測使用者多種不同網頁類型的瀏覽行為，同時提升模型的學習效能。

大部分網頁瀏覽歷史紀錄的時間集中在 2016 年 8 月到 2016 年 12 月，由於考慮聖誕節之前的使用者瀏覽歷史紀錄較為完善，因此我們選擇以聖誕節作為主要研究節日。

其中所有使用者都被要求必須擁有瀏覽購物網頁的歷史紀錄，所以資料集中



圖 4.1: 使用者在 12 月每日點擊購物網頁次數

使用者在購物行為上的表現較為突出，因此以使用者瀏覽購物網頁的角度來做考量，統計所有使用者在 12 月每一天點擊購物網頁的次數（圖 4.1），12 月初使用者點擊次數大約落在 1,000 到 5,000 次區間，我們發現 12 月 12 日點擊次數明顯出現一個高峰值 7,461 次，表示有使用者開始明顯出現購物的需求或普遍使用者開始增加點擊購物網頁，因此將 12 月 12 日至 12 月 25 日定義為使用者在特殊節日前之行為變化日期，同樣的表示取節日前 14 天（包含節日當天）作為產生行為變化的分界點。

我們將每位使用者最早的瀏覽歷史紀錄之日期至聖誕節前 14 天當作為使用者平時瀏覽各類型網頁的習慣，若在特殊節日前瀏覽某一種類型網頁比例高於該使用者平時瀏覽該類型網頁的習慣，則定義使用者瀏覽該網頁類型的比例上升，反之則為減少。其中使用者瀏覽各類型網頁比例，定義為每位使用者在節日來臨前 14 天至聖誕節對某種網頁瀏覽數量除以所有被挑選作為預測特徵網頁之總瀏覽量（式 4.1）。

$$cate_i \text{ ratio} = \frac{\text{number of visits on the } cate_i \text{ websites during holiday period}}{\text{total number of visits during holiday period}} \quad (4.1)$$

因此若模型能有效判斷出擁有某些瀏覽習慣或個人資訊的使用者在特殊節日來臨前可能增加瀏覽某種類型網頁，則能夠幫助電商了解應該要更積極向哪些用戶投放何種類型廣告，增加廣告的實行效益。

4.2 資料集介紹

在本篇論文中，總共使用兩種資料集來進行預測及研究分析，分別是真實的資料集以及擬真資料集。在真實資料集，收集使用者詳細的個人資訊以及他們在 Google Chrome 瀏覽器中完整的瀏覽紀錄，預測使用者在節日來臨前之行爲變化，另外，爲了討論資料量對我們提出的模型在預測效果之影響，我們根據使用者在節日前可能的瀏覽行爲變化建立模仿資料集，以下爲兩組資料集之完整介紹。

4.2.1 真實資料集

本篇論文使用的真實資料集透過 Google Chrome 瀏覽器中的 plugin 詳細收集 672 位使用者留下的網頁瀏覽歷史紀錄，所有使用者共累計 12,837,216 筆瀏覽紀錄，透過（圖 4.2）、（表 4.1）分別表現每位使用者瀏覽網頁量的經驗分布函數及瀏覽網頁數量統計，圖中顯示出使用者瀏覽量介於 44 到 50,000 筆之間。

表 4.1: 使用者瀏覽網頁數量之綜合統計

min	1st Quartile	Median	Mean	3rd Quartile	max
44	4,239	13,335	19,103	26,698	130,992

在詳細收集的 672 位使用者中，我們取得使用者的同意，得到 508 位用戶自主公開的個人資訊，例如：性別、年齡、感情狀態等等，以圓餅圖（圖 4.3）來表

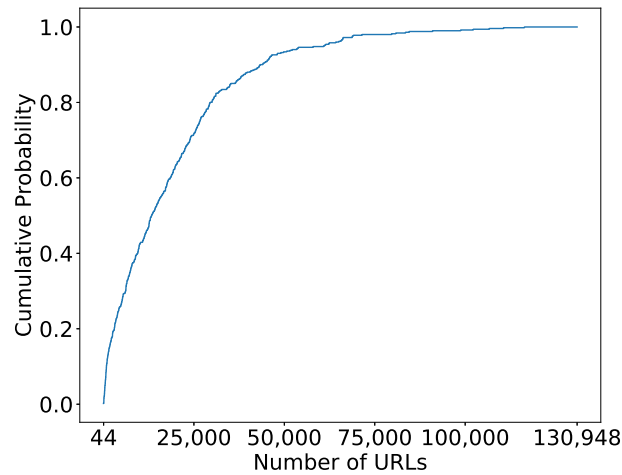


圖 4.2: 使用者瀏覽網頁數量之經驗分布函數

示這些用戶的人口資訊分布統計。而預測目標是特殊節日前 14 天的使用行為，因此在聖誕節前 14 天使用者必須留下瀏覽紀錄才能作為預測目標，因此實際能夠做為分析的使用者僅有 224 位（圖 4.4）。在實際進行研究的使用者當中，聖誕節來臨前 14 天使用者人數與瀏覽個類型網頁之行為變化統計如表 4.2。

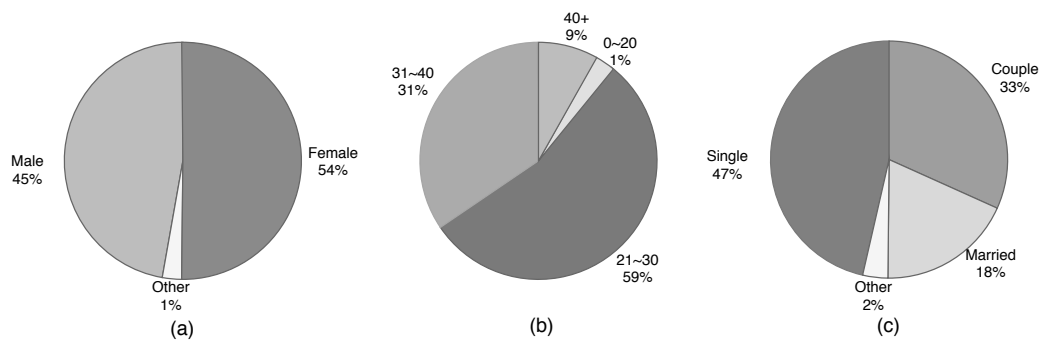


圖 4.3: 使用者相關資訊分布，（a）性別（b）年齡（c）感情狀態

為了更加了解使用者瀏覽網頁的習慣，我們詳細長時間觀察使用者訪問預測目標之六種類型網頁的比例，圖 4.5、圖 4.6 呈現所有使用者在 8 月到 12 月每天瀏覽預測目標之六種網頁比例，以購物網頁為例，我們觀察到所有使用者每日瀏覽購物網頁比例落在 4% 到 10% 區間，而每日瀏覽六種網頁的比例在聖誕節來臨



圖 4.4: 三種資料集之使用者人數集合關係

表 4.2: 使用者人數與瀏覽個類型網頁之行爲變化統計

瀏覽趨勢	餐廳類	購物類	旅行類	教育類	娛樂類	遊戲類
瀏覽比例上升	68	95	79	83	81	69
瀏覽比例下降	156	129	145	141	143	155

前，並沒有發生明顯的上升，此結果說明並不是所有使用者對每一種類型網頁都有需求，因此造成網路商城銷售額增加可能只是源自於少數消費者的需求。若能精準地從所有使用者中找出節日前可能提高某種類型網頁瀏覽率的用戶，增加對這類型使用者投放廣告，便能增加電商的廣告效益。

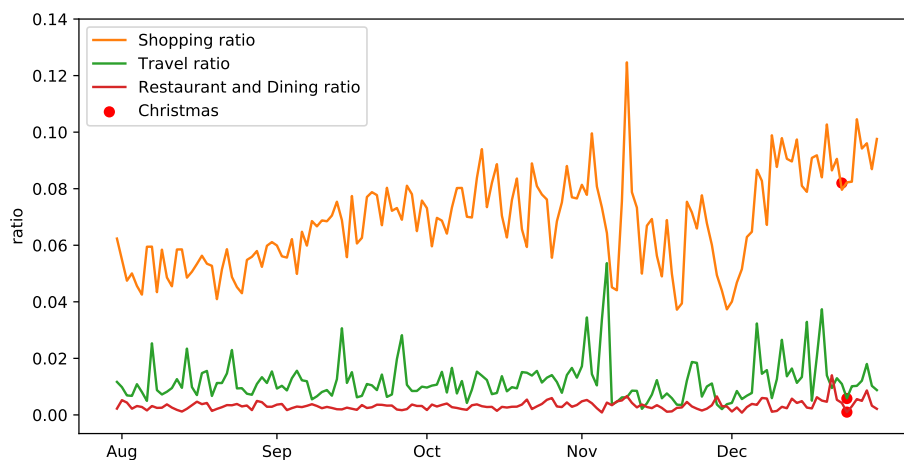


圖 4.5: 所有使用者每天瀏覽預測目標之餐廳、購物、旅遊網頁佔總瀏覽量比例

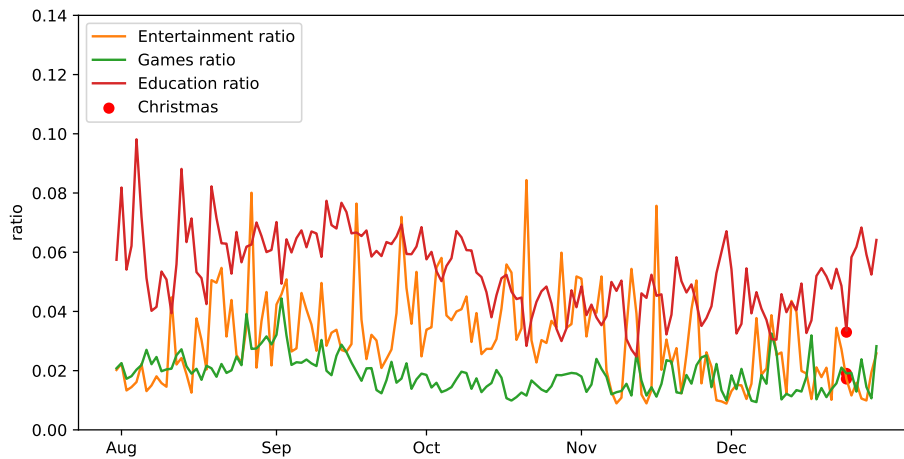


圖 4.6: 所有使用者每天瀏覽預測目標之教育、娛樂、遊戲網頁佔總瀏覽量比例

4.2.2 建立擬真資料集

由於真實資料集中的使用者數受限，而為了討論資料量對 Matrix Factorization based Multi-objective Method 與一般線性方法在預測準確度之影響，因此我們根據真實資料集整理完的架構，生成使用者之個人資訊以及在平時和節日前瀏覽行為變化的資料集，以下為擬真資料集之完整介紹。

我們將資料總數生成 1,000 位使用者來進行分析，其中個人資訊包含：

(1) 性別：男性或女性、(2) 年齡：共分為四個區間 20 歲以下、21～30 歲、31～40 歲或 41 歲以上、(3) 感情狀態：單身、交往中或已婚，以圓餅圖 (圖 4.7) 詳細表示人口資訊的分布統計。

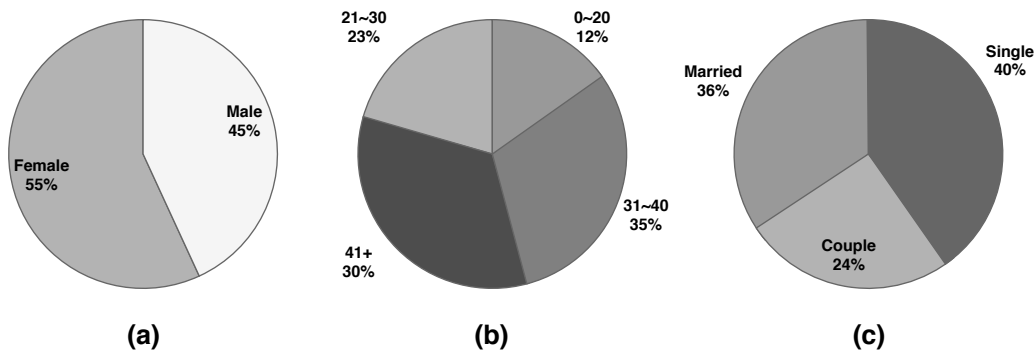


圖 4.7: 使用者人口資訊分布統計，(a) 性別 (b) 年齡 (c) 感情狀態

在使用者個人資訊中，除了性別，其他個人資訊是以樹狀圖的概念生成，如圖 4.8，例如：年齡根據不同性別生成不同比例的使用者、感情狀態根據不同年齡生成不同比例的使用者，目的是為了確定資料集中存在各種人格資訊的使用者。而使用者在各類型網頁的瀏覽率為隨機生成，則能夠盡可能模擬各種使用者的行為表現。

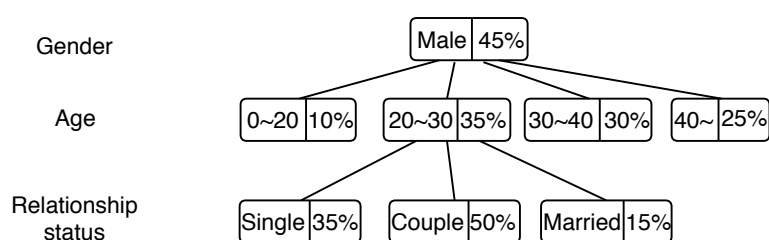


圖 4.8: 使用者人口結構分布概念圖

使用者瀏覽各類型網頁比例，配合真實資料整理完的架構進行建立。分別為不同使用者生成“特殊節日前 14 天”之前瀏覽各種網頁的比例，表示使用者平時瀏覽習慣，並生成使用者在特殊節日前 14 天對各種預測目標網頁的瀏覽率，做為將預測節日前的行為，因此特殊節日前 14 天瀏覽網頁的比例減平時瀏覽習慣，結果值為正數表示增加瀏覽，若結果為負數則表示減少瀏覽，做為待預測的真實目標。

對於預測目標之間並沒有建立影響不同目標的隱性因子，我們透過給定的使用者特徵結合各特徵對目標的重要性，並計算所有使用者對此運算結果之平均，若某位使用者之計算結果值大於平均表示增加瀏覽，反之則表示減少，讓使用者瀏覽各種網頁的行為變化呈現均勻分布。

因此每一種目標單純僅受使用者特徵所影響，並且在目標之間並不存在相互影響的關聯性，使實驗對於我們所提出的方法及學習使用者特徵對預測目標之權重，在擁有不同使用者資料數之實驗分析能夠達到公平。

4.3 真實資料集前處理

在真實資料集中使用兩個部分的資料，分別是使用者瀏覽網頁的歷史紀錄以及他們的個人資訊。從使用者瀏覽網頁的紀錄中我們觀察到用戶訪問之網頁分布不均，大多數使用者的瀏覽行為集中在較受歡迎的網頁，而較為冷門的網頁甚至僅少數使用者訪問，透過圖 4.9 表示網頁依受歡迎程度排序之使用者點擊網頁量累計，例如：最受歡迎的網頁 <https://www.facebook.com/>，占使用者瀏覽網頁總量的 27.6%，因此若直接使用用戶瀏覽網頁作為預測目標，導致網頁總數過多，而且大多數的網頁點擊次數都趨近於 0，且網頁過於繁雜，沒辦法進行有效運用與分析。

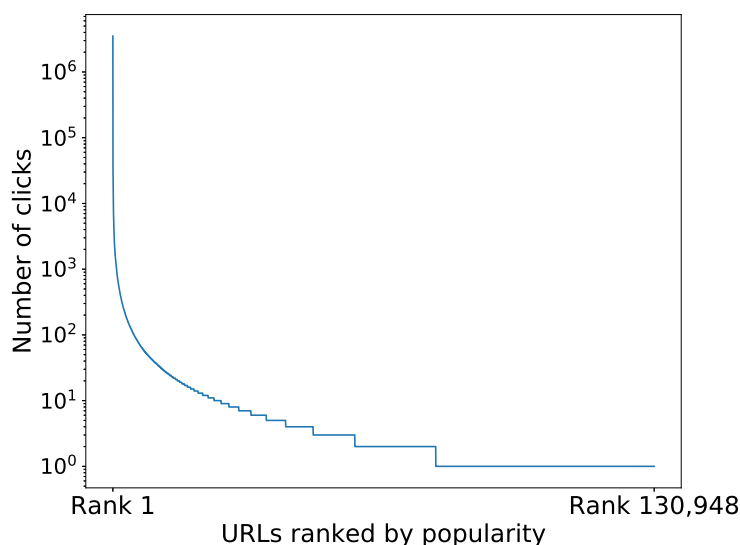


圖 4.9: 網頁依受歡迎程度排序之使用者點擊網頁量累計

因此我們將使用者瀏覽每一筆網址透過網頁分類器¹整理為不同網頁類型，例如：瀏覽網址為 <https://shopping.pchome.com.tw/>，歸類為 Shopping 類型、瀏覽網址為 <https://www.vscinemas.com.tw/>，則歸類為 Entertainment 類型，

¹<http://www.fortiguard.com/webfilter>

共將所有網頁歸納為 88 種類別，其中分別以 Social Networking、Search Engines and Portals、Web-based Email 三種類別的瀏覽量最高，共占全類別的 55%，另外將預測的餐廳、購物、旅遊、教育、娛樂及遊戲類型網頁分別占 0.3%、7%、1.2%、5.4%、3.3% 及 1.9%，透過圓餅圖(圖 4.10)表示五種最受歡迎的網頁類別瀏覽比例。

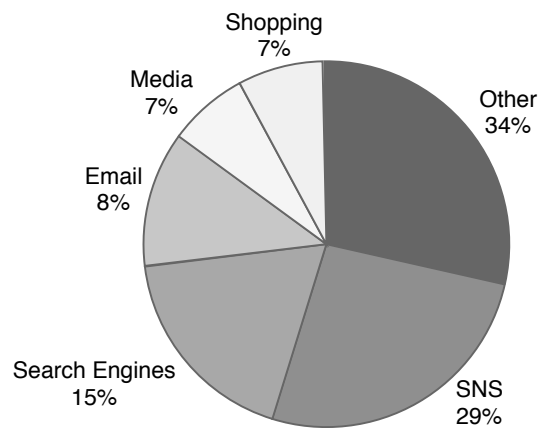


圖 4.10: 五種最受歡迎的網頁類型佔所有網頁的瀏覽比例

其中在使用者瀏覽網頁紀錄中，<https://www.facebook.com/>、<https://www.google.com.tw/>、<https://mail.google.com/> 及 <https://www.youtube.com/> 為較熱門之網頁，分別佔所有網頁紀錄之 27.6%、11.8%、6.6% 及 6.4%，因此我們將此四筆網址分別獨立為一種新的網頁類型，因此所有網頁種類增加為 92 種。

4.4 特徵說明

主要使用的特徵分為兩大類，分別是使用使用者自主提供的個人資訊以及使用者在網頁瀏覽留下的歷史紀錄，以下將詳細介紹特徵之選擇。

其中挑選做為特徵之人格屬性包含(1)性別：男性、女性及其他、(2)年

齡：共分為五個區間 20 歲以下、21 ~ 30 歲、31 ~ 40 歲、41 ~ 50 歲及 51 歲以上、(3)感情狀態：單身、交往中、已婚及其他，使用者對網頁的瀏覽行為可能受到個人資訊不同而有所影響，因此將此三種使用者個人資訊做為預測特徵。而使用者個人資訊之類別沒有順序分別，無法轉換為有大小差異的數值表示，因此以 one hot encoding 的方式表示此三種特徵。

另外使用者所瀏覽的網頁內容可能受到使用者本身的個性或生活事件所影響，以購物行為為例：個性較為外向的使用者，可能更頻繁的使用社群媒體與他人互動，而該使用者在特殊節日更有可能參與社交活動，產生送禮的需求，因此將使用者平時習慣瀏覽各類網頁之比例做為分類特徵。其中我們以生活所需要務共六種網頁作為預測使用者瀏覽行為，因此挑選六種網頁的瀏覽比例做為特徵，並挑選所有使用者總瀏覽次數前三高之熱門網頁做為每位使用者的其中三種特徵，包含 <https://www.facebook.com/>、<https://www.google.com.tw/> 以及 <https://mail.google.com/>，考慮到我們提出的方法在資料受限的情況若使用過多特徵可能導致 overfitting，沒辦法將所有種類的瀏覽類別都做為預測特徵，因此將此九種網頁瀏覽比例做為預測特徵。

每位使用者平時習慣（分別自每位使用者留下最早瀏覽歷史紀錄起，至聖誕節前 14 天）瀏覽不同類型網頁比例（式 4.2），以使用者對某種網頁瀏覽數量除以所有被挑選作為預測特徵網頁之總瀏覽量。

$$cate_i \text{ ratio} = \frac{\text{number of visits on the } cate_i \text{ websites before holiday period}}{\text{total number of visits before holiday period}} \quad (4.2)$$

4.5 預測分類器介紹

預測目標為使用者瀏覽行為可能增加或減少，屬於二元分類問題，因此透過監督式學習分別預測使用者在節日來臨前可能瀏覽各種類型網頁的行為變化，表示每一種監督式學習每次僅預測一種網頁的行為變化。所使用的監督式學習分別有 K-Nearest Neighbor (KNN)、Logistic Regression 以及 Support Vector Machine (SVM)，以下介紹各種方法。

首先，k-NN 依據使用者特徵為座標，選 K 個相鄰最近的使用者行為表現做為預測的參考，並取大部分相鄰使用者的行為表現，定義為使用者的預測結果。

Logistic Regression 透過學習所有特徵不同權重，並將特徵及權重之運算結果相加，與 Linear Regression 相似，差別在於利用 sigmoid 函數將輸出結果壓縮到 0 至 1 區間，表示屬於某一種分類機率，並以 cross entropy 做為損失函數。

SVM 同樣是將 Linear Regression 的結果做為分類依據，與 Logistic Regression 最大差別為 SVM 以 hinge loss 做為損失函數。

五、實驗分析

這個章節總共分成四個部分，首先說明如何分配資料集進行模型學習，以及如何評估所有模型之預測效果。接著介紹真實資料集在所有模型的參數，並對實驗的結果提出分析與討論。最後呈現當擁有的使用者數不同對 Matrix Factorization based Multi-objective Method 之影響。

5.1 實驗介紹

我們將資料以交叉驗證 (5-fold cross validation) 的方法將資料集切割成 5 份子資料集，每次分別取其中 4 份子資料集當作訓練資料，剩下 1 份當作測試資料，反覆進行 5 次的訓練及測試，並且每次進行測試的子資料集不重複，將保證所有資料皆進行測試，最後取 5 次測試結果的評估指標分數平均，進行模型分類效果比較。

5.2 模型預測效果之評估指標

預測使用者在節日來臨前之瀏覽各類型網頁行為變化為上升或下降，屬於二元分類問題，並且在我們的資料集顯示使用者在節日來臨前之行為變化比例失衡，因此我們選擇以 F_1 score 做為模型預測效果的評估指標， F_1 score 同時考慮預測之召回率 (*recall*) 與精確率 (*precision*)，因此若 F_1 score 之分數較高，表示分類模型在預測使用者行為上升與下降同時取得較佳之表現。其中召回率與精確率以圖 5.1 表示之 True Positive (TP)、False Positive (FP)、True Negative (TN)、False Negative (FN) 計算。

$$F_1 \text{ score} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (5.1)$$

$$\text{Precision} = \frac{tp}{tp + fp} \quad (5.2)$$

$$\text{Recall} = \frac{tp}{tp + fn} \quad (5.3)$$

		True condition	
		Condition positive	Condition negative
Predicted condition	Predicted condition positive	True positive (TP)	False positive (FP)
	Predicted condition negative	False negative (FN)	True negative (TN)

圖 5.1: 判斷測試表

5.3 真實資料集之結果分析

在真實資料集，我們比較三種監督式學習（k-NN、Logistic regression 以及 SVM）以及我們提出的 Matrix Factorization based Multi-objective Method 作為分類模型的預測結果表現。並利用網格搜尋（Grid Search）的方式，找出各種分類方法效果相對較佳之參數。

以 k-NN 作為分類模型時，購物網頁考慮鄰近的 6 位使用者，餐廳、旅遊、教育及遊戲網頁考慮鄰近的 3 位使用者，娛樂網頁考慮鄰近的 2 位使用者。

而用來計算相鄰使用者距離的方法有很多，透過改變式 5.4 中的參數 p 為 1 表示使用曼哈頓距離（Manhattan distance）， p 為 2 表示以歐式距離（Euclidean distance）作為距離計算公式。我們在餐廳、購物、娛樂及遊戲網頁選擇 p 為 1，旅遊及教育網頁以 p 為 2 進行距離計算的預測效果較好。

$$\sqrt[p]{\sum_{i=1}^n (|x_i - y_i|)^p} \quad (5.4)$$

在以 Logistic regression 及 SVM 作為分類模型時，我們透過參數 C 調整 regularization term 之權重。參數 C 對餐廳類($C_{lr} = 7.6$ 、 $C_{svm} = 2.9$)、購物類($C_{lr} = 7.6$ 、 $C_{svm} = 1.0$)、旅遊類($C_{lr} = 7.0$ 、 $C_{svm} = 1.0$)、教育類($C_{lr} = 10.0$ 、 $C_{svm} = 0.05$)、娛樂類($C_{lr} = 9.8$ 、 $C_{svm} = 4.8$)及遊戲類($C_{lr} = 1.4$ 、 $C_{svm} = 0.1$)。另外，以 SVM 作為分類模型時，我們以 linear 做為 kernel 得到了較好的模型學習效果。

透過表 5.1 呈現四種模型對不同網頁類型之 F_1 score，在四種分類模型中，k-NN 在旅遊類的預測效果較佳，我們提出的 Matrix Factorization based Multi-objective Method 模型之 F_1 score 在餐廳、購物、教育、娛樂及遊戲類預測表現較好，表示使用者瀏覽不同類型網頁的行為之間確實存在著某種不容易觀察卻相互影響的隱性因子，透過模型能有效學習到這之間存在的關聯性，並幫助模型的預測效果更好。另外，我們於附錄補充以 AUC 做為評估指標之參數選擇並呈現預測結果。

表 5.1: 真實資料集在四種模型預測不同網頁類型之 F_1 score

分類器	餐廳類	購物類	旅行類	教育類	娛樂類	遊戲類
k-NN	0.537	0.574	0.615	0.484	0.44	0.492
Logistic regression	0.501	0.585	0.489	0.437	0.402	0.443
SVM	0.41	0.576	0.391	0.385	0.399	0.409
MF based Multi-objective Method	0.561	0.584	0.57	0.515	0.479	0.531

其中當購物網頁利用 Logistic regression 與我們提出的方法作為預測模型之預測效果較為接近，而餐廳網頁在此兩種模型之預測效果差異較明顯，可能的原因在於使用者對購物類型網頁之瀏覽量佔總瀏覽量 7%，而餐廳網頁之瀏覽量佔總瀏覽量僅 0.3%，相對提供的資訊量較為不充足，因此產生差異，在下一節中，我們將詳細討論使用者資料量對我們提出的方法之影響。

5.4 資料量對 MF based Multi-objective Method 之影響分析

Matrix Factorization based Multi-objective Method（如圖 5.2）透過同時學習隱藏在預測目標背後之隱性因子（矩陣 P 及矩陣 Q ），取代一般的線性方法，學習使用者特徵對預測目標之重要性（矩陣 W ），如圖 5.3。在本節我們利用自己建立的資料集進行實驗，討論資料量對我們提出的方法與一般的線性模型對預測結果之影響。

在自己建立的資料集共生成 1000 位使用者，我們每次取前 u 位使用者為測試資料，並且每次測試相同取最後 100 位使用者為預測資料，以表示訓練使用者數量對測試結果影響之公平測試。其中我們提出的方法能夠同時預測多種目標，而學習使用者特徵重要性每次僅預測一種目標，因此我們將所有預測目標預測結

果取平均作為預測結果進行比較。

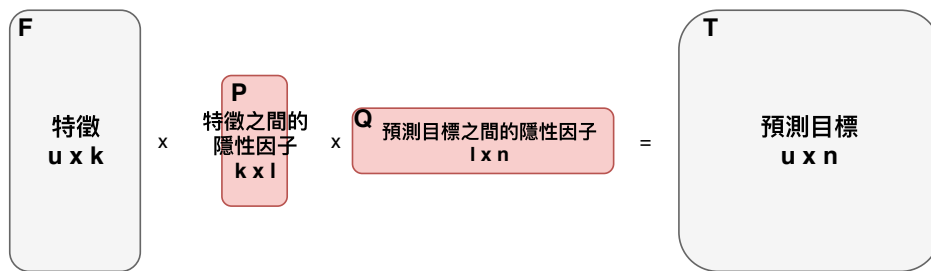


圖 5.2: MF based Multi-objective Method



圖 5.3: general linear model

其中評估使用者資料量對模型影響以均方根誤差（root-mean-square error）做為評估指標，均方根誤差屬於測量數值之間差異的量度，其定義方式如式 5.4，其中 n 為預測樣本數， y 為預測樣本真實結果， \hat{y} 預測樣本預測結果。

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n}} \quad (5.5)$$

一般的線性模型，以 gradient descent 將損失函式對矩陣 W 偏微分，透過反覆迭代學習矩陣 W 的參數，讓使用者特徵與矩陣 W 相乘的結果與真實目標之誤差越小越好。

透過圖 5.4 表現訓練資料量對 Matrix Factorization based Multi-objective Method 與一般的線性方法對預測結果之影響，當訓練資料量增加時，預測準確度能夠有效提升。當訓練資料量不足時，兩種方法在預測結果出現明顯差異，原因在於

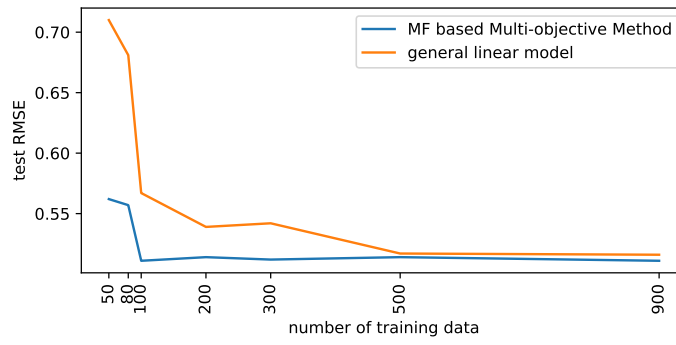


圖 5.4: 訓練資料量對預測結果之影響

同時學習矩陣 P 與矩陣 Q 之參數，且當其中 l 遠小於 k 及 n ，使學習之參數個數為 $l \cdot (k + n)$ ，其參數數量遠小於學習矩陣 W 之參數個數 $k \cdot n$ ，因此當擁有充足訓練資料量，使矩陣 P 、矩陣 Q 及矩陣 W 進行有效訓練，兩種模型之預測效果相近，但我們提出的模型在測試資料量較少時，預測效果可近似於一般線性模型的預測準確度。

六、 結論與未來展望

6.1 結論

大多數電子商務公司表示在特殊假期的銷售額增加，但本篇論文使用之資料集中所觀察到的現象表示銷售額提升可能是來自少數人的行為，因此我們能透過使用者的個人資訊及瀏覽網頁行為正確地分辨出潛在的購買者，並提供不同的行銷策略給不同類型的使用者，便能夠達到更好的廣告效益。

本篇論文之實驗對真實資料集預測使用者在特殊節日來臨前是否會增加瀏覽生活所需要之網頁，透過監督式學習分別對使用者瀏覽不同網頁的行為進行預測，以及 Matrix Factorization based Multi-objective Method 同時對六種網頁瀏覽行為進行多目標預測，並學習不同預測目標之間可能存在的隱性因子。

在真實資料集中，將我們提出的方法與基本的監督式學習之結果進行比較，實驗結果中我們提出的 Matrix Factorization based Multi-objective Method 在餐廳、購物、教育、娛樂、遊戲類別中更是優於其他監督式學習，此結果顯示使用者在瀏覽各種類型網頁的行為之間確實存在著不容易直接觀察卻相互影響的關聯性，因此透過我們的模型能夠有效學習到此的隱性因子，同時使預測效果更加精確，並且能夠準確預測使用者在節日前訪問六種類型網頁的行為變化。

我們透過建立擬真的使用者資料集來比較我們提出的方法與學習使用者特徵對預測目標之重要性在訓練使用者數量不同之狀況，在實驗結果中明顯呈現出當使用者人數受限時，我們提出的方法透過訓練較少的使用者數量，便能夠達到一

般線性模型的預測效果。

Matrix Factorization based Multi-objective Method 主要設計是以 Matrix Factorization 為基礎，適用於同時進行多目標預測的任務，並找出在預測目標之間可能存在著某種隱性因子，將此關係帶入預測目標中，使此模型之預測效果能優於預測單一目標之方法，並且此方法在訓練樣本數受限時，能夠達到相對較佳之預測準確度。

6.2 未來展望

在本篇論文中使用的真實資料集，實際能夠進行實驗的使用者人數過少，若使用過於複雜的模型，容易在訓練時造成 overfitting 的情況發生，並且當模型對其中一、兩位使用者的預測結果預測錯誤，將導致預測準確度大幅度下滑。若未來能夠收集更多使用者詳細個人資訊以及瀏覽網頁的歷史紀錄，將能夠讓模型之預測效果更加提升，並且針對不同行為變化的使用者進行更詳細的行為研究與討論，例如：對於瀏覽餐廳與購物網頁比率同時增加的使用者，在節日後是否可能增加對社群媒體的瀏覽率。

而使用者的購買意願會隨著時間慢慢累積，若使用者數量充足，便能夠結合深度學習之長短期記憶（Long Short-Term Memory），藉由使用者瀏覽網頁歷史紀錄擁有時序性之特性，將使用者平時瀏覽習慣區分為多個不同時期之瀏覽行為表現，對特殊節日前之行為變化進行研究，將可能取得更好的研究成果。

我們發現使用者瀏覽不同網頁的行為之間存在著某種潛藏的關聯性，使同時預測不同目標的預測效果比分別單獨預測更好，所以未來應該嘗試套用不同多目標學習的辦法 [8]，將可能使預測效果更加提升。

另一方面，我們的研究能有效預測使用者在特殊節日來臨前可能瀏覽網頁的行為，但是無法確認使用者更頻繁訪問某種類型網頁的行為，是否造成更多的購買行為產生，因此若能夠與電子商務之廠商合作，驗證使用者在瀏覽網頁之後續行為，以便彌補此遺憾，並且使整體研究更具推薦之價值。

參考文獻

- [1] “E-Commerce Continues To Be The Bright Spot For Holiday Sales,”
<https://www.forbes.com/sites/shoptalk/2016/12/27/ecommerce-continues-to-be-the-bright-spot-for-holiday-sales/28b1f6bb2780>, accessed: 2017-06-15.
- [2] “Alibaba’s Singles’ Day: What We Know About The World’s Biggest Shopping Event,” <https://www.forbes.com/sites/franklavin/2016/11/06/alibabas-singles-day-what-we-know-about-the-worlds-biggest-shopping-event/687e1e636da7>, accessed: 2017-06-15.
- [3] C. Lo, D. Frankowski, and J. Leskovec, “Understanding behaviors that lead to purchasing: A case study of pinterest,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016, pp. 531–540.
- [4] D. Kempe, J. Kleinberg, and É. Tardos, “Maximizing the spread of influence through a social network,” in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2003, pp. 137–146.
- [5] J. Leskovec, A. Singh, and J. Kleinberg, “Patterns of influence in a recommendation network,” in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 2006, pp. 380–389.

- [6] Y. Koren, R. Bell, and C. Volinsky, “Matrix factorization techniques for recommender systems,” *Computer*, vol. 42, no. 8, 2009.
- [7] T. K. Landauer, P. W. Foltz, and D. Laham, “An introduction to latent semantic analysis,” *Discourse processes*, vol. 25, no. 2-3, pp. 259–284, 1998.
- [8] S. Ruder, “An overview of multi-task learning in deep neural networks,” *arXiv preprint arXiv:1706.05098*, 2017.

附錄一

以下將呈現已以 AUC 當作評估指標之參數選擇：

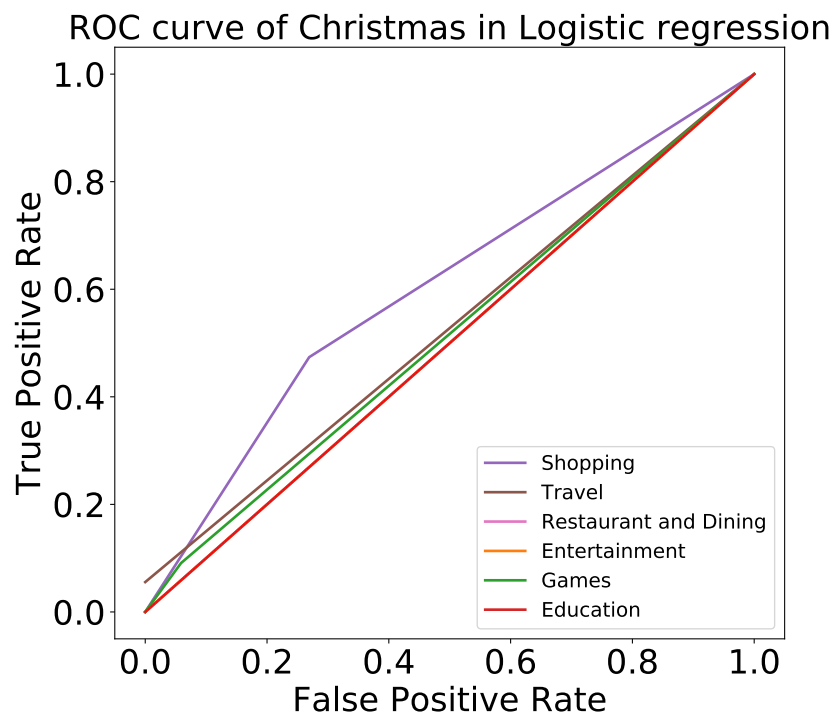
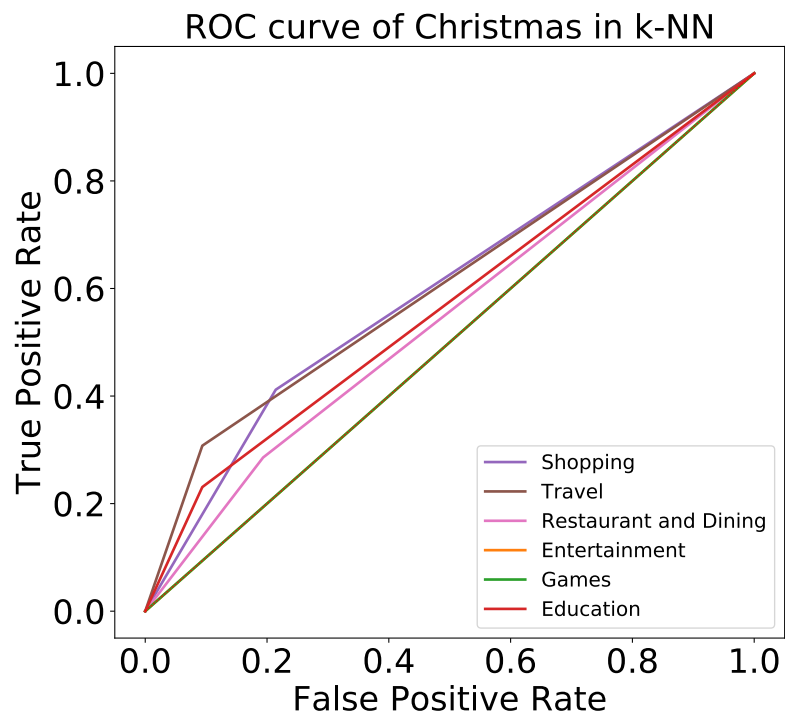
以 k-NN 當作分類模型時，餐廳及旅遊網頁考慮鄰近的 3 位使用者，購物及教育網頁考慮鄰近的 6 位使用者，娛樂網頁考慮鄰近的 24 位使用者，遊戲網頁考慮鄰近的 15 位使用者。在餐廳、購物、教育及娛樂網頁選擇 p 為 1，旅遊及遊戲網頁以 P 為 2 進行距離計算的預測效果較好。

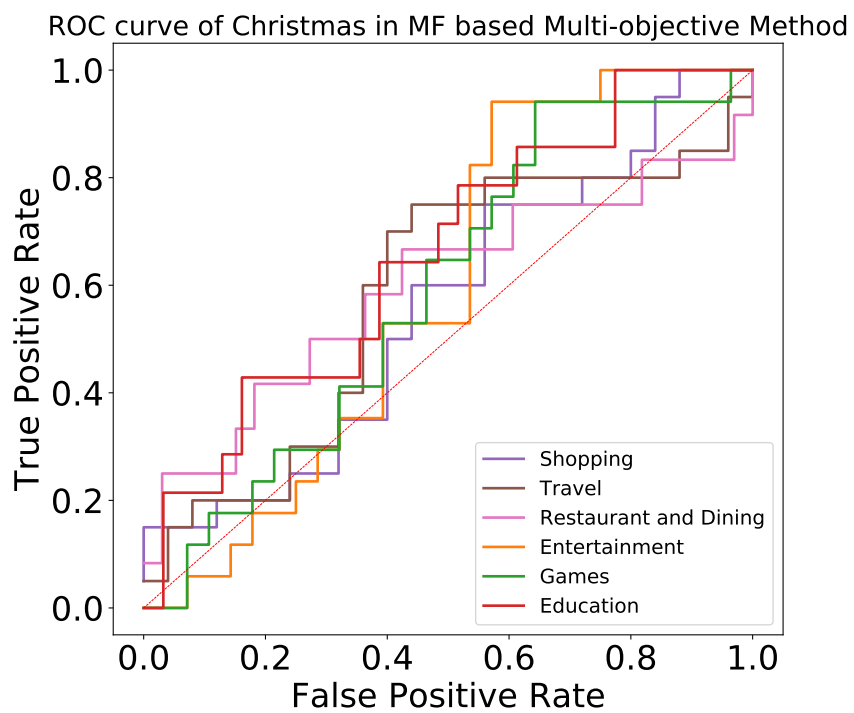
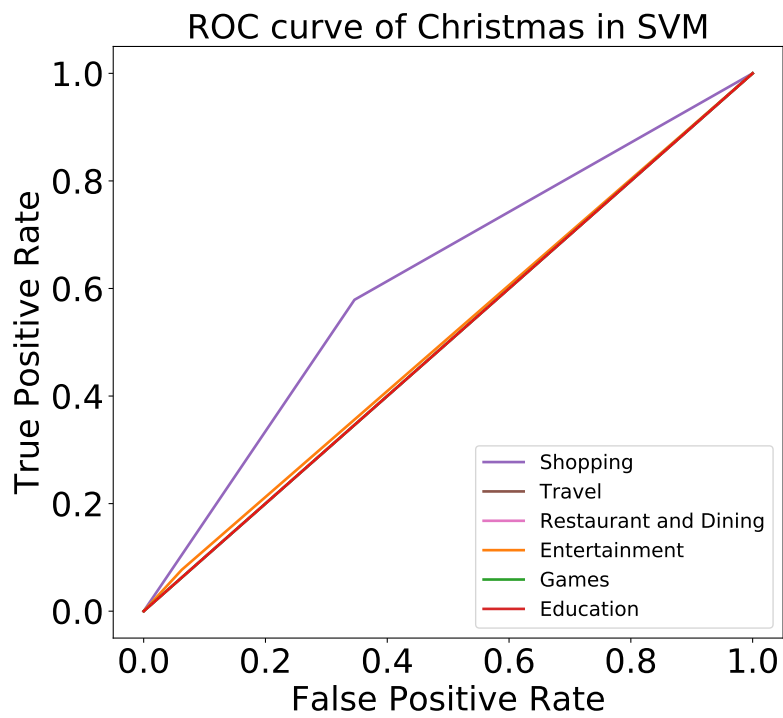
以 Logistic regression 及 SVM 當作分類模型時，參數 C 對餐廳類 ($C_{lr} = 7.1$ 、 $C_{svm} = 0.1$)、購物類 ($C_{lr} = 7.6$ 、 $C_{svm} = 0.1$)、旅遊類 ($C_{lr} = 7.0$ 、 $C_{svm} = 9.6$)、教育類 ($C_{lr} = 0.1$ 、 $C_{svm} = 0.1$)、娛樂類以 ($C_{lr} = 0.2$ 、 $C_{svm} = 1.7$) 及遊戲類 ($C_{lr} = 1.4$ 、 $C_{svm} = 0.1$)。另外，以 SVM 當作分類模型時，我們以 linear 當作 kernel 得到了較好的模型學習效果。

我們提出的方法當作分類模型時，學習率為 0.001，學習迭代次數為 150 次。

以下將呈現以 AUC 當作評估指標之預測結果以及 ROC 曲線圖：

分類器	餐廳類	購物類	旅行類	教育類	娛樂類	遊戲類
k-NN	0.542	0.584	0.62	0.512	0.499	0.504
Logistic regression	0.538	0.586	0.528	0.506	0.5	0.505
SVM	0.5	0.595	0.5	0.5	0.501	0.5
MF based Multi-objective Method	0.575	0.626	0.584	0.518	0.532	0.542





實驗資料集：<https://goo.gl/LmH46H>

實驗程式碼：<https://github.com/ivy2350442/MF-based-Multi-objective-Method>

實驗步驟：

1. 利用 `supervise_knn.py`、`supervise_log.py`、`supervise_svm.py` 以三種督式學習為模型，預測使用者於特殊節日前之瀏覽行為變化。
2. 利用 `mfmo_f1_score.py` 以 Matrix Factorization based Multi-objective Method 為模型，預測使用者於特殊節日前之瀏覽行為變化。
3. 利用 `generative_data.py` 生成擬真資料集。
4. 利用 `mfmo_rmse.py`、`general_linear_model.py` 比較不同使用者資料量對 Matrix Factorization based Multi-objective Method 及一般線性模型之 RMSE。

實驗詳細步驟將在實驗程式碼之網址上敘述。

