

Best deal, really? Review influence in deal-sharing community

朱瑤章[†]

Information Management
National Taiwan University
R07725032@mail.ntu.edu.tw

陳俊達[†]

Data Science degree program
National Taiwan University
R08946014@mail.ntu.edu.tw

施星宇[†]

Information Management
National Taiwan University
B04605052@mail.ntu.edu.tw

林挺生[†]

Information Management
National Taiwan University
D08725001@mail.ntu.edu.tw

Background and Motivation

Introduction of deal-sharing website

隨著電子商務網站在2000年開始逐步流行,越來越多的線上購物管道提供了消費者多元化的選擇。在選擇眾多情況之下,自然催生了不少類型的導購網站。所謂的導購網站,就是作為電商與消費者之間的仲介網站,提供消費者有興趣的購買資訊,以更有效率的商品推薦模式滿足消費者購買需求。通常推薦模式包含推薦高性價比商品、熱門商品(高評論度、高閱覽率、高商品點選率)、高評論商品、高相關連結度商品(買了A就會同時買B)、活動商品(電商活動如雙十一、週年慶)。

這其中,包含一類所謂Deal Sharing網站。Deal Sharing是一種結合商品導購以及社群討論的網站,商品資訊提供者為消費者或商家或協力廠商,每篇商品開放給消費者評論或點選是否值得購買。若是獲得多數消費者肯定並點選「值得」,或者引起高度討論,就會以熱門度指標逐步被推升至網站首頁的熱門推薦商品區。

因為商品推薦並不一定來自廠商自主推薦,或者皆由廠商付費廣告促銷,因此對消費者而言,避免了商家老王賣瓜自賣自誇,對商品品質及價格信任度大幅提升,推薦商品的購買意願也增強許多。

Pros and Cons of deal-sharing

這類導購網站當中, Slickdeals.net(以下簡稱SD)是美國最大最活躍的Deal Sharing促銷網站,也是Amazon.com目前最大的流量引導網站。其成立於1999年,宗旨是用戶分享、評論折扣,兼備購物引導與社群討論功能。當初成立初衷是挖掘BUG Price(標錯價),進而衍申為搜尋高性價比商品及優惠商品匯集搜尋功能。網站特色是利用社群討論功能強化對商品/價格的正反意見表達,並以投票評論決定每篇商品促銷資訊的好壞,若是好的促銷資訊會自動推上首頁熱門排行榜(Forum □ Trending deals □ Popular deals □ Frontpage deals)。好的促銷資訊由多方面決定:比原來價格降低了多少,產品的評論,促銷是否有地域、數量限制,在什麼樣的電商平臺。SD能獲得消費者的信任,來自使用者為主的推薦體系,以及嚴格的監管機制,一旦發現人為操弄商品評論或惡意欺騙,就會立即下架該商品資訊。

Motivation of research

隨著互聯網技術的發展, e-WOM(electronic word-of-mouth), 如線上評論和微博, 變得越來越流行。許多電子商務網站, 如亞馬遜和淘寶, 已經建立了線上評論系統, 以鼓勵消費者發佈產品評論, 並因此逐漸改變了消費者的行為模式, 影響了消費者的購買決策。例如, 消費者在決定看哪部電影、應該投資哪支股票等時, 越來越關注線上意見。(Wysocki, 2000; Ryu & Han, 2010)

過去商品評論研究只能間接評估網路商品評論對網路消費者購物意圖之影響(Daniel S.K., Jochen R., Martin N., Daniel K., 2016), 有些透過問卷調查法評估購買意圖受到商品評論影響, 有些透過評估商品評論預測交易數量(Zhi-Ping Fan, Yu-Jie Che, Zhen-Yu Chen, 2017), 但普遍的問題是, 很難評估潛在買家如何受到評論之正負面影響。因一般電商網站無法直接判定使用者評論對購物意圖之效應(例如消費者看完負面評論而離開, 並未留下任何證據得知這類行為)。

此一問題, 在SD這類導購網站可以從消費者點選"Deal score"而評估潛在買家受商品評論的影響程度。喜歡該商品的消費者可以點選正面的"Thumb up", 不喜歡該商品或價格的消費者也可以點選"Thumb down"來表達他的不滿意意見。而SD網站的開放性, 允許所有消費者都可以進行評論與評論deal score, 所以也避免了電商網站只有購買者得以評論的限定意見發表缺點。如此一來, 消費者夠過這類協力廠商購物分享網站, 可獲得更客觀公正的評價, 以協助購物決策。

以研究的角度而言, 豐富的資訊讓我們可以透過SD網站的deal score, 與許多相關的變數, 觀察有哪些可能影響deal score分數的變動。

其中, 我們認為有幾個可能造成影響的變數:

1. 使用者的評論comment篇數
2. 使用者的正面評論comment篇數
3. 使用者的負面評論comment篇數
4. 商品的正面評論篇數之總影響(亦即多篇的正面評論造成的綜效)
5. 商品的負面評論篇數之總影響(亦即多篇的負面評論造成的綜效)
6. 商品文章瀏覽次數

我們關注幾個議題:

1. 評論對deal score是否有影響? 又影響程度為何?

2. 正面評論與負面評論對deal score造成的影響是否有差異？
3. 是否可以依據初期評論及瀏覽情況預估deal score的分數？

Literature Review

Online product review effects

e-WOM指的是“潛在的、實際的或之前的消費者對一個產品或公司所做的任何正面或負面的陳述，這些陳述可以通過互聯網提供給許多人和機構” (Hennig-Thurau, Gwinner, Walsh, & Gremler, 2004, p. 39)。消費者評論是e-WOM最容易獲得和最普遍的形式(Chatterjee, 2001)。Online customer reviews(以下簡稱OCRs)是一個評論者和許多讀者之間的單向非同步通信(一對多通信) (Litvin et al., 2008)。OCRs可定義為任何正面、負面或中性的評論、評級、產品、服務、品牌的排名，或由前客戶提出的，並以非結構化格式(如Twitter)或更結構化的格式(如發佈在獨立消費者評論網站如TripAdvisor或Slickdeals上的消費者評論)，協力廠商電子商務網站(如Booking.com)，或一般電子商務網站(如Amazon.com)。旅遊學者非常重視與旅遊相關的評論，例如顯示它們對酒店銷售的影響(Ye et al., 2009)，它們如何影響消費者對酒店的認識和態度(Vermeulen & Seegers, 2009)，以及消費者的購買意圖(Cantalops & Salvi, 2014; Cox et al., 2009; Yoo & Gretzel, 2009)。

網上信任被認為是消費者使用和購買電子商務網站的主要影響因素(e.g. Bart, Shankar, Sultan, & Urban, 2005)。這項研究側重於對OCRs系統的信任，OCRs系統是一種e-WOM的形式，在文獻中也被稱為使用者生成的內容。下面我們回顧一下e-WOM文獻的來源可信度和資訊可信度。

來源可信度是指消費者認為溝通來源可靠、真誠和誠實，而來源可信度來自於消費者對來源在特定領域擁有的知識、技能或專長的認知(Ohanian, 1990)。當消費者缺乏對產品或服務的瞭解時，他們通常依賴專家或公正的來源。已經發現，在offline WOM，來源可信度和可信賴性會強烈影響客戶對來源消息的意見改變和接受(Brown & Reingen, 1987; McGinnies & Ward, 1980)。

在線上環境中，使用者可以基於電子資訊的連結內容創建對他人的完整印象(Brown, Broderick, & Lee, 2007)，這意味著購物者分析之前消費者評論的內容來評估可信度。目前對於使用者張貼內容對消費者行為的影響最有爭議的問題之一是消費者對發佈在使用者生成的內容網站上的資訊的信任程度(Cox et al., 2009)，然而，很少有研究探討這個問題。例如，Cheung et al. (2009)研究了e-WOM在中國感知可信性的決定因素，發現來源可信性、先前信念的確認、推薦一致性、推薦評級和論證強度影響感知的e-WOM評論可信性。Qiu, Pang, and Kim (2012)在學生中進行了實驗，發現通過評論歸因的中介作用，相互衝突的綜合評分降低了正面評論的評論可信度和診斷性，而負面評論則不然。Kusumasondjaja, Shanka, and Marchegiani (2012)對639名旅遊消費者進行了實驗，發現當評論者(reviewer)的身份被揭露時，負面評論比正面評論審查更可信。然而，當評論者的身份未被揭露時，正面評論和負面

評論在感知的可信度或對消費者信任的影響方面沒有顯著差異。

消費者信任負面評論多於正面評論，認為極端評論更有可能是不可信的評論(Raffaele, 2016)。事實上，消費者認為負面評論比正面評論更值得信賴，這與之前在離線和線上環境下的研究結果是一致的(Kusumasondjaja et al., 2012; Papathanassis & Knolle, 2011)。線上評論的來源可信度是通過評論者提交的評論數量來評估的，但它的評估也取決於評論資訊的極端性、效價(Valence)和內容。我們發現，來自一次性貼文(全新貼文者的單一篇張貼文)的極端評論更有可能被旅遊消費者視為不可信(Raffaele, 2016)。

許多研究表明e-WOM影響消費者行為和產品銷售(Liu, 2006; Godes & Mayzlin, 2004; Chevalier & Mayzlin, 2006)。他們發現線上評論的屬性，例如線上評論的數量(Duan, Gu, & Whinston, 2008; Ye, Law, & Gu, 2009; Liu, 2006)，評級(Chevalier & Mayzlin, 2006; Segal et al., 2012)和評論中分享的觀點(Ye et al., 2009)對產品銷售有影響。幾位研究人員探索了線上評論和產品銷售之間的關係(Chevalier & Mayzlin, 2006; Dellarocas, Awad, & Zhang, 2007)。

Sentiment analysis

在過去的十年裡，情感分析技術已經被用來衡量透過線上評論內容傳達的情感(Pang & Lee, 2005; Prabowo & Thelwall, 2009)。正如Yu, Liu, Huang, and An (2012)所指出的，通過情感分析技術從線上評論的內容中提取的情感指數可以用於預測許多社會經濟現象，包括產品市場份額、票房出席率、資訊或疾病的傳播(Culotta, 2010)和政治選舉的結果(Lee, 2009)。情緒指數也可以用來分析宏觀經濟。

使用機器學習技術對語意分析的發展，始發展於2002年，嘗試根據整體情感來對文檔進行分類以取代過去人工分類的作法(Pang, Lee & Vaithyanathan, 2002)。隨後提出情緒分析(Sentiment analysis)這一詞在NLP領域中(Nasukawa & Yi, 2003)，並提出意見挖掘及語意分類方法(Dave, Lawrence & Pennock, 2003)。後續研究嘗試改變早先的基礎兩極意見研究，將電影影評分類並預測為3至4星的多重級別(Pang, B, Lee, 2005)，開始出現了緒強弱不同程度的語意分析作法。為避免用關鍵單詞判斷情緒的單一作法，開始思考如何參考前後文語意，例如利用否定詞語和轉折詞語，提升了基於詞典的情感分類方法(Ding X., Liu B. & Yu P., 2008)，以及利用twitter為語料庫，並使用n-gram進行了情緒分析和意見挖掘(Pak A. & Paroubek P., 2010)。在2011年，提出語義定向計算器(SO-CAL, Semantic Orientation Calculator)，利用詞語的情感強度以及情感加強和否定規則判斷篇章的情感極性與強度(polarity and strength)(Taboada M., Brooke J., Tofigolski M., Voll K., Stede M., 2011)。隨著神經網路興起，提出多個基於樹結構的情感樹庫上語義組成的遞迴深層模型(Socher R., Perelygin A., Wu J., 2013)，以及描述了一個卷積體系結構，稱為動態磁碟區積神經網路(DCNN)，他們採用它來進行句子的語義建模(Kalchbrenner N., Grefenstette E., Blunsom P., 2014)。在序列化的LSTM (Long Short-Term Memory)模型的基礎上加入了句法結構的因素(Tai K. S., Socher R., Manning C. D., 2015)，利用語言資源和神經網路相結合來提

升情感分類問題的精度(Qian Q., Huang M., Lei J., Zhu X.,2016)。

研究步驟

為研究本次主題:評論對deal score的影響。我們設計以下研究步驟:

1. 訓練正負面評論情緒分類器
 - 尋找合適Training set
 - Classifier Method1: BOW + MLP
 - Classifier Method2: BERT
2. 爬蟲抓取Slickdeals.net網站近兩個月的所有商品文章過濾無評價的商品, 抓取近期約5000筆商品貼文及其所有評論。然後使用上述第一步的classifier, 對所有評論進行正負面情緒分析並給予一個情緒分數(正面評價分數為0.5~1, 值越大代表情緒越正面; 負面評價分數為0~0.5, 值越小代表情緒越負面)
3. 對Deal score與各comment相關變數進行回歸分析

以下進行詳細說明。首先在訓練資料集的選擇, 我們選擇了¹Multi-Domain Sentiment Dataset (2009)

1. 此資料集搜集了Amazon.com 上共17個domain的產品評論, 各種domain間的review數是不平衡的, 有些domain像是書籍、DVD等, 包含review數較多, 可達到十萬的量級; 有些domain像是樂器, 可能就只有數百個review。
2. 每個review會label為1 or 0, 1 代表此review的情緒為正面, 0代表此review的情緒為負面。
3. 每個domain針對label 1和 0 最多各選擇1000筆, 故每個domain最多有2000筆, 而訓練資料總共有28431筆。
4. 訓練資料的label比例如下所示(Figure 1.)

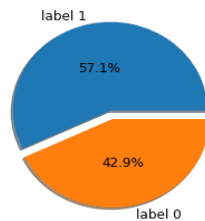


Figure 1. 訓練資料label比例 (Multi-Domain Sentiment Dataset 2009)

再來是以BOW + MLP訓練Classifier method 1, 步驟及結果如下:

Preprocess Method

先對評論內容進行前處理。我們刪除省略符號並還原, 例如: let's → let us。將網路常用縮寫還原, 例如: ppl → people。刪除非字母, 並斷詞使用 spaCy。處理後的字典共62062字。

Model: BOW + MLP

1. 使用 Bag of Word 向量代表每一個review, 向量中的每一個元素代表相對應位置的字在該review中的詞頻。
2. 使用 Multi-layer Perceptron 作為最後的分類器, 其中維度變化為[62062, 1024, 256, 64, 1]
3. 最後將MLP的結果再通過激勵函數 sigmoid, 跟 threshold比較即為最後的output

Parameter choose

1. 優化器使用 Adam, learning rate設為1e-4 且為避免太早過擬合將 weight decay設為1e-5
2. 將minibatch設為1024
3. 最多跑20個 epoch, 不過差不多第4個epoch後就開始有過擬合的情況發生
4. 在MLP 中的 dropout rate 皆設為 0.3
5. 使用 binary cross entropy 計算loss, 並因應label 不平衡的狀況將threshold 設為 0.55

Classifier validation result

1. Threshold 0.5 → Validation result: f1 0.866, precision 0.838, recall 0.892
2. Threshold 0.55 → Validation result: f1 0.869, precision 0.846, recall 0.892
3. Threshold 0.60 → Validation result: f1 0.864, precision 0.851, recall 0.877

¹ <http://www.cs.jhu.edu/~mdredze/datasets/sentiment/>

Training process visualization

以下為best validation F1 0.869, 即threshold設為0.55 之訓練過程(Figure 2. and Figure 3.)

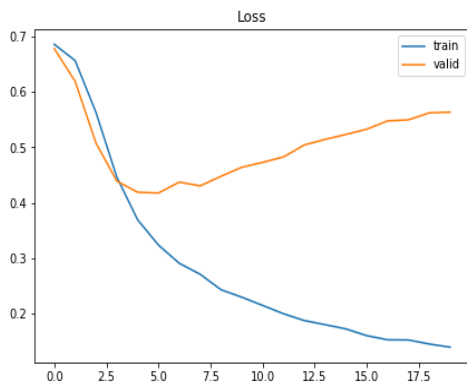


Figure 2 : Loss curve

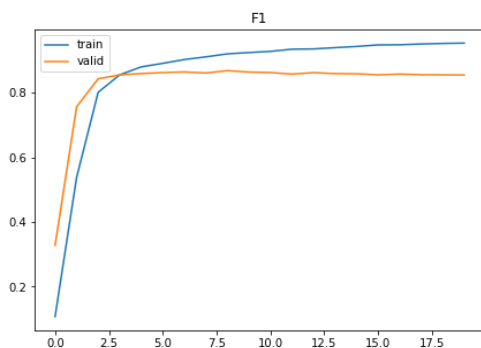


Figure 3. F1 Curve

接著是第二個分類器訓練方法, 以BERT訓練Classifier 2. 首先介紹BERT。BERT 是一個在大規模文本上預訓練的語言模型, 目標是讓各種NLP任務不要浪費時間和模型參數來學習艱難但卻共同需要的語言結構, 改為透過使用BERT預訓練的向量, 再加入簡單的fine tuning結構後, 就可節省理解力氣, 直接使模型架構專注在解決特定NLP任務, 也避免過擬合。

BERT的訓練方式為非監督式, 包含:

1. 輸入一句話, 隨機移除幾個詞, 讓BERT預測移除的詞。
2. 輸入兩句話, 可能是文章中連續的兩句話或者不相干的兩句, 讓BERT預測是否兩句是連續或不相干的。藉由預測上下文關係, BERT 得以學到豐富的語言模型。

BERT模型由Transformer組成(Figure 4.), Transformer主要是由Multi-head attention取代傳統seq2seq RNN的attention, 原因是seq2seq一次只能輸入一個time step, 訓練慢, 且只有捕捉來源句和目標句的關係, 沒捕捉來源句內部的關係與目標句內部的關係。因此Transformer分別針對來源句內部、目標句內部、來源與目標間分別做了attention。而Multi-head attention相較傳統attention則是讓當前詞(query)與要注意的詞(key和value)分別做了不同線性轉換

再過attention並組合, 亦即希望建立多種相異的注意力模式, 而非只有單一種。另外, transformer也加入位置編碼, 讓模型能瞭解詞的前後順序, BERT-BASE的模型就是採用多個Transformer模型當基本元件組合起來, 深化為12層的模型。

此次任務上, 我們採用 BERT 再接上兩層 fully-connected layer 作為情感分析的模型(Figure 5.), 採用 80% 的資料作為 train data, 用 20% 的資料作為 validation data, 預測一篇文的情感是正面或負面。另外, 我們採用 huggingface 的 BERT 預訓練模型來作為 fin-tuning 的基礎。

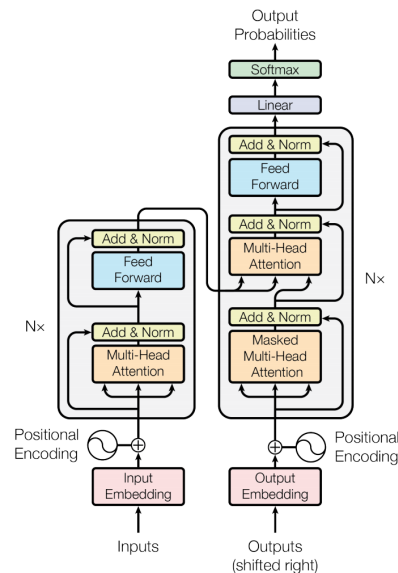


Figure 4. The transformer – model and architecture

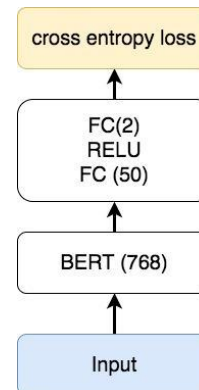


Figure 5. BERT process

Parameter choose

1. BERT 的 Learning Rate 從 3e-6 向下每層遞減, 乘以上一層的 0.95 倍。在BERT 之外的 Learning Rate 則是 1e-3。
2. 前 200 個 step 固定 BERT 的參數, 從 0 % 線性增加 Learning Rate 直到 2000 step(100 %), 再線性遞減到 1 %
3. Optimizer: Adam

- 4. epoch: 2
- 5. Classifier validation result
- 6. f1 score: 0.92, precision: 0.909, recall: 0.940

Method selection

比較方法BOW+MLP以及BERT, 因BERT validation分數較高(F1 score: 0.86對0.92), 因此決定用BERT 作為下一步研究的情感分析基礎, 針對每則貼文的回覆, 都計算其情感正負的分數。

Research Results

分析方法

為了達到能夠達到根據過往的商品評論中的評論情緒分析以預測該商品的deal score, 根據測試資料集選取了各個商品中時間先後前10的評論去進一步計算各項參數, 並訓練參數與deal score的線性迴歸模型, 了解deal score與各個參數的關係以得到較佳的預測模型。

參數說明

參數	說明
D	商品數目
C	$C = 10$, 前十篇評論的數量
$comment_score$	單一comment的情緒分數
$reputation$	單一comment發表人的聲望值
$dealScore$	資料集中的 $dealScore$, 作為預測的y值
$sigmoid_y$	$Sigmoid(dealScore)$
$sigmoid_y/comment$	$Sigmoid(\frac{dealScore}{C})$
num_reply	資料集中統計的reply數
$real_num_reply$	與參數 C 相同
$dealViews$	資料集中的瀏覽次數
$dealScore/dealView$	$\frac{dealScore}{dealViews}$
$dealScore/real_num_reply$	$\frac{dealScore}{real_num_reply}$
pos_num	$\sum_{i=1}^C k, \forall k = \{1 \text{ if } comment_score > 0.5 \text{ } 0 \text{ if } comment_score \leq 0.5\}$
neg_num	$\sum_{i=1}^C k, \forall k = \{1 \text{ if } comment_score \leq 0.5 \text{ } 0 \text{ if } comment_score > 0.5\}$
pos_sum	$\sum_{i=1}^C s, \forall s = \{comment_score_i \text{ if } comment_score > 0.5 \text{ } 0 \text{ if } comment_score \leq 0.5\}$
neg_sum	$\sum_{i=1}^C s, \forall s = \{comment_score_i \text{ if } comment_score \leq 0.5 \text{ } 0 \text{ if } comment_score > 0.5\}$
pos_num_rep	$\sum_{i=1}^C k, \forall k = \{1 * reputation \text{ if } comment_score > 0.5 \text{ } 0 \text{ if } comment_score \leq 0.5\}$
pos_sum_rep	$\sum_{i=1}^C s, \forall s = \{comment_score_i * reputation \text{ if } comment_score > 0.5 \text{ } 0 \text{ if } comment_score \leq 0.5\}$
neg_num_rep	$\sum_{i=1}^C k, \forall k = \{1 * reputation \text{ if } comment_score \leq 0.5 \text{ } 0 \text{ if } comment_score > 0.5\}$
neg_sum_rep	$\sum_{i=1}^C s, \forall s = \{comment_score_i * reputation \text{ if } comment_score \leq 0.5 \text{ } 0 \text{ if } comment_score > 0.5\}$
$pos_sigmoid_sum$	$\sum_{i=1}^C Sigmoid(s), \forall s = \{comment_score_i \text{ if } comment_score > 0.5 \text{ } 0 \text{ if } comment_score \leq 0.5\}$
$neg_sigmoid_sum$	$\sum_{i=1}^C Sigmoid(s), \forall s = \{comment_score_i \text{ if } comment_score \leq 0.5 \text{ } 0 \text{ if } comment_score > 0.5\}$
$pos_sigmoid_sum_rep$	$\sum_{i=1}^C Sigmoid(s), \forall s = \{comment_score_i * reputation \text{ if } comment_score > 0.5 \text{ } 0 \text{ if } comment_score \leq 0.5\}$
$neg_sigmoid_sum_rep$	$\sum_{i=1}^C Sigmoid(s), \forall s = \{comment_score_i * reputation \text{ if } comment_score \leq 0.5 \text{ } 0 \text{ if } comment_score > 0.5\}$

$$score_sum = \sum_{i=1}^C comment_score_i$$
$$score_mean = \frac{score_sum}{C}$$
$$score_sigmoid_sum = \sum_{i=1}^C Sigmoid(comment_score_i)$$
$$score_sigmoid_sum_rep = \sum_{i=1}^C Sigmoid(comment_score_i * reputation)$$
$$score_sigmoid_mean = \frac{score_sigmoid_sum}{C}$$
$$difference = (pos_sum - neg_sum) * C$$
$$difference2 = (pos_num - neg_num) * C$$

參數處理

- 觀察dealScore與情緒分析的結果, dealScore為一整數而情緒分析結果為介於0~1間的數值：
 - $dealScore = \{x \mid \forall x \in N\}$
 - $comment_score = \{x \mid \forall 0 \leq x \leq 1\}$
 - 因此可以用sigmoid function將dealScore轉為一介於0~1間的值以方便進行回歸分析。
 - $Sigmoid(X) = \frac{1}{1+e^{-x}}$
- 因為發現在正面情緒和負面情緒中pos_sum和neg_sum有稍微明顯的差距, 又透過乘上comment數量來顯著此效果：
 - $difference = (pos_sum - neg_sum) * C$
- 由於dealScore變化很大, 也可以將dealScore除以dealView以及real_num_reply減少其變化程度得到： $dealScore/dealView$ 、 $dealScore/real_num_reply$
- 由於各個留言者的聲譽值不相同, 因此可根據不同留言者的聲譽值對於該則留言的情緒值做加權, 進而計算出： pos_num_rep 、 pos_sum_rep 、 neg_num_rep 、 neg_sum_rep 、 $pos_sigmoid_sum_rep$ 、 $neg_sigmoid_sum_rep$ 、 $score_sigmoid_sum_rep$

參數範例

參數	數值	參數	數值
dealScore	75	neg_num_rep	1827
sigmoid_y	1	neg_sum_rep	72.819015
sigmoid_y/comment	0.72343	pos_sigmoid_sum	53.148577
num_reply	39	neg_sigmoid_sum	-58.602633
real_num_reply	77	pos_sigmoid_sum_rep	1178.4044
		p	98
dealViews	20121	neg_sigmoid_sum_rep	-5846.9279
dealScore/dealView	0.003727	p	32
dealScore/real_num_reply	0.961538	score_sum	17.466622
pos_num	18	score_mean	0.447862
neg_num	21	score_sigmoid_sum	-5.454056
pos_sum	15.85469	score_sigmoid_sum_rep	-4668.5234
neg_sum	1.611925	rep	34
pos_num_rep	384	score_sigmoid_mean	-0.139848
pos_sum_rep	356.5951	difference	555.46811
	47	difference2	3
			-117

先挑選與 $dealScore$ 無相關之參數分別計算之間的VIF值已判定是否存在共線性：

其中可發現 pos_num 與 pos_sum 存在共線性因此剔除 pos_num ，又因為 pos_num 越多 pos_sum 必然越大為合理現象，因此保留 pos_num ， $pos_sigmoid_sum$ 與 pos_sum 存在共線性因此剔除 $pos_sigmoid_sum$ ， $score_sum$ 、 $score_mean$ 與 pos_sum 存在共線性因此剔除 $score_sum$ 、 $score_mean$ ， $difference$ 與 pos_sum 存在共線性因此剔除 $difference$ ， $score_sigmoid_sum$ 與 $score_sigmoid_mean$ 間存在共線性因此剔除 $score_sigmoid_mean$ ，最後 pos_num_rep 與 pos_sum_rep 存在共線性因此剔除 pos_num_rep 。

$$S_t = \beta_0 + \beta_1 * dealViews + \beta_2 * pos_num + \beta_3 * neg_num + \beta$$

將資料集中的20%作為測試資料集，以訓練資料集訓練回歸式
得以下結果：

$score_sigmoid_sum(\beta_{12})$	0.7789	0.388	2.009	0.045	0.018	1.54
$score_sigmoid_sum_rep(\beta_{13})$	-1.23E-06	4.88E-05	-0.025	0.98	-9.71E-05	9.46E-05
$difference2(\beta_{14})$	0.1926	0.332	0.58	0.562	-0.459	0.845

參數測試

1. 將 pos_num 、 neg_num 作為X,
 $dealScore/real_num_reply$ 為y做線性迴歸, 得以下
結果: $R^2 = 0.027$

	<i>constan</i>	<i>pos_nu</i>	<i>neg_nu</i>
	<i>t</i>	<i>m</i>	<i>m</i>
<i>coef</i>	15.0977	-1.2441	-1.3034

1. 將 pos_num 、 neg_num 作為X, $dealScore/dealView$ 為y做線性迴歸, 得以下結果: $R^2 = 0.023$

	<i>constan</i> <i>t</i>	<i>pos_nu</i> <i>m</i>	<i>neg_nu</i> <i>m</i>
<i>coef</i>	0.0054	-0.0002	-0.0003

2. 將 pos_num 、 neg_num 作為 X ， $dealScore$ 為 y 做線性迴歸，得以下結果： $R^2 = 0.002$

	constant	pos_num	neg_num
coef	-381.9378	40.9148	39.9758

3. 將 pos_sum 、 neg_sum 作為 X ， $dealScore$ 為 y 做線性迴歸，得以下結果： $R^2 = 0.004$

	constant	pos_sum	neg_sum
coef	38.1636	2.7742	15.4754

4. 將 $pos_sigmoid_sum$ 、 $neg_sigmoid_sum$ 作為 X ， $dealScore$ 為 y 做線性迴歸，得以下結果： $R^2 = 0.009$

	constant	pos_sigmoid_sum	neg_sigmoid_sum
coef	98.3399	-0.8718	1.3230

5. 將 $score_sum$ 作為 X ， $dealScore$ 為 y 做線性迴歸，得以下結果： $R^2 = 0.001$

	constant	score_sum
coef	54.8987	1.7290

6. 將 $score_mean$ 作為 X ， $sigmoid_y$ 為 y 做線性迴歸，得以下結果： $R^2 = 0.021$

	constant	score_mean
coef	0.6682	0.0966

7. 將 $difference$ 、 $difference2$ 作為 X ， $dealScore$ 為 y 做線性迴歸，得以下結果： $R^2 = 0.004$

	constant	difference	difference2
coef	108.8712	-1.0328	0.5693

8. 取前述較具有影響的參數： $dealViews$ 、 $score_sigmoid_sum$ 作為 X ， $dealScore$ 為 y 做線性迴歸，得以下結果： $R^2 = 0.728$ ，而以測試資料集得到之 $R^2 = 0.6981$ 、 $RMSE = 37.9384$

參數	數值
constant	2.1873
dealViews	0.0022
score_sigmoid_sum	0.4533

Ridge Regression

$dealViews$ 、 pos_num 、 neg_num 、 pos_sum 、 neg_sum 、 $neg_sigmoid_sum$ 、 pos_sum_rep 、 neg_num_rep 、 neg_sum_rep 、 $pos_sigmoid_sum_rep$ 、 $neg_sigmoid_sum_rep$ 、 $score_sigmoid_sum_rep$ 、 $score_sigmoid_sum$ 、 $difference2$ 參數作為 X 並以 $dealScore$ 作為 y ，透過 ridge regression 去差異化不同參數對於迴歸線的影響，其中在 $\alpha = 1$ 時得到如下結果：訓練 $R^2 = 0.7339$ ，測試 $R^2 = 0.6997$ ， $RMSE = 48.5761$ ，參數結果如下：

參數	$\alpha = 1$
dealViews	1.004021
pos_num	0.002131
neg_num	3.186374
pos_sum	3.168364
neg_sum	-6.858621
neg_sigmoid_sum	3.344892
pos_sum_rep	0.117768
neg_num_rep	0.000008
neg_sum_rep	-0.000002
pos_sigmoid_sum_rep	0.000229
neg_sigmoid_sum_rep	-0.000015
score_sigmoid_sum	0.000013
score_sigmoid_sum_rep	0.759445
difference2	-0.000001

又分別測試在不同 α 值下各個參數的變化，得以下圖表 (Figure 6.):

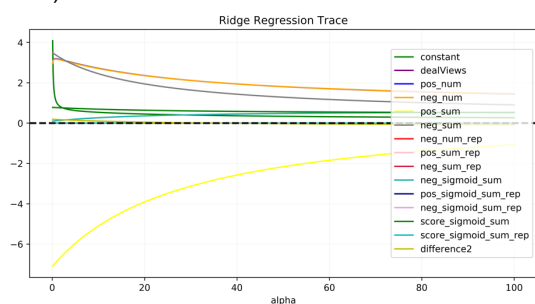


Figure 6. Ridge regression Trace

又根據Ridge Regression於 $\alpha = 40$ 時的各個參數值繪製影響力程度圖表(Figure 7.):

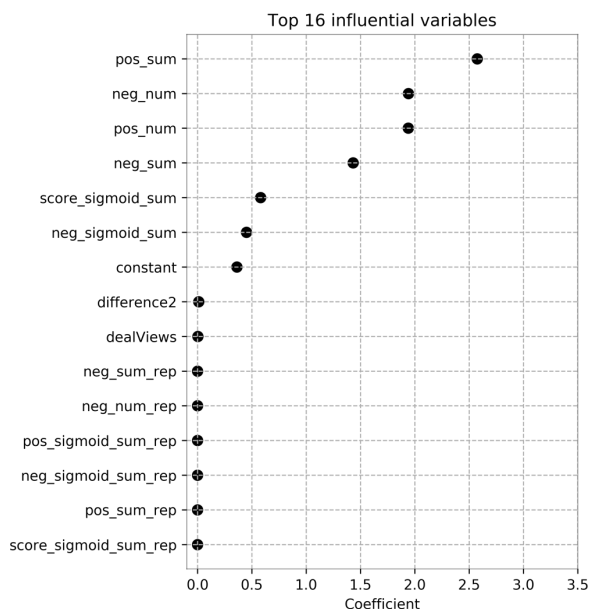


Figure 7. Top 16 Influential variables of Ridge regression($\alpha = 40$)

其中遞減越迅速的參數意味著在模型中較為不重要，相對的能夠保留到最後者則為影響力較高的參數，根據上圖結果，neg_num、pos_sum、pos_num這三個參數在模型中具有較高的影響力。

Lasso Regression

dealViews、pos_num、neg_num、pos_sum、neg_sum、neg_sigmoid_sum、pos_sum_rep、neg_num_rep、neg_sum_rep、pos_sigmoid_sum_rep、neg_sigmoid_sum_rep、score_sigmoid_sum_rep、score_sigmoid_sum、difference2參數作為X並以dealScore作為y，透過lasso regression去差異化不同參數對於迴歸線的影響，又分別測試在不同alpha值下，各個參數的變化，得以下圖表(Figure 8.):

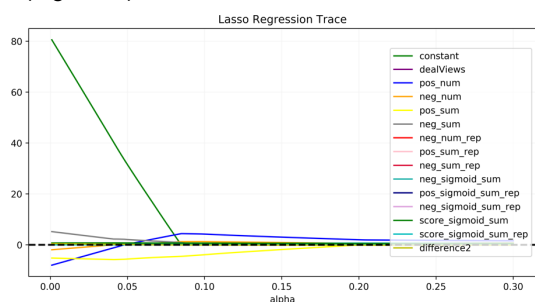


Figure 8. Lasso regression Trace

其中根據參數收斂至零的順序可以視為該參數對於模型的重要性，可根據收斂至零的順序製作以下圖表：

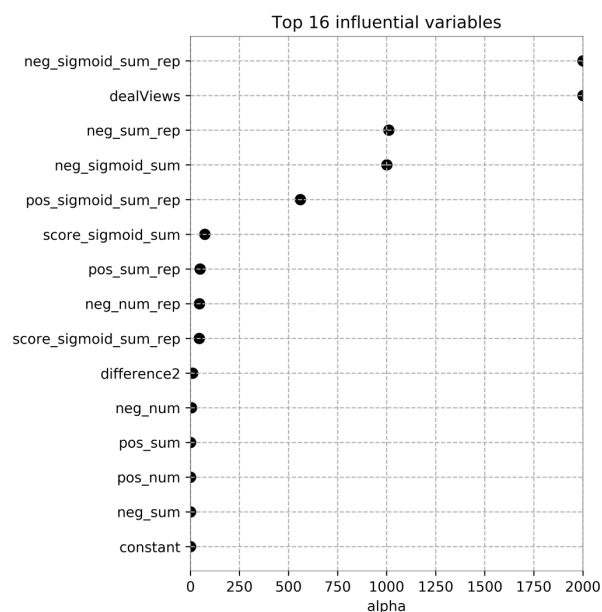


Figure 9. Top 16 Influential variables of Lasso regression

根據上圖可知較為有影響力的參數有dealViews、neg_sigmoid_sum_rep、neg_sum_rep、neg_sigmoid_sum

結論

對於建構一個模型透過分析商品評論的情緒程度以預測商品的deal score，採用簡單線性迴歸並透過各個參數的p-value去挑選影響力大的參數，所得到的回歸結果：對於訓練資料集的 $R^2 = 0.734$ ，對於測試資料集的 $R^2 = 0.6998$ 、 $RMSE = 48.5691$ ，而採用Ridge Regression於 $\alpha = 1$ 時得到如下結果：對於訓練資料集 $R^2 = 0.7339$ ，對於測試資料集 $R^2 = 0.6997$ ， $RMSE = 48.5761$ 。

又近一步分析各個參數的重要性於簡單線性迴歸p-value的分析中dealViews、score_sigmoid_sum具有較大的影響力，於Ridge Regression中neg_num、pos_sum、pos_num具有較大的影響力，於Lasso Regression中dealViews、neg_sigmoid_sum_rep、neg_sum_rep、neg_sigmoid_sum具有較大的影響力，總結上述結果dealViews對於模型影響力最大，其中dealScore的大小與商品的comment數量有相關性，comment數量越大dealScore的值就會越大，而dealViews與商品的comment數目有一定的關係因此可預期有相當的影響力，而從分析結果中可發現負面情緒的評論影響力高於正面情緒評論的影響，推論是因為在實際情況中dealScore為負值的商品較少，因此與負面情緒相關的參數越能夠決定該商品dealScore是否為負值。

未來研究

本次研究使用BERT訓練情感分析分類器，雖然可驗證正負面不同情緒評論對潛在消費者的購物影響，但經過進一步分析評論內容，發現對於價格討論的文章較不容易判斷出情緒。例如：「這價格比上個月便宜10元」、「去年至今只跌價了10%」、「運費還需另外計算」等等對於價格的評論，無法經由我們訓練的分類器判斷出正負面情緒，相對會偏在中性情緒評價。但這類對於價格的評論，卻又對消費者的滿意程度佔了很重要的角色。因此期望在進一步的研究中，可以針對這類價格評論也給予一個價格情緒分類，應可更準確的預估deal score的消費者滿意程度。

REFERENCES

- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, pages 6000–6010.
- Bart, Y., Shankar, V., Sultan, F., & Urban, G. L. (2005). Are the drivers and role of online trust the same for all web sites and consumers? A large-scale exploratory empirical study. *Journal of Marketing*, 69(4), 133–152.
- Brown, J., Broderick, A. J., & Lee, N. (2007). Word of mouth communication within online communities: Conceptualizing the online social network. *Journal of Interactive Marketing*, 21(3), 2–20.
- Brown, J. J., & Reingen, P. H. (1987). Social ties and word of mouth referral behavior. *Journal of Consumer Research*, 14(3), 350–362.
- Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, 43(3), 345–354.
- Cox, C., Burgess, S., Sellitto, C., & Buultjens, J. (2009). The role of user-generated content in tourists' travel planning behavior. *Journal of Hospitality Marketing & Management*, 18(8), 743–764.
- Cheung, C. M. K., Lee, M. K. O., & Rabjohn, N. (2008). The Impact of e-WOM: The adoption of online opinions in online customer communities. *Internet Research*, 18(3), 229–247.
- Cantalalpo, A. S., & Salvi, F. (2014). New consumer behavior: A review of research on eWOM and hotels. *International Journal of Hospitality Management*, 36, 41–51.
- Chatterjee, P. (2001). Online reviews: Do consumers use them? *Advances in Consumer Research*, 28(1), 129–133.
- Cox, C., Burgess, S., Sellitto, C., & Buultjens, J. (2009). The role of user-generated content in tourists' travel planning behavior. *Journal of Hospitality Marketing & Management*, 18(8), 743–764.
- Culotta, A. (2010). Towards detecting influenza epidemics by analyzing twitter messages. *Proceedings of the first workshop on social media analytics (SOMA'10)* (pp. 115–122). New York: ACM.
- Daniel S. Kostyra, Jochen Reinera, Martin Nattera, Daniel Klapper, (2016), *International Journal of Research in Marketing*, 33(1), 11–26.
- Dave, K., Lawrence, S. and Pennock, D. M. (2003). Mining the peanut gallery: Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the internet? *Journal of Interactive Marketing*, 18(1), 38–52.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- Kalchbrenner, N., Grefenstette, E., Blunsom, P. (2014). A Convolutional Neural Network for Modelling Sentences. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*. pp 655–665.
- Kusumasondajaja, S., Shanka, T., & Marchegiani, C. (2012). Credibility of online reviews and initial trust: The roles of reviewer's identity and review valence. *Journal of Vacation Marketing*, 18(3), 185–195.
- Lee, Y. (2009). Internet election 2.0? Culture, institutions, and technology in the Korean presidential elections of 2002 and 2007. *Journal of Information Technology & Politics*, 6(3), 312–325.
- Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management*, 29(3), 458–468.
- Liu, Y. (2006). Word-of-mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing*, 70(3), 74–89.
- Nasukawa, T. and Yi, J. (2003). Sentiment analysis: Capturing favorability using natural language processing. In *Proceedings of the K-CAP-03, 2nd International Conference on Knowledge Capture*.
- Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll and Manfred Stede (2011). Lexicon-based methods for sentiment analysis. *Computational linguistics*. 37(2):267–307.
- Ohanian, R. (1990). Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. *Journal of Advertising*, 19(3), 39–52.
- Pak, A.; Paroubek, P. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining.
- Pang, B.; Lee, L. (2005). Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. *Proceedings of the Association for Computational Linguistics (ACL)*: 115–124.
- Pang, B., & Lee, L. (2005). Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales. *Proceedings of the 43rd annual meeting on association for computational linguistics*. Association for Computational Linguistics. 115–124.
- Pang, B.; Lee, L. and Vaithyanathan, S. (2002). Thumbs up? Sentiment Classification using Machine Learning Techniques. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*: 79–86.
- Papathanassis, A., & Knolle, F. (2011). Exploring the adoption and processing of online holiday reviews: A grounded theory approach. *Tourism Management*, 32(2), 215–224.
- Prabowo, R., & Thelwall, M. (2009). Sentiment analysis: A combined approach. *Journal of Informetrics*, 3(2), 143–157.
- Qian, Q.; Huang, M.; Lei, J.; Zhu, X. (2016). Linguistically Regularized LSTMs for Sentiment Classification. *arXiv:1611.03949*
- Raffaele Filieri (2016). What makes an online consumer review trustworthy?. *Annals of Tourism Research*, 58(2016), 58–61.
- Ryu, K., & Han, H. (2010). Influence of the quality of food, service, and physical environment on customer satisfaction and behavioral intention in quick-casual restaurants: Moderating role of perceived price. *Journal of Hospitality and Tourism Research*, 34(3), 310–329.
- Socher, R.; Perelygin, A.; Wu, J. et al. (2013). Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*. pp 1631–1642.
- Tai, K. S.; Socher, R.; Manning, C. D. (2015). Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics*. pp 1556–1566.
- Vermeulen, I. E., & Seegers, D. (2009). Tried and tested: The impact of online hotel reviews on consumer consideration. *Tourism Management*, 30(1), 123–127.
- Wysocki, P. D. (2000). Private information, earnings announcements and trading volume, or stock chat on the internet: A public debate about private information. Working paper. University of Michigan Business School https://www.researchgate.net/publication/228431130_Private_Information_Earnings_Announcements_and_Trading_Volume_or_Stock_Chat_on_the_Internet_A_public_Debate_about_Private_Information_Working_Paper.
- Ye, Q., Law, R., & Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management*, 28(1), 180–182.
- Yoo, K.-H., & Gretzel, U. (2009). Comparison of deceptive and truthful travel reviews. In W. Höpken, U. Gretzel, & R. Law (Eds.), *Information and communication technologies in tourism 2009* (pp. 37–48). Vienna, Austria: Springer.
- Yu, X., Liu, Y., Huang, J. X., & An, A. (2012). Mining online reviews for predicting sales performance: A case study in the movie domain. *IEEE Transactions on Knowledge and Data Engineering*, 24(4), 720–734.
- Zhi-Ping Fan, Yu-Jie Che, Zhen-Yu Chen, 2017, *Journal of Business Research*, 74(2017), 90–100.